POLYPUBLIE
Polytechnique Montréal

POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

Titre: Analyse du recours aux correspondances dans un réseau de transport en commun à l'aide des données de cartes à puce
Title:

Auteur: Benjamin Disson
Author:

Date: 2020

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Disson, B. (2020). Analyse du recours aux correspondances dans un réseau de transport en commun à l'aide des données de cartes à puce [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/5214/
Citation:

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

URL de PolyPublie: https://publications.polymtl.ca/5214/
PolyPublie URL:

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

Programme: Maîtrise recherche en génie industriel
Program:

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal

https://publications.polymtl.ca

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Analyse du recours aux correspondances dans un réseau de transport en commun à l'aide des données de cartes à puce

BENJAMIN DISSON
Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Avril 2020

© Benjamin Disson, 2020.

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Ce mémoire intitulé :

Analyse du recours aux correspondances dans un réseau de transport en commun à l'aide des données de cartes à puce

présenté par **Benjamin DISSON**
en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Bruno AGARD, président
Martin TRÉPANIER, membre et directeur de recherche
Catherine MORENCY, membre et codirectrice de recherche
Vincent CADORET, membre

REMERCIEMENTS

En tout premier lieu, je tiens à remercier Martin Trépanier, directeur de recherche au sein du département de mathématiques et de génie industriel, et Catherine Morency, codirectrice de recherche au sein du département des génies civil, géologique et des mines, pour m'avoir permis de réaliser cette maîtrise à Polytechnique Montréal. Leurs connaissances du sujet et leurs nombreux conseils m'ont aidé à découvrir le monde de la recherche et à me passionner pour le domaine du transport.

Je remercie également Polytechnique Montréal et l'Institut National des Sciences Appliquées de Lyon, écoles dans lesquelles j'ai eu une formation complète par le biais de ce double diplôme. Les accords entre ces deux institutions m'ont permis d'enrichir à la fois mes connaissances techniques et culturelles au travers de la découverte d'un nouveau pays.

Ensuite, je souhaite remercier les partenaires de ce projet, à savoir le groupe Keolis, la Société Nationale des Chemins de fer Français (SNCF) et l'organisme Mitacs Accélération, pour leur apport financier et technique au sein de cette maîtrise. De même, je veux remercier la Société de transport de Montréal (STM) qui a rendu les données de cartes à puce disponibles, sans quoi le projet de recherche n'aurait pas pu être réalisé. J'espère que ces travaux pourront apporter de nouvelles connaissances aidant à leur amélioration. Pendant ce projet, j'ai également pu bénéficier du soutien technique de la Chaire Mobilité et du Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport (CIRRELT). Merci à ces deux structures et à leurs membres.

Cette maîtrise recherche m'a de plus permis de rencontrer plusieurs acteurs tout au long des événements et je remercie chacun d'entre eux. Ils m'ont apporté des connaissances, des idées, des conseils et du soutien, qui m'ont aidé à construire et finaliser mon projet. Parmi ces personnes, j'accorde une pensée particulière à mes professeurs, à Jean-Simon Bourdeau pour son aide avec les bases de données et à mes collègues de bureau Suzanne Pirie, Arame Niang et Camille Garnier pour leurs idées et leur patience.

Cette partie ne peut pas être complète sans que je remercie mes parents, ma famille et mes amis pour leurs encouragements et leur soutien moral tout au long de ce projet. Rien n'aurait été possible sinon. Merci tout particulièrement à Thomas Charpentier qui m'a beaucoup accompagné ces dernières semaines. Enfin, un énorme merci à Sabine Lefévère, qui m'a tant apporté.

RÉSUMÉ

Les sociétés de transport proposent des réseaux de plus en plus vastes et connectés pour répondre efficacement à la diversification des couples origine – destination. Seulement, un itinéraire direct n'étant pas possible pour chaque déplacement, les usagers doivent dans certaines situations recourir à une ou plusieurs correspondances. La correspondance est définie comme le passage d'un véhicule à un autre au cours d'un déplacement. Elle entraîne à la fois un effort physique et mental, une perte de temps et du stress pour les usagers, occasionnant une discontinuité du déplacement et diminuant l'attractivité du transport en commun (TC) au profit d'autres alternatives, notamment l'automobile particulière. Si les sociétés de transport veulent améliorer leur offre de service, il est essentiel de trouver des solutions pour limiter l'impact des correspondances.

Afin de proposer des solutions efficaces, il est nécessaire de comprendre comment les usagers ont recours aux correspondances. Historiquement, ces recherches sont réalisées à l'aide d'enquêtes auprès des voyageurs. Elles sont utilisées pour déterminer les attributs qui influencent les usagers dans leurs choix de faire une correspondance. Mais ces méthodes manquent de profondeur, car elles ne permettent pas d'observer les fluctuations de comportement des usagers. L'utilisation des données de cartes à puce apporte de nouvelles possibilités pour ces études, à la fois car l'échantillonnage obtenu est plus large que celui des enquêtes et car ce sont des données spatio-temporelles. Chaque fois qu'un voyageur valide sa carte à puce sur une borne de perception, la transaction est enregistrée avec des informations provenant de la carte et de la borne.

Ce mémoire de maîtrise développe une méthodologie universelle d'analyse de l'utilisation des correspondances à l'aide des données de transactions par cartes à puce. Cette méthode est capable de gérer des données récoltées quotidiennement sur l'intégralité d'un réseau multimodal. Pour ses expérimentations, le projet utilise les données de 10 000 cartes à puce de la Société de transport de Montréal (STM), pendant toute l'année 2016. Les cartes ont un abonnement mensuel ou annuel et sont sélectionnées aléatoirement. L'échantillon, non nécessairement représentatif de l'utilisation du réseau de la STM, est constitué de 9 550 725 transactions.

L'analyse est construite à travers quatre objectifs. En premier, un algorithme de transformation des transactions en déplacements est créé afin de détecter, comptabiliser et localiser l'ensemble des correspondances faites sur le réseau. Il permet de fournir les jeux de données nécessaires à la suite du projet. En particulier, l'algorithme traite 97,48 % des transactions,

fournissant une table de données de 5 887 074 déplacements avec 4 983 278 correspondances. Pour enrichir les données, deux éléments méthodologiques sont utilisés. Premièrement, les déplacements sont catégorisés selon leur composition en correspondances. Une série de modes est attribuée à chaque enregistrement, permettant de décrire avec un seul attribut le nombre de correspondances qui composent le déplacement, ainsi que leurs types et l'ordre dans lequel elles sont réalisées. Ensuite, le réseau est décrit à l'aide d'une typologie des lignes de bus. Chaque ligne est caractérisée selon son niveau de service, sa localisation et ses possibles interactions avec les autres lignes du réseau.

Le second objectif s'intéresse à la fluctuation quotidienne de l'utilisation des correspondances. Il est question d'expliquer comment les usagers de l'échantillon ont recours aux correspondances dans leurs déplacements quotidiens. Pour cela, la répartition des déplacements est d'abord étudiée. Elle met en évidence les variations d'utilisations du réseau entre tous les jours de la semaine. Les heures de pointe sont identifiées, montrant un écart entre les jours ouvrés et les jours de fins de semaine. Ensuite, l'analyse de la proportion des déplacements de chaque série de modes par heure montre que le bus et le métro sont autant utilisés (respectivement 33,04 % et 32,51 % des déplacements), mais que le métro est favorisé en pointe alors que le bus est favorisé hors pointe. De même, les usagers de l'échantillon ont tendance à faire plus de correspondances pendant leurs déplacements en pointe. Ensuite, une analyse de la symétrie des déplacements successifs est réalisée. Elle permet de décrire la répartition des activités générées par le TC et révèle que 61,11 % des passagers avec deux déplacements quotidiens font un aller-retour, dont 50,09 % avec la même combinaison de lignes dans les deux directions. Enfin, elle témoigne que l'utilisation du bus est plus variable que celle du métro.

Pour poursuivre l'analyse de l'utilisation des correspondances, le troisième objectif est la caractérisation du comportement des usagers de l'échantillon. Une segmentation par la méthode des k-moyennes est réalisée sur un vecteur composé séquences de correspondances dans les déplacements quotidiens de chaque carte. Il en résulte sept groupes marqués par une utilisation différente des correspondances dans les déplacements. La variabilité intragroupe montre que les voyageurs favorisant des déplacements avec une correspondance sont moins stables que ceux favorisant des trajets directs (en moyenne respectivement $\pm$ 12,07 % contre $\pm$ 6,05 % de variation). De plus, une fluctuation plus élevée est observée chez les passagers favorisant le métro plutôt que le bus. Une matrice de probabilité de changement de groupe explique que les usagers favorisent en priorité un mode de transport plutôt qu'un type de correspondance. Ensuite, un regroupement des données en semaines et un diagramme de Sankey illustrent la régularité des passagers de l'échantillon. Ils sont stables pendant les jours ouvrés, mais ont un comportement variable la fin de semaine, où ils se déplacent moins et favorisent davantage

le bus. Enfin, l'analyse de l'utilisation des lignes permet de positionner géographiquement les groupes. Elle souligne que les voyageurs favorisant le bus sont répartis sur tout le territoire de l'Île de Montréal alors que ceux privilégiant le métro restent principalement dans les régions centrales.

Enfin, le dernier objectif est la caractérisation de l'utilisation des lignes du réseau. Il s'agit de les étudier afin de détecter celles qui génèrent le plus de correspondances. Un indicateur d'autonomie est créé pour classer les lignes et des graphes sont développés pour représenter schématiquement leurs interactions. Cette étude montre que l'indicateur dépend à la fois du service et de la géolocalisation des lignes. Celles dont l'autonomie est grande sont caractérisées par un service spécifique, une large couverture spatiale ou par peu d'interactions avec le reste du réseau. En moyenne, avec l'échantillon étudié, les lignes de bus sont plus autonomes que les lignes de métro, respectivement 0,68 contre 0,55. De plus, une régression linéaire multiple prouve que l'indice d'autonomie d'une ligne peut être expliqué à l'aide des variables de sa typologie. Cela permet à la fois d'identifier les lignes à prioriser pour augmenter l'offre de service et d'estimer l'influence des futurs parcours proposés. Enfin, à l'aide des graphes il devient possible de localiser les principaux arrêts impliqués dans les correspondances bus – bus et de classer les stations selon leur taux de correspondances bus – métro, métro – bus et métro – métro. Cette visualisation met en évidence les lignes importantes et montre que les stations générant le plus de correspondances sont celles situées sur les grands boulevards et aux terminus.

En conclusion, ce mémoire contribue à l'amélioration des méthodes d'analyses comportementales des usagers en incluant dans les modèles existants la prise en compte des correspondances effectuées. Il fournit une nouvelle méthodologie de traitement des données de cartes à puce apportant des connaissances sur l'impact des correspondances, à la fois sur le réseau et sur ses usagers. Toutefois, des limites techniques liées au manque d'informations disponibles dans les données brutes et au développement du modèle offrent de nouveaux défis à relever et des perspectives d'études élargies.

ABSTRACT

Transit organizations offer connected and large networks to effectively answer the diversification of origin – destination pairs. But because a direct trip is not always possible, travelers must sometimes use several transfers to reach their destination. A transfer is defined as the change from a vehicle to another during a trip. It generates a physical and mental effort, a waste of time as well as some stress for users. All of which induce a trip discontinuity and reduce the attractiveness of public transportation in favor of other alternatives, especially private cars. If public transit agencies want to improve their service, it is fundamental for then to find solutions to reduce the impacts caused by transfers within the networks.

In order to propose efficient solutions, it is necessary to understand how travelers use transfers. Historically, such researches are achieved with passenger surveys. They are conducted to determine the attributes that influence the choice to make a transfer for users. But those methods are limited because they do not allow to observe passengers behaviors fluctuations. The use of smart card data brings new possibilities for these studies, both because the sample is larger and because data are spatiotemporal. Each time a traveler validates his card on a fare collection device, the transaction is registered with information from the card and the device.

This master's thesis develops a global methodology to analyse the use of transfers through smart cards fare collection data. This method is able to handle daily recorded data from an entire network. The practical applications are made from data of 10,000 smart cards from the Société de transport de Montréal (STM) for the year 2016. These cards have a monthly or a yearly subscription and are randomly selected. The sample, not necessarily representative of the use of the STM's network, is made up of 9,550,725 transactions.

The analysis is built around four objectives. The first is the development of an algorithm to transform the original transaction database into a trip database, in order to detect, count and locate all the transfers made into the network. It allows to construct the databases useful for the rest of the project. Particularly, the algorithm processes 97.48 % of the transactions, giving 5,887,074 trips with 4,983,278 transfers. To enrich the databases, two methodological elements are used. First, trips are put into categories according to their transfers composition. A sequence of modes is assigned to each trip, allowing to know with only one attribute the number of transfers, as well as their types and the order in which they are made. Afterwards, the network is described using a bus line typology. Each line is characterized according to its level of service, its location and the feasible connections with the other lines of the network.

The second objective takes an interest in the daily fluctuation of the use of transfers. It is about explaining how travelers from this sample resort to transfers in their daily trips. To do so, the distribution of trips is firstly studied. It highlights the variations in the network use. The peak times are identified, showing a difference between weekdays and weekends. Then the analysis of the proportion of trips according to the time of the day and the sequence of modes reveals that the bus and the metro (subway) are equally used in this sample, with respectively 33.04 % and 32.51 % of the trips. But the metro is privileged during the peak periods while the bus is privileged during off-peak periods. Likewise, users in this sample tend to do more transfers during the peak periods. Next, the symmetry between successive trips is considered. It allows to describe the distribution of activities generated by public transit and reveals that 61.11 % of passengers from this sample with two trips in a day use it to make return trips, among which 50.09 % with the same combination of lines in both directions. Finally, it shows that in this sample there are more fluctuation in the bus utilisation than in the metro utilisation.

To continue the analysis of the use of transfers, the third objective is the description of the passengers behaviors. One segmentation with the k-means method is made on a vector built with the daily trips types from each card. This leads to seven clusters showing a different use of transfers in the trips. The within-group variability proves that in this sample, users who favor trips with one transfer are less stable than the ones which favor directs trips (on average respectively $\pm$ 12.07 % to $\pm$ 6.05 % of variation). Additionally, a higher fluctuation is observed among travelers that privilege the use of metro rather than the bus. A group change probability matrix explains that users from this sample favor in priority a transportation mode rather than a transfer type. Then, a data merging from days to weeks and a Sankey diagram illustrate the passengers' regularity. They are stable within the weekdays but adopt a different behavior in the weekend, when they make fewer trips and favor the use of the bus. Finally, the analysis of the lines used allows to determine the geographical position of the clusters. It highlights that travelers who favor the bus are distributed all around the island of Montreal while the ones who favor the metro stay primarily in the central areas.

Finally, the last objective is the characterization of the network use. It is about studying the lines in order to detect the ones that generate more transfers. An independence indicator is created to rank the lines and graphs are developed to schematically represent the interactions between them. This study shows that the independence is linked to both the service and the geolocation of the lines. The ones with the highest independence indicator are the ones defined by a specific service, a large coverage or by few interactions with the rest of the network. On average from this sample, bus lines are more independent than the metro

ones, respectively 0.68 to 0.55. Furthermore, a multiple linear regression proves that the independence indicator of a line can be explained using the variables of its typology. This allows both to determine which lines to prioritize to provide a better service and to estimate the influence of future routes. Last, lines needing an improvement are deduced from the graphs and some solutions are listed. It is then possible to identify the principal stops used for bus to bus transfers and to classify each subway station according to its bus to metro, metro to bus and metro to metro transfers ratios. This view highlights important lines and shows that stations who generate the most transfers are located on the main boulevards and on terminuses.

To conclude, this master's thesis contribute to the improvement of the methods to analyse users behavior, by including in the existing models the consideration of the transfers realized. It provides a new methodology to process smart cards fare collection data bringing knowledge about the impact of transfers, both on the network and its travelers. Nevertheless, technical limitations linked with the lack of information in the original database and the model's development offer new challenges and perspectives for future research.

TABLE DES MATIÈRES

REMERCIEMENTS iii

RÉSUMÉ iv

ABSTRACT vii

TABLE DES MATIÈRES x

LISTE DES TABLEAUX xiii

LISTE DES FIGURES xiv

LISTE DES SIGLES ET ABRÉVIATIONS xvi

LISTE DES ANNEXES xviii

CHAPITRE 1 INTRODUCTION 1
- 1.1 Mise en contexte 1
- 1.2 Problématique et objectifs 2
- 1.3 Structure du mémoire 2

CHAPITRE 2 REVUE DE LITTÉRATURE 4
- 2.1 Cartes à puce 4
  - 2.1.1 Présentation générale 5
  - 2.1.2 Utilisation des cartes à puce en planification des transports 6
  - 2.1.3 Avantages et inconvénients 8
- 2.2 La correspondance en transport en commun 10
  - 2.2.1 Définition d'une correspondance 10
  - 2.2.2 Attributs influençant la correspondance 13
  - 2.2.3 Pénalité de correspondance 15
  - 2.2.4 Perception du temps 16
  - 2.2.5 Solutions mises en place pour limiter l'impact des correspondances . . 18
- 2.3 Éléments méthodologiques 19
  - 2.3.1 Comportement des usagers 19
  - 2.3.2 Performance des réseaux 20
- 2.4 Conclusion de la revue de littérature 21

CHAPITRE 3 DONNÉES ET MÉTHODOLOGIE 23
3.1 Structure de l'étude 23
3.2 Description des données 24
3.2.1 Réseau de la STM 25
3.2.2 Échantillon des données de cartes à puce 28
3.2.3 Informations disponibles dans la table des données brutes 29
3.3 Algorithme de transformation des transactions 32
3.3.1 Étape 1 : Préparation des données 33
3.3.2 Étape 2 : Création des séquences de transactions 33
3.3.3 Étape 3 : Déduction des possibilités de correspondance 34
3.3.4 Étape 4 : Localisation et comptage des correspondances en métro . . 35
3.3.5 Étape 5 : Enrichissement de la table de données 36
3.3.6 Étape 6 : Passage en déplacements 36
3.3.7 Exemple explicatif des étapes de l'algorithme 37
3.4 Typologie des lignes de bus 40
3.5 Classification des déplacements : séries de modes 42
3.6 Analyse des déplacements quotidiens 44
3.6.1 Répartition des correspondances 44
3.6.2 Symétrie des déplacements pendulaires 45
3.7 Analyse du comportement des usagers de l'échantillon 48
3.7.1 Création des vecteurs de la segmentation 48
3.7.2 Choix du nombre de groupes 50
3.7.3 Segmentation des cartes-jours 54
3.7.4 Analyse des résultats de la segmentation des cartes-jours 54
3.8 Analyse de l'utilisation des lignes du réseau 55
3.8.1 Indice d'autonomie d'une ligne 56
3.8.2 Couples de lignes liées par une correspondance 58
3.8.3 Principaux arrêts et stations de correspondance 59

CHAPITRE 4 EXPÉRIMENTATIONS ET RÉSULTATS 61
4.1 Analyse descriptive des données 61
4.1.1 Résultats de l'algorithme VALIDEP 61
4.1.2 Distribution annuelle des déplacements 62
4.1.3 Distribution quotidienne des déplacements 64
4.2 Composition des déplacements quotidiens 65
4.2.1 Répartition des correspondances 65

4.2.2 Écart de temps entre deux déplacements 70
4.2.3 Symétrie des déplacements pendulaires 73
4.3 Comportement des usagers de l'échantillon étudié 79
4.3.1 Groupes de la segmentation 80
4.3.2 Fluctuation de composition des groupes 82
4.3.3 Répartition hebdomadaire des cartes-jours 87
4.3.4 Positionnement géographique des groupes 89
4.4 Utilisation des lignes du réseau par les usagers de l'échantillon 93
4.4.1 Indice d'autonomie des lignes 93
4.4.2 Interactions entre les lignes 100
4.4.3 Détection des principaux arrêts et stations de correspondances 107

CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS. 113
5.1 Synthèse des travaux 113
5.2 Limitations 117
5.3 Perspectives de recherches 120

RÉFÉRENCES 122

ANNEXES 128

LISTE DES TABLEAUX

Tableau 2.1 Usages des données de cartes à puce à des fins de planification 7
Tableau 2.2 Avantages et inconvénients de l'utilisation de la carte à puce 8
Tableau 2.3 Avantages et inconvénients des données de cartes à puce 9
Tableau 2.4 Attributs diminuant l'attractivité des correspondances 14
Tableau 3.1 Utilisation du réseau par type de support en 2016 27
Tableau 3.2 Informations recueillies lors de chaque transaction 30
Tableau 3.3 Transactions du cas d'étude 37
Tableau 3.4 Table des transactions du cas d'étude après l'étape 2 39
Tableau 3.5 Table des transactions du cas d'étude après l'étape 4 39
Tableau 3.6 Déplacement obtenu à partir des transactions du cas d'étude 39
Tableau 3.7 Typologie des lignes de bus 42
Tableau 3.8 Classification des déplacements : séries de modes 43
Tableau 3.9 Vecteurs des cartes-jours pour la segmentation 50
Tableau 3.10 Table des correspondances entre chaque paire de lignes 58
Tableau 4.1 Proportion des cartes-jours selon l'heure des deux déplacements . . . 74
Tableau 4.2 Symétries des séries de modes 75
Tableau 4.3 Proportion des cartes-jours avec un ajout ou un retrait 76
Tableau 4.4 Symétries des lignes utilisées pour les deux déplacements 77
Tableau 4.5 Lignes majoritairement impliquées dans les symétries 78
Tableau 4.6 Centres des groupes de la segmentation des cartes-jours 80
Tableau 4.7 Matrice de probabilité de passage d'un groupe à un autre 84
Tableau 4.8 Combinaisons de régions des lignes de la STM 90
Tableau 4.9 Variables de la régression linéaire multiple 99
Tableau 4.10 Résultats globaux de la régression linéaire multiple 99
Tableau 4.11 Contribution des variables de la régression linéaire multiple 100
Tableau D.1 Table des transactions pendant l'algorithme VALIDEP 134
Tableau D.2 Table des transactions enrichies avec les correspondances 135
Tableau D.3 Table des déplacements enrichis avec les correspondances 136

LISTE DES FIGURES

Figure 2.1 Exemple de déplacement avec une correspondance 11
Figure 3.1 Méthodologie générale de l'étude 24
Figure 3.2 Organisation du transport en commun de la grande région de Montréal 26
Figure 3.3 Processus de sélection de l'échantillon de données 29
Figure 3.4 Algorithme de transformation des transactions en déplacements . . . 32
Figure 3.5 Succession des transactions du cas d'étude 38
Figure 3.6 Régions de l'Île de Montréal 41
Figure 3.7 Arbre de décision pour évaluer la symétrie des séries de modes 46
Figure 3.8 Arbre de décision pour évaluer la symétrie des lignes utilisées 47
Figure 3.9 Méthodologie utilisée pour la segmentation des usagers 49
Figure 3.10 Analyse de la distorsion des groupes 52
Figure 3.11 Analyse des sauts d'inertie du dendrogramme 53
Figure 3.12 Dendrogramme avec $k = 6$ groupes 53
Figure 3.13 Nomenclature des graphes d'interactions des lignes 59
Figure 4.1 Distribution annuelle des déplacements 62
Figure 4.2 Proportion des déplacements contenus dans l'échantillon étudié . . . 63
Figure 4.3 Distribution quotidienne des déplacements 64
Figure 4.4 Répartition des déplacements selon la série de modes 66
Figure 4.5 Répartition des déplacements des séries "B,M" et "M,B" 68
Figure 4.6 Répartition des déplacements des séries "B,M,M" et "M,M,B" 69
Figure 4.7 Nombre de déplacements par carte-jour 70
Figure 4.8 Écarts de temps entre deux déplacements consécutifs 71
Figure 4.9 Matrice de répartition des durées d'activités 72
Figure 4.10 Écarts de temps entre deux déplacements selon les tranches horaires . 74
Figure 4.11 Lignes fréquemment utilisées par les usagers symétriques 79
Figure 4.12 Caractéristiques des déplacements de chaque groupe 81
Figure 4.13 Variation de la taille des groupes 83
Figure 4.14 Proportion de cartes dans les groupes chaque jour 86
Figure 4.15 Répartition quotidienne des cartes dans les groupes 87
Figure 4.16 Diagramme de Sankey de la semaine 42 88
Figure 4.17 Répartition spatiale des lignes utilisées dans chaque groupe 91
Figure 4.18 Distribution de l'indice d'autonomie des lignes selon leur classification 94
Figure 4.19 Variabilité de l'indice d'autonomie des lignes selon leurs régions . . . 96

Figure 4.20 Matrice de corrélation des variables de la typologie des lignes de bus . 97
Figure 4.21 Graphe des interactions de tous les types de correspondances 101
Figure 4.22 Carte des lignes qui ont le plus de correspondances 102
Figure 4.23 Graphe des correspondances entre deux bus 103
Figure 4.24 Graphe des correspondances d'un bus vers un métro 104
Figure 4.25 Graphe des correspondances d'un métro vers un bus 105
Figure 4.26 Graphe des correspondances entre deux métros 106
Figure 4.27 Répartition géographique des correspondances bus – bus 108
Figure 4.28 Répartition géographique des correspondances bus – métro 109
Figure 4.29 Répartition géographique des correspondances métro – bus 110
Figure 4.30 Répartition géographique des correspondances métro – métro 111
Figure A.1 Configurations de stations optimisant les correspondances 128
Figure A.2 Horaires cadencés avec quatre points focaux 128
Figure A.3 Réseaux de métro optimisant les correspondances 129
Figure B.1 Plan du réseau de bus et de métro de la STM 130
Figure B.2 Plan du métro de Montréal 131
Figure C.1 Cartes de transport de la STM 132
Figure C.2 Grille tarifaire de la STM 133
Figure E.1 Lignes selon leur localisation dans les régions de Montréal 137
Figure E.2 Lignes selon leur localisation dans les régions de Montréal (suite) . . 138

LISTE DES SIGLES ET ABRÉVIATIONS

AFC	*Automated fare collection*
ARTM	Autorité régionale de transport métropolitain
BB	Correspondance d'un bus vers un bus
BM	Correspondance d'un bus vers un métro
CEB	Code d'emplacement billettique
CHA	Classification hiérarchique ascendante
CIRRELT	Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport
CMJ	Carte magnétique jetable
CMM	Communauté métropolitaine de Montréal
CPCT	Carte OPUS commune de transport
CPO	Carte à puce occasionnelle
CSV	*Comma-separated values*
CTM	Commission de transport de Montréal
CV	Centre-ville
GIS	*Geographic information system*
GPS	*Global positioning system* – Système de localisation mondial
GTFS	*General Transit Feed Specification*
IVT	*In Vehicle Time*
MB	Correspondance d'un métro vers un bus
MM	Correspondance d'un métro vers un métro
MSE	*Mean squared error* – Erreur quadratique moyenne
OD	Origine – destination
OPTC	Organisme public de transport en commun
QGIS	*Quantum Geographic Information System*
RDS	`R` *data serialization*
REM	Réseau express métropolitain
RFID	*Radio Frequency Identification*
RTC	Réseau de transport de la Capitale
RTL	Réseau de transport de Longueuil
SCFCD	*Smart card fare collection data*
SNCF	Société Nationale des Chemins de fer Français
STL	Société de transport de Laval

ST LÉVIS	Société de transport de Lévis
STM	Société de transport de Montréal
TC	Transport en commun
TCL	Transport en commun lyonnais

LISTE DES ANNEXES

Annexe A Exemples d'optimisation des correspondances 128

Annexe B Plan du réseau de bus et de métro de la STM 130

Annexe C Titres de transport proposés par la STM 132

Annexe D Tables des transactions et des déplacements 134

Annexe E Lignes de chaque combinaison de régions 137

Annexe F Liste des jours atypiques de l'année 2016 139

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

L'exode rural associé à un accroissement démographique élevé entraîne depuis plusieurs années un agrandissement et une densification des villes. À cette date, plus de la moitié de la population mondiale habite en zone urbaine et ce chiffre devrait atteindre deux tiers avant 2050 [1]. Cette expansion rapide représente à la fois un enjeu social et politique incontournable. Elle est également un défi pour la mobilité des personnes puisqu'elle génère une hausse du nombre de déplacements ainsi qu'une diversification des couples origine – destination. Pour répondre à cette croissance, les sociétés de transport travaillent à proposer des réseaux plus efficaces et mieux connectés. Néanmoins, un itinéraire direct n'étant pas possible pour chaque déplacement, les usagers doivent dans certains cas recourir à une ou plusieurs correspondances.

Une correspondance est définie par un changement de véhicule au cours d'un déplacement, qu'il soit question de deux véhicules du même mode (deux bus par exemple) ou de deux modes différents (un bus et un métro) [2]. Elle est reconnue par de nombreux auteurs comme étant un frein à l'utilisation du transport en commun, car elle entraîne à la fois un effort physique, un temps d'attente et une discontinuité dans le déplacement. Elle est une source de stress pour les usagers, diminuant l'attractivité du transport en commun au profit d'autres moyens de transport. Dans les dernières années, afin de minimiser l'impact de la correspondance, de nombreux chercheurs ont souhaité expliquer comment les voyageurs la perçoivent et la façon dont ils l'utilisent.

Les analyses des correspondances sont majoritairement réalisées à l'aide de données d'enquêtes auprès des voyageurs. Cependant, une nouvelle technologie liée à l'arrivée des données massives (*Big Data*) agrandit le champ des possibles. Il s'agit de la carte à puce. Cette innovation apparue au cours des années 1990 pour améliorer les systèmes de perception permet aujourd'hui la production de nombreuses informations spatio-temporelles sur une partie des déplacements des voyageurs. Encore sous-exploitées pour étudier les correspondances, les données de cartes à puce deviennent graduellement indispensables aux yeux des planificateurs, qui y voient un nouveau moyen d'analyser l'utilisation du réseau et les habitudes comportementales des usagers.

Ce mémoire s'inscrit dans une problématique de valorisation des données opérationnelles en transport collectif. Il résulte d'un projet relatif à l'étude de l'offre et de la demande en partenariat avec la Société Nationale des Chemins de fer Français (SNCF) et la société Keolis. Sa réalisation est basée sur des données de cartes à puce mises à disposition par la Société de transport de Montréal (STM). Il vise ainsi le développement d'algorithmes d'analyse de l'utilisation des réseaux de bus et de métro pour apporter de nouvelles connaissances liées aux correspondances entre les deux modes.

1.2 Problématique et objectifs

Ce projet donne suite à un partenariat entre Polytechnique Montréal et la Société de transport de Montréal, dans le cadre du programme de recherche de la Chaire Mobilité. L'objectif principal est d'utiliser des données de transactions par cartes à puce (dites "brutes") de la STM afin d'élaborer une méthode universelle d'étude de l'utilisation des correspondances. Cette méthode devra être capable de gérer des données récoltées quotidiennement dans les bus et les stations de métro d'un réseau au complet.

Il s'agit de fournir aux planificateurs du transport collectif une meilleure compréhension de l'impact des correspondances sur le réseau et ses usagers. Pour cela, cette recherche se base sur quatre objectifs spécifiques :

- développer un algorithme de détection, de comptage et de localisation des correspondances afin de transformer les données de transactions en données de déplacements;
- mesurer la variabilité d'utilisation des correspondances au sein des déplacements quotidiens afin de caractériser la pendularité des usagers;
- élaborer une méthode permettant l'analyse comportementale des usagers du transport en commun selon leur utilisation des correspondances;
- proposer une étude du réseau afin de repérer les lignes qui génèrent des correspondances et dénombrer leurs interactions.

1.3 Structure du mémoire

Le présent mémoire est constitué de cinq chapitres. Tout d'abord, cette introduction définit le contexte de l'étude et le domaine d'application. La problématique et les objectifs qui s'en suivent sont clairement énoncés.

Le chapitre 2 propose une revue de littérature sur les thèmes liés au domaine d'étude. La technologie des systèmes de cartes à puce est expliquée et ses applications en planification des transports sont stipulées. De la même façon, une mise en avant des travaux effectués au sujet des correspondances est faite. Elle vise à clairement définir l'objet étudié ainsi que les créneaux de recherche à exploiter. Enfin, une partie de la revue de littérature est consacrée aux méthodes utilisées pour caractériser les usagers du transport en commun et les réseaux.

Ensuite, le chapitre 3 décrit les données de cartes à puce utilisées dans le cadre de cette recherche, la méthodologie générale du projet ainsi que les éléments méthodologiques associés à chaque objectif. Ce chapitre suit l'ordre chronologique du raisonnement employé pour résoudre les différentes problématiques de la recherche.

Les expérimentations et les résultats d'application de la méthodologie appliquée sur un échantillon de données de cartes à puce de la STM sont finalement présentés dans le chapitre 4. L'application de l'algorithme de transformation des transactions permet d'obtenir une table des déplacements enrichie avec les correspondances. Une mesure de la variation d'utilisation des correspondances selon l'heure des déplacements met en avant les activités générées par le transport en commun (TC) et montre que les usagers de l'échantillon effectuent majoritairement des allers-retours symétriques. Le comportement des usagers face aux correspondances est aussi analysé, mettant en avant plusieurs groupes composés de caractéristiques différentes et illustrant les changements quotidiens d'agissements des voyageurs. L'analyse de l'utilisation des lignes permet quant à elle la détection des lignes à forte autonomie et la mise en évidence des couples les plus influents et des arrêts les plus utilisés dans des correspondances.

Finalement, le chapitre 5 conclut ce mémoire en exposant les contributions de cette recherche ainsi que ses limitations. Plusieurs perspectives d'application et d'amélioration sont proposées pour de futures recherches.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce deuxième chapitre présente un historique de la littérature en lien avec le sujet de recherche. Son objectif est de dresser un état de l'art des connaissances et des méthodologies développées lors de travaux ultérieurs. Cette revue permet ainsi d'apporter une meilleure compréhension des questions abordées dans ce mémoire. De plus, la mise en avant des lacunes dans les travaux cités permet de justifier les approches méthodologiques qui seront développées dans la suite de ce document.

La revue de littérature débute par une présentation générale de la carte à puce. Il s'agit d'exposer son fonctionnement, son utilité dans les études en transport en commun (TC), ainsi que de détailler ses forces et ses faiblesses. Dans un second temps, un portrait des études portant sur les correspondances en transport en commun est brossé. La correspondance est définie suivant plusieurs aspects. Les attributs qui impactent les voyageurs sont énoncés et les notions de pénalité de correspondance et de valeur du temps perçu sont expliquées. Pour finir, ce chapitre s'intéresse aux méthodes principalement employées pour modéliser le comportement des usagers du TC ainsi que l'utilisation du réseau.

2.1 Cartes à puce

Bien que de plus en plus répandue dans les réseaux de transport du monde entier, la carte à puce n'est pas uniquement utilisée pour se déplacer. L'idée de la carte à puce a été brevetée pour la première fois en 1968 par les inventeurs allemands Helmut Gröttrup et Jürgen Dethloff. Elle a été plus amplement développée dans les années suivantes à la fois au Japon et en France. Sa forme finale est due à l'ingénieur français Rolland Moreno en 1974 [3]. La carte à puce consistait initialement en un petit circuit mémoire intégré à l'intérieur d'une fine carte en plastique de quelques centimètres carrés. Ses premières applications commerciales ont vu le jour dans les années 1990. D'abord utilisée par les structures bancaires pour simplifier les méthodes de paiement, elle a très vite suscité de l'intérêt dans plusieurs secteurs industriels. Quarante-cinq ans plus tard, la carte à puce existe sous une multitude de formes et est utilisée mondialement dans de nombreux domaines. Entre autres, elle prend la forme de porte-monnaie électronique, de carte d'identité, de carte de fidélité, etc. [4]. Au début du XXIe siècle, plus de trois milliards de cartes à puce étaient utilisées [5]. Dans ce mémoire, l'intérêt est porté sur l'utilisation des cartes à puce comme outil de planification des transports en commun. C'est cette application de la technologie que présente la suite de cette section.

2.1.1 Présentation générale

Les sociétés de transport ont intégré les cartes à puce dans leurs systèmes au cours des années 1990. Elles sont semblables aux cartes de crédit et intègrent une puce RFID ayant pour but de stocker, de traiter et d'écrire des données. Lorsqu'une carte à puce est en contact avec un lecteur, elle communique ses informations par le biais d'une onde électromagnétique. Plusieurs technologies de cartes existent, dont les différences sont principalement liées à la conception et l'implantation du microprocesseur [6]. Le fonctionnement est quant à lui similaire d'un système à l'autre.

Un système de paiement par carte à puce se compose de deux entités : la carte (1) et le lecteur (2). D'un côté, chaque voyageur possède une carte avec un identifiant unique qu'il utilise pour stocker ses titres de transport. De l'autre, des bornes automatiques sont installées à l'entrée de tous les bus (et les trams selon les villes) et dans les stations de métro. Lorsque le voyageur utilise un de ces modes, il n'a qu'à valider sa carte sur la borne qui autorisera ou non l'accès selon le titre disponible. Dans certaines villes, des bornes se situent également à la sortie des moyens de transport. Les voyageurs valident ainsi en entrant et en sortant du mode. C'est en général le cas dans les réseaux de transport où la tarification du déplacement se fait en fonction de la distance parcourue [7].

Le rôle premier des systèmes de paiement par cartes à puce est le contrôle des recettes clientèle en raison de la simplification des grilles tarifaires [8]. De surcroît, les cartes à puce remplacent progressivement les traditionnels tickets et cartes magnétiques en offrant des options de paiement plus sécuritaires et fiables [6]. Cela en fait un système avantageux à la fois pour les voyageurs et les opérateurs. Chaque carte est rechargeable selon les conditions tarifaires des sociétés de transport. Elles s'adaptent ainsi aux usagers (étudiant, adulte régulier, personne âgée, etc.) et à leurs besoins de déplacement (passage unique, journée, abonnement mensuel, etc.). De plus, ce système diminue le taux de fraude et simplifie le travail des chauffeurs de bus qui ont moins souvent besoin de gérer de la monnaie [4]. Cette méthode de perception permet ainsi de diminuer le temps d'embarquement des usagers dans les véhicules, ce qui fiabilise le réseau.

Les systèmes de perception par cartes à puce sont également employés pour récolter en continu des données transactionnelles individualisées offrant de nouvelles perspectives d'études et de planification [8]. Lors de la validation de la carte sur le lecteur, un échange d'informations se fait, permettant de fusionner à la fois les données contenues par la borne (les données opérationnelles telles que l'heure, le mode, la ligne, etc.) et les données contenues par la carte (les données relatives à l'utilisateur telles que le titre utilisé). Chaque transaction est ainsi

enregistrée, permettant d'analyser l'utilisation du réseau. Ces données ont un fort potentiel pour les planificateurs, car elles apportent des informations en temps réel servant à la fois l'opération quotidienne et la planification à long terme [4]. De plus, la qualité de ces données dépasse celle des systèmes de comptage tout en respectant l'anonymat des usagers [9].

Pour toutes ces raisons, l'utilisation des cartes à puce est maintenant légion partout dans le monde. C'est en Europe que ces systèmes sont les plus courants, notamment en France, au Royaume-Uni et en Italie. De même, plusieurs villes du Québec les ont adoptés, en particulier Gatineau, Québec et Montréal. Enfin, ces systèmes se répandent petit à petit dans les pays émergents, tels qu'à Santiago au Chili [10].

2.1.2 Utilisation des cartes à puce en planification des transports

L'emploi des données de cartes à puce pour la recherche et le développement du transport en commun ne cesse d'augmenter. Comme elles fournissent à la fois des informations sur les cartes et sur les véhicules, elles servent autant pour modéliser le comportement des passagers que pour modéliser l'utilisation d'un réseau [4]. Cette section recense quelques-unes des études réalisées à partir des données de cartes à puce. Ces recherches sont positionnées dans le tableau 2.1 selon leurs principaux objectifs.

Deux familles d'études se distinguent. D'une part, certaines visent à analyser et expliquer les comportements des voyageurs. Pour cela, beaucoup d'auteurs se sont intéressés aux techniques d'exploration de données (de l'anglais *data mining*), pour obtenir des groupes d'usagers similaires en fonction d'attributs variant d'une étude à l'autre. Dans certains cas, des données exogènes sont également ajoutées. De même, des chercheurs mesurent la variabilité du comportement des passagers. Le but est d'expliquer le changement de comportement des voyageurs selon le temps, l'espace ou d'autres caractéristiques. Ils s'intéressent pareillement à la loyauté des utilisateurs, c'est-à-dire leur constance dans l'utilisation du réseau.

D'autre part, de nombreuses études ont recours aux données de carte à puce pour analyser l'aspect opérationnel du transport en commun. Il s'agit d'obtenir des informations sur l'utilisation des lignes du réseau ainsi que de mesurer les impacts d'attributs exogènes. Ces études sont regroupées en quatre groupes. Tout d'abord, plusieurs chercheurs ont développé des méthodes d'estimation du lieu de descente. Cela permet d'enrichir les modèles de prédiction de la demande et le suivi des utilisateurs. Des études sont également menées pour mesurer le taux d'utilisation des lignes. Elles permettent d'identifier les lignes fortement achalandées et d'observer l'évolution des charges à bord selon l'heure de la journée, le type de jour ou encore

Tableau 2.1 Usages des données de cartes à puce à des fins de planification

Auteurs	Comportement des usagers		Utilisation du réseau			
	Segmentation des usagers	Fidélité et variabilité	Lieu de descente	Utilisation des lignes	Fiabilité et performance	Correspondances
Agard et al. [11]	×	×				
Briand et al. [12]	×					
Deschaintres et al. [13]	×	×				
Descoimps et al. [14]		×		×		
Devillaine et al. [15]		×	×			
Dubos-Golain et al. [16]	×				×	
Ghaemi et al. [17]	×					
Hao et Lee [18]					×	×
He et al. [19–21]	×		×	×		
Hofmann et O'Mahony [22]			×			×
Li et al. [23]	×		×			×
Morency et al. [24]	×	×				
Nishiuchi et al. [25]					×	×
Sun et al. [26]	×					×
Sun et Schonfeld [27]					×	×
Toqué et al. [28]				×		
Trépanier et al. [29]			×			
Trépanier et al. [30]				×	×	
Trépanier et Morency [31]		×		×		
Trépanier et al. [32]		×				
Viallard et al. [33]	×	×				
Yen et al. [34]	×					×

selon la météo. D'autres chercheurs créent des indicateurs pour mesurer la performance et la fiabilité des lignes. Par exemple, il s'agit de calculer la vitesse commerciale d'un bus ou de comptabiliser les arrêts de service. Enfin, plusieurs études intègrent les données de cartes à puce dans la modélisation des correspondances. Ces dernières applications de la carte à puce sont présentées plus en profondeur à la section 2.2.

Il est à noter que le tableau 2.1 présente des recherches utilisant les cartes à puce comme source de données, car c'est cet aspect de la technologie qui justifie son importance dans le projet. En revanche, des domaines de recherche s'intéressent à la carte à puce pour d'autres raisons. Il existe par exemple des études sur le développement commercial de la carte à puce [35] ou sur la sécurité du système.

2.1.3 Avantages et inconvénients

L'utilisation de la carte à puce apporte un gain opérationnel et améliore la qualité de service perçue par les usagers. Cependant, sa mise en place est un défi technique qui requiert un investissement majeur de la part des sociétés de transport [4]. L'hésitation des décideurs est donc compréhensible et c'est pourquoi il est essentiel de connaître les apports positifs et négatifs liés à une telle technologie. Dans cet objectif, le tableau 2.2 propose une liste des principaux avantages et inconvénients liés à l'utilisation de la carte à puce.

Tableau 2.2 Avantages et inconvénients de l'utilisation de la carte à puce

Avantages de la carte à puce	Inconvénients de la carte à puce
Offre une meilleure intégration des systèmes [6]	Les coûts de développement, d'installation et de maintenance sont élevés [8, 36]
Est plus adaptée pour le transport régional [6] grâce à l'interopérabilité [6, 8]	Le prix des infrastructures de perception est important [36]
Permet de proposer des programmes de fidélisation [6] par la mise en place d'une structure tarifaire plus complexe [8]	Le succès économique peut être long à venir et dépend de l'adhésion des usagers à la technologie [36]
Est économique sur le long terme [9] grâce en partie à la longue durée de vie des cartes [4]	Plus la complexité des systèmes augmente, plus la fiabilité diminue [6]
Évite la manipulation d'argent liquide [6] ce qui simplifie le travail des chauffeurs [8]	
Simplifie et accélère les transactions, ce qui améliore la vitesse d'opération et la flexibilité des unités [6]	
Est simple d'utilisation pour les voyageurs [9] ce qui améliore la qualité perçue [4] et diminue le taux de fraude [6, 8]	

Les auteurs rapportent comme principaux bénéfices une réduction des coûts à long terme, une meilleure gestion des revenus ainsi qu'une bonne fiabilité de fonctionnement. À l'inverse, ils reprochent au système sa complexité technique et ses coûts d'implantation élevés [4].

De surcroît, la mise en place de ces systèmes est justifiée par le besoin de récolter des données. Il faut donc mesurer les forces et les faiblesses liées à cet objectif. Le tableau 2.3 recense les principaux avantages et inconvénients de la carte à puce comme source de données.

Tableau 2.3 Avantages et inconvénients des données de cartes à puce

Avantages des données de cartes à puce	Inconvénients des données de cartes à puce
Il s'agit d'une grande quantité de données sur les déplacements [7, 8]	La précision spatiale est limitée [7], car le lieu de descente est souvent inconnu [8]
Les données sont localisées dans le temps et dans l'espace [8] et couvrent une grande proportion d'usagers [7, 8]	La distance des déplacements, l'utilisation de correspondances et les allers-retours sont difficilement mesurables [9]
Elles fournissent des informations spatio-temporelles et longitudinales [8] sur une longue durée [7, 8]	Il n'y a pas d'information sur le profil des usagers [7, 8], sur leur satisfaction ainsi que sur les non-usagers [7]
Les données sont sécurisées [6] et garantissent la confidentialité des usagers [8]	Les données sociodémographiques sont bloquées en raison de la confidentialité [8]
Les informations sont de bonne qualité [6, 9] et peu coûteuses [8]	Le motif des déplacements est inconnu [7, 9]
Elles remplacent ou enrichissent d'autres sources de données [7, 8]	La continuité des données est limitée dans le temps [8]
Elles permettent d'examiner le comportement des voyageurs [7]	Il peut y avoir présence d'erreurs systématiques et aléatoires [8]
Elles offrent la possibilité de lier les transactions avec des données personnelles pour recréer des chaînes de déplacements [7]	Les données sont non universelles et nécessitent de développer des modèles adaptés pour chaque objectif [8]
Elles apportent des données passives sur l'achalandage [8]	La diffusion des résultats est une étape difficile [8]

En conclusion, la carte à puce offre un potentiel énorme pour réaliser de nouvelles recherches. Les données qu'elle apporte permettent d'obtenir de nouveaux résultats donnant aux planificateurs des pistes d'amélioration pour le transport en commun. Ainsi, même si traiter ces données nécessite le développement de nouvelles méthodes de travail, utiliser les cartes à puce devient de plus en plus inévitable.

2.2 La correspondance en transport en commun

Les usagers du transport en commun ont souvent besoin de faire une correspondance au cours de leurs déplacements. Cependant, il a été démontré de nombreuses fois qu'elle est un frein à l'utilisation du transport en commun [37]. Elle représente une contrainte pour les usagers puisqu'elle les expose à la fois à une consommation de temps, à un effort physique et à un risque supplémentaire de manquer une alternative de transport. Pour une agence de transport urbain, réduire la pénibilité d'une correspondance fait ainsi partie des actions essentielles liées à la planification. Nombreuses sont les recherches qui ont été menées dans ce sens, visant à lister et évaluer l'influence des attributs considérés par les voyageurs, afin de comprendre comment diminuer l'effet des correspondances.

Cette partie commence par une définition générale de la correspondance et liste ensuite les attributs jugés influents sur le ressenti des usagers. Les notions de pénalité de correspondance et de perception du temps sont exposées et plusieurs exemples de solutions mises en place pour diminuer l'impact des correspondances sont énoncés.

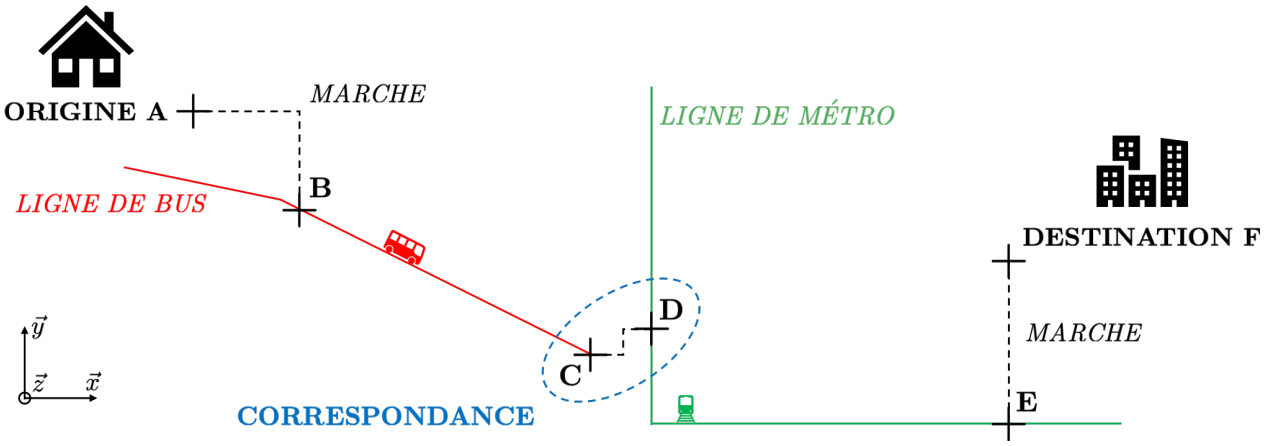
2.2.1 Définition d'une correspondance

En matière de mobilité, un déplacement est l'action de se rendre d'un point de départ à un point d'arrivée [38]. Dans certains cas, une seule ligne de transport est suffisante pour lier l'origine et la destination (OD). Dans toutes les autres situations, plusieurs étapes sont nécessaires. Il y a donc une correspondance (aussi appelée "transfert" par certains auteurs). La littérature permet de proposer la définition 2.1 pour expliquer une correspondance.

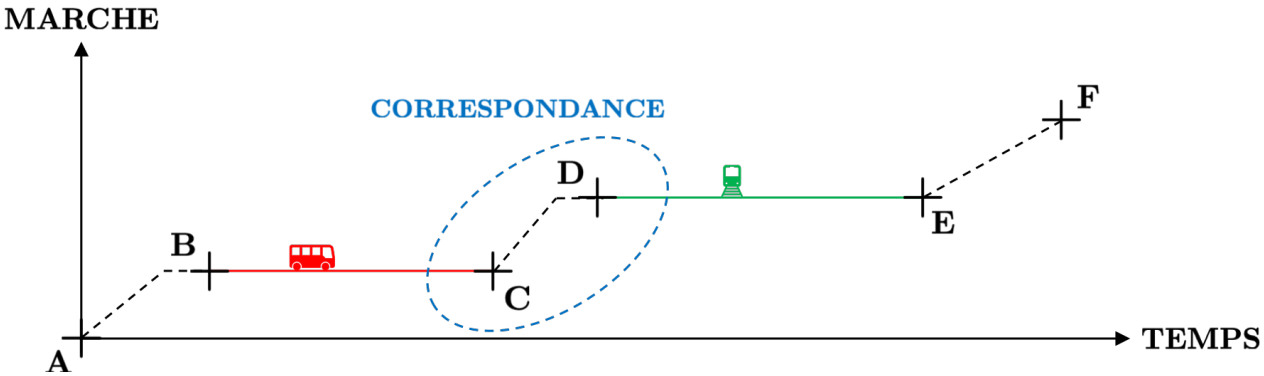
Définition 2.1. La correspondance
Une correspondance est l'action faite par un usager de changer de ligne de transport, en passant d'un véhicule à un autre. Ce changement peut être intramodal (entre deux bus par exemple) ou intermodal (entre un bus et un métro). L'intégration des correspondances dans un réseau comprend le design fonctionnel des lignes et des arrêts, la gestion des horaires, la diffusion de l'information et l'adaptation de la tarification [2].

Ainsi, un déplacement peut être constitué d'une ou plusieurs correspondances. La figure 2.1 illustre le déplacement réalisé par un voyageur pour rejoindre la destination F à partir de l'origine A. Le déplacement commence par une période de marche où le voyageur rejoint la première ligne de transport dans laquelle il embarque au point B. Arrivé au point C, il débarque et marche vers le point D pour embarquer sur une nouvelle ligne, c'est une

correspondance. Une fois au point E, il descend et marche de l'arrêt jusqu'à sa destination, le point F. Dans cet exemple, le voyageur fait un déplacement avec une correspondance intermodale, puisqu'il prend d'abord une ligne de bus et change pour une ligne de métro.



(a) Liste des étapes du déplacement positionnées dans l'espace



(b) Positionnement des étapes en fonction du temps et de la marche

Figure 2.1 Exemple de déplacement avec une correspondance

Comme le montre la figure 2.1, le déplacement est un ensemble d'étapes qui sont positionnées dans le temps et dans l'espace. La correspondance est ainsi un point de rupture spatio-temporelle du déplacement [39]. Elle nécessite à la fois de la marche et de l'attente, ce qui conduit à une diminution de l'attractivité du transport en commun au profit d'autres modes.

Typologies des correspondances

Modes Dans un premier temps, une correspondance peut être définie en fonction du ou des modes impliqués. Les possibilités de correspondre vont dépendre des alternatives disponibles dans le réseau étudié. Les études portent majoritairement sur des correspondances entre des modes lourds (bus, métro, tramway et train). Encore peu de recherches ont été faites en

considérant des correspondances entre des modes lourds et l'automobile ou des modes actifs (marche ou vélo) [40]. Une raison expliquant cette lacune est le manque de lien entre les différents acteurs, ce qui rend complexe l'obtention de données sur ces transferts.

Intervalles La fréquence des lignes joue un rôle important dans la réalisation d'une correspondance. Ainsi, Vuchic (2005) [2] propose de caractériser les correspondances suivant quatre catégories en fonction de la fréquence des deux lignes :

A. intervalles court – court *court : $\leq$ 10 min / long : $>$ 10 min*

B. intervalles long – court

C. intervalles court – long

D. intervalles long – long

   1. intervalles égaux, synchronisés
   2. intervalles égaux, non synchronisés
   3. intervalles différents, non synchronisés (temps d'attente aléatoire)

Dans les cas A et B, la correspondance est commode puisqu'elle implique peu d'attente. Les voyageurs n'ont pas besoin de planifier leur transfert, car ils pourront embarquer rapidement dans un véhicule. Le cas C est l'inverse du cas B, mais le temps d'attente est plus variable, ce qui augmente son impact. Le retard causé par la correspondance est incertain et il est essentiel de fournir aux passagers des informations pour planifier leur changement. Le cas D se décompose en trois situations. La première consiste en deux lignes ayant des intervalles égaux et synchronisés. Les véhicules des deux lignes arrivent au même moment en station et restent à l'arrêt suffisamment longtemps pour permettre aux usagers de changer. Le temps d'attente est court et la correspondance est alors commode. La deuxième situation consiste en deux lignes avec des intervalles égaux, mais non synchronisés. Le temps d'attente est fixe, mais diffère selon le sens de la correspondance (A vers B ou B vers A). La commodité de cette situation dépend ainsi du sens du transfert. Enfin, dans la troisième situation, il n'y a pas de coordination entre les deux lignes. Le retard causé par la correspondance est purement aléatoire, ce qui est très incommode pour les voyageurs.

Politique des sociétés de transport

D'un point de vue opérationnel, la définition 2.1 manque de finesse puisqu'elle ne fait pas la différence entre un nouveau déplacement et une correspondance. Ainsi, chaque société de transport doit préciser les conditions à respecter pour que l'embarquement dans un véhicule soit considéré comme une correspondance. Cette distinction est essentielle pour avoir un

système de tarification intégré. Si les systèmes de perception ne considèrent pas la correspondance, un nouveau passage est débité à chaque validation, ce qui n'est pas souhaité par les usagers. La définition 2.2 précise ce qu'est une correspondance pour une société de transport.

Définition 2.2. La correspondance opérationnelle
Une validation sur une borne de perception est vue comme une correspondance si :

- l'intervalle de temps entre cette validation et la précédente est inférieur à la limite fixée par la société de transport;
- la ligne associée à cette validation est soit différente de la précédente, soit dans la direction opposée;
- au moins une des deux validations n'est pas en métro.

Il faut noter que la définition 2.2 est générale, mais n'est pas adoptée par toutes les sociétés de transport. Certaines villes n'ont pas de tarification intégrée pour les correspondances et à l'inverse, d'autres ajoutent des critères plus stricts. Habituellement, l'intervalle de temps autorisé entre deux validations varie entre 60 et 120 minutes.

2.2.2 Attributs influençant la correspondance

Comme la correspondance diminue l'attractivité du transport en commun en faveur d'autres modes, de nombreuses études cherchent à évaluer son impact. L'objectif est de cerner les faiblesses qu'elle cause sur le réseau pour proposer de nouvelles solutions. Il s'agit d'expliquer comment les usagers raisonnent pour choisir de faire ou non une correspondance. Les effets des correspondances sont associés à plusieurs facteurs que les chercheurs veulent étudier. Le tableau 2.4 présente les principaux éléments cités dans la littérature en soulignant leurs impacts sur les voyageurs.

Tableau 2.4 Attributs diminuant l'attractivité des correspondances

Attribut	Auteurs	Remarques
Information	[27,37,41–46]	Les usagers ont besoin d'avoir de l'information en station pour correctement faire leur correspondance. Le manque d'information diminue la facilité de transférer et augmente le stress généré.
Attente	[37, 41–44, 46, 47]	Le temps d'attente a un impact non négligeable sur le choix d'itinéraire. Un usager ne veut pas attendre plus de dix minutes lors de sa correspondance [37] et le temps idéal est évalué à quatre minutes [43].
Marche	[37, 41–44, 46, 47]	La marche de la correspondance demande un effort physique et mental que les usagers n'apprécient pas. Le temps de marche maximum recommandé pour correspondre est de dix minutes [37].
Sécurité	[37, 41, 42, 44–46]	Il est essentiel que les passagers soient et se sentent en sécurité lors du transfert. Sinon, ils ne voudront pas faire la correspondance.
Équipement	[37, 41–43]	Une station doit être une zone de vie. L'équipement présent permet aux usagers de gagner du temps et de se sentir mieux. L'amélioration du temps "hors véhicule" est essentielle pour améliorer la perception globale du déplacement [41].
Fiabilité	[25, 37, 41, 43]	Plus le risque de ne pas réussir la correspondance est élevé, moins celle-ci est attirante. Les usagers vont avoir tendance à choisir l'itinéraire le moins risqué [48].
Personne	[26, 27, 44, 47]	L'âge influence la compréhension des informations et le genre impacte la perception de sécurité.
Achalandage	[25–27, 47]	Plus il y a du monde dans la station, plus la correspondance est difficile.
Heure	[23, 25–27]	En fonction de l'heure de la journée, les usagers n'ont pas le même regard sur la correspondance.
Nombre	[25, 37, 47]	Chaque ajout d'une correspondance à faire dans un déplacement diminue l'attractivité générale de celui-ci.
Temps	[25, 27, 37]	L'impact d'une correspondance semble diminuer avec l'augmentation de la longueur du déplacement.
Fréquence	[41, 43, 45]	Plus la fréquence des lignes est élevée, moins il y a d'attente. Par conséquent, la fréquence des lignes influence le choix d'itinéraire des usagers.
Habitude	[41, 45, 47]	La connaissance du réseau et de l'itinéraire facilite la correspondance.
Incertitude	[27, 42, 43]	Le risque de ne pas avoir une correspondance est un stress pour les usagers.
Confort	[37, 41]	Plus la station est confortable, moins la correspondance a d'impact.
Dépendance	[25, 41]	Certains usagers n'ont pas plusieurs choix d'itinéraires. Ils sont obligés de faire une correspondance.
Tarification	[37, 46]	Une correspondance ne doit pas engendrer de frais supplémentaires.

En résumé, beaucoup d'attributs sont cités dans la littérature. Comme ils influencent le choix de faire une correspondance, il est important pour les planificateurs du transport en commun d'en considérer le plus possible dans leurs modèles. Cependant, tous les attributs n'ont pas le même impact. Les auteurs s'accordent tous pour dire que l'information disponible, le temps de marche et d'attente ainsi que la sécurité sont les éléments principaux qui rendent les correspondances peu attrayantes pour les voyageurs. Il faut noter que peu de ces recherches intègrent des données de cartes à puce [26, 27, 34]. Les études des attributs sont majoritairement réalisées à l'aide de données d'enquêtes auprès des utilisateurs du TC. Ceci s'explique par le fait que les données de cartes à puce caractérisent l'utilisation du réseau, mais ne contiennent pas d'information sur le ressenti des usagers.

2.2.3 Pénalité de correspondance

La partie précédente liste les attributs qui rendent la correspondance pénible pour les usagers, mais l'impact de chacun d'entre eux n'est pas quantifié. Certaines études en transport utilisent des algorithmes de choix modal pour modéliser les déplacements. Le but est de prédire quel mode est choisi par un usager pour aller d'une origine à une destination. Cela permet ensuite d'affecter chaque paire OD à un itinéraire afin de modéliser les retombées d'une stratégie. C'est une des quatre étapes clés de la procédure séquentielle classique [49]. Cependant, pour gagner en précision, il faut que les correspondances soient prises en compte. C'est le rôle de la pénalité de correspondance (également nommée "lourdeur", "pénibilité" ou "coût") qui est expliquée par la définition 2.3. Elle permet d'enrichir le calcul de la fonction d'utilité d'un itinéraire. Cette fonction est utilisée pour quantifier l'attirance d'un choix par rapport à un autre. Pour cela, elle somme un ensemble de facteurs en leur donnant une valeur financière (ou temporelle) à l'aide d'un coefficient. Plus le coût généralisé d'un itinéraire augmente, plus la probabilité que l'usager fasse ce choix diminue. Le coût généralisé d'un itinéraire est inversement proportionnel à sa fonction d'utilité [49].

Définition 2.3. Pénalité de correspondance
La pénalité de correspondance est un facteur additionnel de la fonction de coût généralisé d'un déplacement qui traduit la lourdeur causée par une correspondance sur l'itinéraire d'un usager. Autrement dit, la pénalité de correspondance exprime l'impact d'une correspondance sur l'ensemble d'un déplacement. Elle regroupe dans un même terme certains attributs propres à la correspondance qui sont connus pour influencer le choix des voyageurs.
La pénalité de correspondance prend la même unité que le reste du coût généralisé. Selon les algorithmes, elle s'exprime en temps (minutes) ou en valeur monétaire ($).

Ainsi, plusieurs recherches visent à estimer la pénalité de correspondance. Premièrement, Iseki et Taylor (2009) [41] centralisent les attributs en trois catégories, opérationnel, physique et personnel, et proposent de donner un coût à chaque élément. Ils réalisent une régression linéaire sur des données d'enquête pour savoir quel coefficient donner à chaque attribut. Leurs résultats permettent de quantifier l'impact de chaque attribut et de comprendre comment minimiser la pénalité de correspondance. Une année plus tard, ils font une enquête Importance – Satisfaction et montrent que l'élément à améliorer en priorité est l'aspect opérationnel des stations [50].

Ensuite, Guo et Wilson (2011) [42] s'intéressent à estimer le coût total perdu dans les correspondances. Leur objectif est de déterminer où, comment et combien investir pour améliorer le service offert. Ils montrent que le coût des correspondances provient de trois sources principales : la marche, l'attente et l'environnement. Cependant, ces trois sources sont indépendantes et nécessitent des solutions différentes.

Enfin, Ceder et Chowdhury (2012) [37] expriment la pénalité de correspondance en fonction de deux catégories : la fiabilité du changement et l'anxiété des voyageurs. Avec une enquête, ils concluent que le temps avant qu'un passager commence à stresser est de dix minutes. Ils expliquent que la pénalité de correspondance diminue uniquement si le service est fiable et que les usagers se sentent en sécurité.

Les chercheurs portent beaucoup d'intérêts sur la pénalité de correspondance puisque comprendre comment la réduire permet d'améliorer le service offert. Le principe de la pénalité de correspondance est fortement lié avec celui de la perception du temps, présenté dans la partie suivante.

2.2.4 Perception du temps

La notion de durée dépend de la situation vécue. Parfois, les minutes semblent s'étirer alors qu'à d'autres moments elles semblent passer rapidement. Cet écart entre la durée réelle d'un événement et la durée ressentie par l'usager s'explique par le fait que l'humain n'a pas une perception constante du temps qui passe. En transport, cette perception joue un rôle important puisque les voyageurs cherchent souvent à minimiser la durée de leur déplacement. Dans leurs modèles, les planificateurs doivent ainsi considérer le temps perçu plutôt que le temps réel. La notion de perception du temps (définition 2.4) est créée dans cet objectif.

Définition 2.4. Perception du temps
En transport, la perception du temps est une notion utilisée pour mesurer le temps tel qu'il est ressenti par un usager. La perception du temps est un facteur donnant l'équivalence entre la durée exacte d'un événement et sa durée subjective. Son unité est ainsi le rapport entre le temps perçu et le temps de référence (généralement en min/min).
La perception du temps peut également servir pour transformer une valeur temporelle (une durée) en une autre unité chiffrable (un coût par exemple). Dans ce cas, elle est nommée valeur du temps perçue et a une unité de conversion ($/min, etc.).

Plusieurs études veulent mesurer et expliquer la perception du temps lors d'une correspondance puisqu'elle impacte le choix d'itinéraire des usagers. Effectivement, des chercheurs ont montré que faire une correspondance change la valeur du temps perçu d'un déplacement.

Parmi les premiers, Wardman (2004) [51] propose de fusionner les données de beaucoup d'enquêtes du Royaume-Uni pour comparer la valeur du temps en transport en commun face à la voiture. Il utilise pour cela une unité nommée IVT (*In Vehicle Time*) qui exprime le ratio entre le temps ressenti en transport en commun avec le temps réel en voiture. Il montre ainsi que la valeur du temps varie en fonction de l'usager et du mode et que le bus (1,79 IVT) est plus pénalisant que le train (1,53 IVT). De même, la marche semble moins pénalisante que l'attente (1,68 IVT contre 1,76 IVT). Ces résultats confirment également que la valeur du temps diminue avec l'augmentation des distances parcourues en transport en commun.

Ensuite, Hadas et Ranjitkar (2012) [52] utilisent des données GTFS et une enquête sur les déplacements afin de séparer les attributs temporels et les attributs propres aux correspondances. Ils calculent d'un côté la valeur du temps et de l'autre la qualité des transferts sur plusieurs nœuds du réseau. Cela permet d'identifier les zones ayant une mauvaise connectivité et de les trier pour savoir lesquels retravailler en premier. Ils montrent qu'il existe une corrélation entre les deux attributs.

Chowdhury et al. (2015) [53] poursuivent par une étude de sensibilité. Ils souhaitent déterminer quelle est la variation de coût ou de temps qui encourage à passer d'un itinéraire direct à un trajet avec une correspondance. À l'aide de données d'enquête, ils montrent qu'en moyenne un usager est prêt à faire une correspondance si elle permet de diminuer de 33 % le temps du déplacement ou de 16 % son coût. Les usagers sont alors plus sensibles au coût du déplacement qu'au temps. En gagnant en confort lors du déplacement, les usagers acceptent une correspondance si elle fait gagner en temps et en prix respectivement 25 % et 10 %.

Pour continuer, Jin et al. (2017) [47] calculent l'équivalence entre le temps perçu d'un trajet direct (T_d) et celui d'un trajet avec une correspondance (T_t). Cela leur permet de développer un modèle de choix d'itinéraire qui prend en compte l'impact des correspondances. Avec les données d'une enquête, ils arrivent à la conclusion que le temps perçu d'un trajet avec une correspondance est une fois et demi plus élevé que celui d'un trajet direct : $T_t = 1{,}5 \times T_d$. Autrement dit, pour un même temps de trajet, l'usager a la sensation que son déplacement dure une fois et demie plus longtemps lorsqu'il fait une correspondance. Ces résultats sont confirmés par Garcia-Martinez et al. (2018) [45] qui montrent également que la valeur du temps diffère selon le sexe, les habitudes et le type de changement.

En conclusion, connaître l'impact des correspondances sur la valeur du temps permet de mesurer la qualité des transferts du réseau pour déterminer ceux qui ont besoin d'une amélioration. De plus, ces valeurs permettent d'enrichir les algorithmes de calcul de chemin qui pourront ainsi conseiller aux usagers leur itinéraire préféré au lieu du plus rapide. En revanche, les données de cartes à puce ne sont pas utilisées pour ces études.

2.2.5 Solutions mises en place pour limiter l'impact des correspondances

Plusieurs stratégies sont déjà mises en place dans des réseaux de transport pour limiter la pénibilité des correspondances, en voici quelques exemples.

Information dans les stations Certaines sociétés fournissent de l'information dans les stations afin de limiter la marche et l'attente. Par exemple, des applications GPS telles que `Citymapper` (`https://citymapper.com`) intègrent un indicateur de wagon conseillant à l'usager une porte de la rame de métro. Cela permet au voyageur de gagner du temps lors de son changement puisqu'il descend de la rame au bon endroit. L'information diminue aussi le stress et le risque de manquer la correspondance. À Lyon, dans le réseau des Transports en commun lyonnais (TCL), des panneaux d'affichage indiquent dès la sortie d'une ligne le temps d'attente pour toutes les directions des autres lignes. Ainsi, les passagers savent s'ils doivent se dépêcher ou non pour faire leur correspondance.

Structure des stations Vuchic (2005) [2] propose des configurations de stations pensées pour faciliter les correspondances. Plusieurs d'entre elles sont schématisées à la figure A.1 (annexe A). Des applications réelles existent. Par exemple, dans la station Lionel-Groulx à Montréal, les lignes orange et verte s'arrêtent au même étage et suffisamment longtemps pour permettre le passage d'une ligne à l'autre. Pour correspondre, la majorité des usagers doivent juste traverser le quai, ce qui prend quelques secondes.

Horaires cadencés Il est possible de synchroniser les lignes les plus utilisées afin de diminuer le temps d'attente lors de la correspondance. Sur des arrêts définis (les points focaux), les unités arrivent et repartent au même moment [2]. Elles s'arrêtent pendant une durée programmée suffisamment grande pour rejoindre l'autre ligne. Ainsi, les usagers qui passent d'une ligne à l'autre n'ont pas besoin d'attendre. Ce concept diminue grandement la lourdeur des correspondances, mais est complexe dans sa mise en place. Il est illustré sur quatre points focaux à la figure A.2.

Configuration du réseau Le tracé des lignes a un impact sur la qualité des correspondances autant que sur leur nombre. Par exemple, à Chicago, toutes les lignes du métro ont un tronçon commun dans le centre-ville. C'est le *Loop* (figure A.3.a). Ce système permet aux usagers de rejoindre le centre-ville depuis toutes les lignes, sans avoir besoin de faire une correspondance. En addition, les changements d'une ligne à l'autre sont plus efficaces et avec moins d'attente. De même, plusieurs réseaux se dotent de lignes circonférentielles puisqu'elles favorisent à la fois la distribution de la charge sur les lignes et diminuent le nombre de correspondances. C'est le cas de Londres avec la *Circle Line* (figure A.3.b).

2.3 Éléments méthodologiques

Les sections précédentes de la revue de littérature apportent des éléments théoriques concernant à la fois la technologie des cartes à puce et l'usage des correspondances. Cette section s'attarde aux éléments méthodologiques permettant de répondre aux quatre objectifs de cette recherche.

2.3.1 Comportement des usagers

Nombreuses sont les recherches qui s'intéressent au comportement des usagers du transport en commun. Il s'agit d'observer et d'expliquer les agissements des voyageurs. Avec le développement des méthodes d'exploration de données (*data mining*), de nouvelles possibilités ont permis de remplacer les enquêtes par des données opérationnelles afin de déduire des comportements [11]. La segmentation semble être la méthode la plus répandue. En se basant sur un vecteur descriptif (plusieurs caractéristiques déduites des données), les passagers dont le comportement est similaire sont regroupés et ceux dont il est hétérogène sont séparés.

Le principal objectif d'une segmentation est d'obtenir des groupes de comportements typiques qui s'expliquent en fonction de divers attributs. Par exemple, Deschaintres et al. (2019) [13]

s'intéressent à définir la variabilité interpersonnelle et intrapersonnelle du comportement des usagers en se basant sur l'organisation des déplacements dans la semaine. Ils utilisent une segmentation en 12 groupes par la méthode des k-moyennes, ce qui leur permet de définir une typologie de semaines. Ils identifient aussi différentes périodes d'activités en regroupant toutes les semaines d'une année. En parallèle, Viallard et al. (2019) [33] tentent d'identifier l'évolution de la structure des déplacements afin de prédire les comportements. Pour cela, ils utilisent la méthode des k-moyennes sur un an, mais réalisent une segmentation chaque semaine. Cette méthode permet de suivre à la fois l'évolution des usagers dans les groupes et l'évolution des groupes.

Certaines recherches modélisent les comportements des usagers sur une très longue période. Le but est de mesurer la variabilité des comportements, car « un réseau avec peu de changement d'activité des passagers va nécessiter moins d'attention en termes d'opération » [12].

En résumé, étudier le comportement des usagers est essentiel pour comprendre les besoins des utilisateurs et pour améliorer les services offerts. Par exemple, identifier différents groupes de comportements est un moyen d'enrichir la grille tarifaire proposée.

2.3.2 Performance des réseaux

Le coût d'exploitation des lignes de transport constitue une part importante du budget des sociétés de transport qui veulent le minimiser. Dans cet objectif, des chercheurs développent des méthodes pour caractériser les lignes. Le but est de déterminer les liens qui ont besoin d'une attention particulière et, si possible, de fournir des idées d'amélioration.

Parmi les premiers, Mishra et al. (2012) [54] mesurent la connectivité des éléments dans un réseau multimodal avec l'approche théorique des graphes. Pour cela, ils décomposent le réseau en quatre éléments : les nœuds, les lignes, les centres de correspondances et les régions. Ils arrivent ainsi à identifier chaque élément qui diminue la connectivité générale du réseau. Plus tard, Hadas (2013) [55] reprend cette méthode en utilisant des données GTFS. Il propose cinq indicateurs qu'il utilise pour noter quatre attributs : arcs et nœuds du réseau et arcs et nœuds des itinéraires. Cette méthodologie permet très rapidement de lister les lieux problématiques du réseau en sachant quel attribut est en cause.

Ces méthodes ne prennent pas en considération le volume de passagers sur les lignes. Ainsi, Ceder et Perera (2014) [56] proposent une méthode basée sur un algorithme de maximisation des flux afin de mesurer la capacité de chaque lien. Ils détectent alors les goulots d'étranglement du réseau, c'est-à-dire les zones qui affaiblissent le potentiel des lignes.

Plus tard, Dubos-Golain et al. (2017) [16] étudient la variabilité d'utilisation des lignes de transport avec une segmentation par les k-moyennes. Ils séparent ainsi l'utilisation des lignes en fonction du jour de l'année et de la période de la journée. Ils montrent que certaines lignes sont surtout employées par un type d'usagers particulier, que des lignes servent différemment en fonction de la journée et que la météo a un impact non négligeable sur le réseau.

Pour terminer, Frappier et al. (2018) [57] présentent deux indicateurs qui permettent d'évaluer la qualité et la diversité des alternatives de transport pour un déplacement. Ils définissent un indicateur de similarité entre les lignes qui permet d'aider les usagers à choisir leur itinéraire.

En définitive, toutes ces méthodes permettent de prioriser des actions de développement sur des lignes en particulier, pour diminuer les coûts d'exploitation tout en répondant le mieux possible aux besoins des usagers.

2.4 Conclusion de la revue de littérature

En tout premier lieu, cette revue de littérature définit les notions essentielles de ce projet. Le fonctionnement des systèmes de cartes à puce est expliqué, ainsi que les fondements du déplacement et de la correspondance. La structure du déplacement basée sur une échelle spatio-temporelle est décrite ce qui sert pour énoncer les différents paramètres et acteurs de la correspondance. La revue de littérature expose le fort potentiel des données de cartes à puce dans un contexte de planification des transports. Les systèmes de perception par cartes à puce ont l'avantage de fournir une grande quantité de données spatio-temporelles réputées de bonne qualité. Ainsi, les cartes à puce ont permis aux chercheurs de mener des études à la fois sur le comportement des usagers et sur l'utilisation des réseaux, même si leurs traitements nécessitent le développement de nouveaux algorithmes complexes. La possibilité de travailler sur un réseau complet avec une grande couverture temporelle est l'argument principal justifiant l'utilisation des cartes à puce dans ce projet.

Dans un second temps, la revue de littérature explique pourquoi la correspondance contraint les usagers dans leurs déplacements en transport en commun. Beaucoup d'études visent à identifier les attributs qui impactent le choix d'itinéraire des usagers et les scientifiques s'accordent à dire que l'information disponible, les temps de marche et d'attente ainsi que la sécurité sont les éléments les plus pénalisants pour les voyageurs. Leurs impacts sont quantifiés dans plusieurs études à la fois en mesurant la pénalité de correspondance ou la valeur du temps perçu. Cependant, les données d'enquêtes sont limitées dans le temps et dans l'espace, et les études à partir des données de cartes à puce sont encore minoritaires,

car elles nécessitent des algorithmes de repérage des correspondances. Conséquemment, il apparaît que les études spatio-temporelles de l'utilisation des correspondances représentent, encore aujourd'hui, un domaine à approfondir.

Ce projet vise alors à apporter des connaissances sur le comportement des voyageurs et leur utilisation du réseau, avec comme élément central de réflexion la correspondance. Pour permettre de mener à bien cette étude, la revue de littérature présente également plusieurs recherches sur la caractérisation du comportement des usagers et sur la mesure de la performance des réseaux. Elle influence ainsi le choix de la méthode employée dans ce projet, la segmentation des usagers par la méthode des k-moyennes.

CHAPITRE 3 DONNÉES ET MÉTHODOLOGIE

Le chapitre 2 a montré que les cartes à puce fournissent des données exploitées dans diverses études du transport en commun. Cependant, des données brutes sont générées à chaque transaction. C'est pourquoi il faut appliquer des transformations pour obtenir, avec la plus grande précision possible, des informations sur les correspondances. Ce chapitre détaille la méthodologie proposée pour caractériser l'utilisation des correspondances dans le transport en commun. Elle peut être appliquée sur tous les réseaux de transport équipés d'un système de perception par cartes à puce, ce qui rend le projet généralisable à beaucoup d'autres villes.

Ce chapitre établit dans un premier temps une description des données utilisées dans l'étude puis détaille le fonctionnement de l'algorithme de transformation des transactions en déplacements. Ensuite, les éléments techniques utilisés sont exposés pour finalement s'intéresser aux méthodes d'analyses des déplacements et des usagers.

3.1 Structure de l'étude

L'ensemble du projet est structuré en plusieurs étapes qui s'enchaînent successivement, comme l'illustre le schéma 3.1. La première étape consiste à appliquer l'algorithme de transformation des validations en déplacements (VALIDEP) aux données originales (section 3.3). Le résultat de cet algorithme est la production de deux tables de données, respectivement celle des validations et celle des déplacements. Avant les analyses, une typologie des lignes de bus et une classification des déplacements sont développées (sections 3.4 et 3.5). Ces données sont d'abord utilisées pour obtenir la répartition des déplacements quotidiens (section 3.6), ce qui permet d'analyser l'évolution de l'utilisation des correspondances et de caractériser la symétrie des itinéraires pendulaires. En parallèle, les deux tables sont combinées pour créer le profil de déplacement de chaque carte. L'ensemble des profils est alors employé pour segmenter les usagers de l'échantillon et obtenir davantage d'informations sur les attributs qui influencent le choix de correspondre (section 3.7). Enfin, les jeux de données sont utilisés pour mesurer l'indice d'autonomie de chaque ligne et pour identifier les paires de lignes comptabilisant le plus de correspondances, de sorte à localiser les nœuds et les liens nécessitant un intérêt particulier (section 3.8).

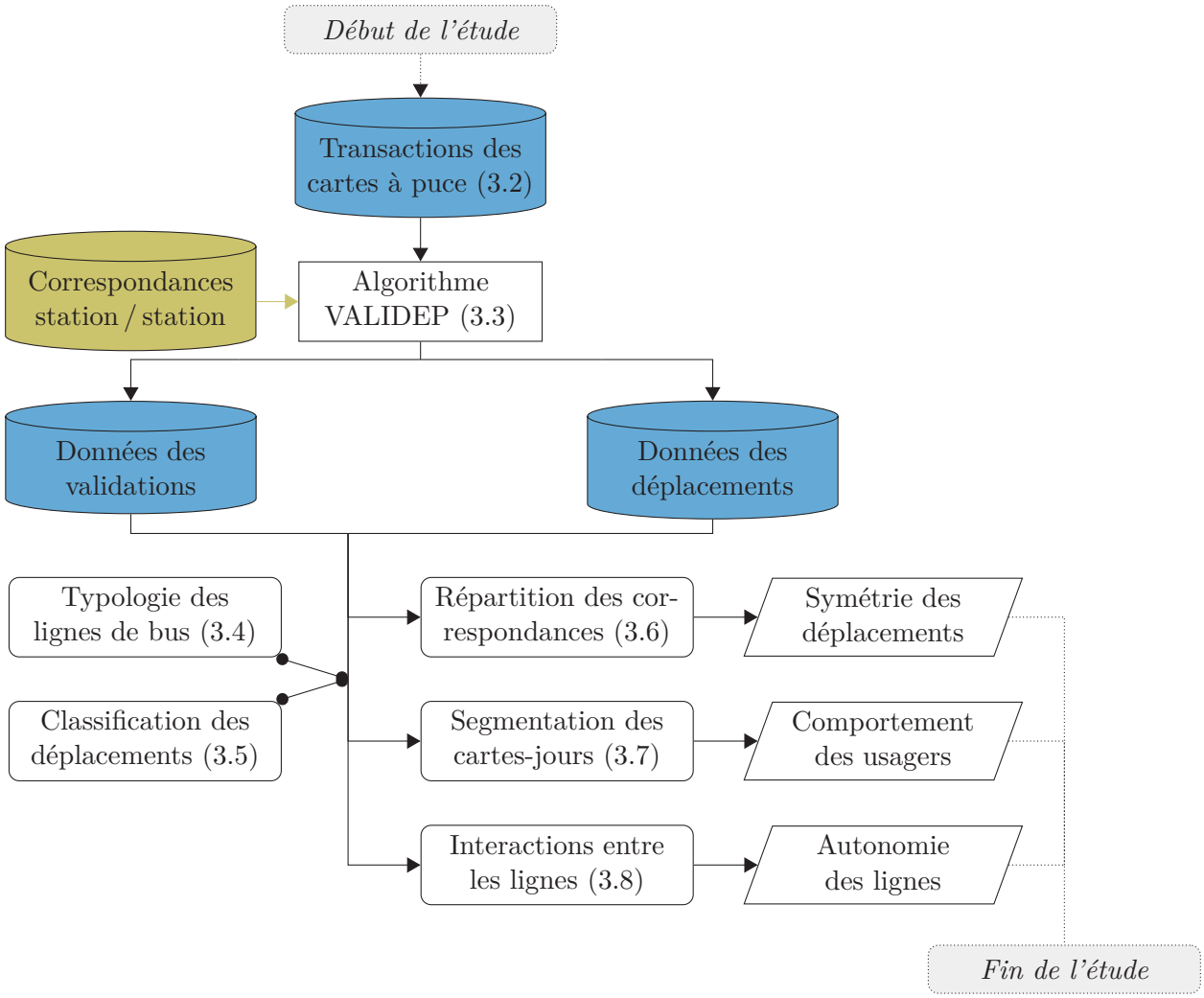


Figure 3.1 Méthodologie générale de l'étude

3.2 Description des données

Les données utilisées tout au long de ce projet proviennent d'un échantillon de transactions fourni par la Société de transport de Montréal (STM), à partir des cartes à puce de l'année 2016. La section actuelle permet donc de présenter le réseau de la STM ainsi que son système de perception. Les données sont alors détaillées, tout comme le processus d'échantillonnage.

3.2.1 Réseau de la STM

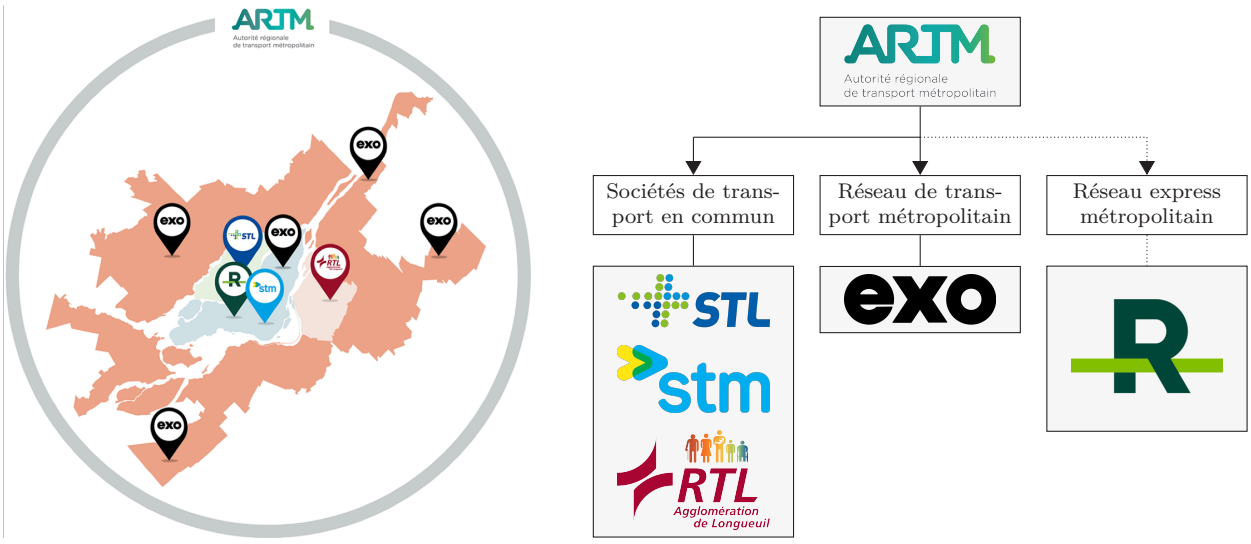
La Société de transport de Montréal

La Société de transport de Montréal (STM) est une entreprise publique de transport collectif. Créée en 1951 sous le nom de Commission de transport de Montréal (CTM), l'agence vise à regrouper l'ensemble des services de transport en commun présents sur l'Île de Montréal. Cette commission a changé de fonctionnement et d'appellation plusieurs fois au cours de son évolution pour prendre sa forme finale, la Société de transport de Montréal, en 2002. Aujourd'hui, la STM « met en place, exploite et maintient un réseau performant et intégré qu'elle contribue à développer » [58]. C'est la principale agence de transport qui opère sur l'Île de Montréal. En 2016, elle était la 15e entreprise en importance au Québec, avec une valeur de remplacement des actifs de 26 G$ [59]. Elle est passée 13e au classement et a été nommée 4e entreprise dans le secteur du transport en 2019 [60].

Par ailleurs, la grande région de Montréal s'étend au-delà de l'Île. Le territoire métropolitain est décomposé en 5 régions : l'Île de Montréal, Laval, Longueuil, la couronne nord et la couronne sud. Par conséquent, plusieurs agences de transport opèrent des réseaux différents. Dans l'objectif d'uniformiser l'offre et d'encourager les interactions entre les organismes, l'Autorité régionale de transport métropolitain (ARTM) a été créée. Son rôle est de « planifier, organiser, financer, promouvoir » le transport en commun de sorte à favoriser l'intégration des services entre les différents acteurs [58]. Ainsi, le transport collectif de la Communauté métropolitaine de Montréal (CMM) s'articule autour de quatre organismes publics de transport en commun (OPTC) supervisées par l'ARTM : la Société de transport de Montréal (STM), la Société de transport de Laval (STL), le Réseau de transport de Longueuil (RTL) et exo. Cette hiérarchisation est illustrée à la figure 3.2. À ces quatre OPTC s'ajoutera éventuellement le Réseau express métropolitain (REM), un réseau de trains légers, à partir de 2021.

Le réseau de transport de la STM

La STM opère un réseau de bus, un réseau de métros ainsi qu'un service de transport adapté. Une carte du réseau est présentée à l'annexe B. Le métro comporte 4 lignes, la verte, la orange, la jaune et la bleue, pour une longueur totale de 71 kilomètres et 68 stations. En 2016, son parc comptait 759 voitures qui ont totalisé 79,6 millions de kilomètres. Souhaitant remplacer petit à petit les vieillissantes rames MR-63 et MR-73, la STM a accueilli en 2018 ses premiers AZUR (54 trains) qui font monter le nombre de voitures à 909 et augmentent de 16 % le nombre total de kilomètres parcourus en 2019 [59, 60].



(a) Territoire du Grand Montréal [61]

(b) Hiérarchie du transport collectif

Figure 3.2 Organisation du transport en commun de la grande région de Montréal

Le réseau de bus couvre la majorité du territoire de Montréal (près de 500 km^2), avec un parc de 1748 bus dont 35 hybrides et 257 articulés. En 2016, il y avait 220 lignes, presque toutes accessibles aux personnes à mobilité réduite. De plus, environ 274 kilomètres de voies comportaient des mesures préférentielles. Au total, le service de bus a offert 85 millions de kilomètres en 2016 [59]. Ces chiffres ne cessent de croître grâce aux investissements de la STM pour améliorer le réseau.

La STM opère six types de lignes différents : le réseau local, le réseau 10 minutes max, le réseau de nuit, le réseau express, les navettes spéciales et les navettes OR (conçues pour les aînés). À ces lignes s'ajoute un service de taxis collectifs mis en place dans les quartiers où il n'y a pas de bus régulier.

Enfin, pour les personnes ayant des limitations fonctionnelles, la STM propose un service de transport adapté porte-à-porte. Il est rendu possible grâce à 86 minibus de la STM et 14 fournisseurs de services de taxis réguliers. Ce service a contribué à plus de 3,8 millions de déplacements en 2016 pour 30 000 voyageurs [59].

Le système OPUS

Le réseau de la STM est accessible à tout utilisateur grâce à un système automatisé de perception des titres (AFC, de l'anglais *automated fare collection*). Lorsque l'usager veut

embarquer dans un bus ou entrer dans une station de métro, il valide une carte sur une des bornes de perception présentes. Par une communication RFID (de l'anglais *radio frequency identification*), la borne peut ainsi lire les informations de la carte et autoriser ou non l'entrée dans le réseau selon le titre de transport disponible.

Depuis 2008, le système utilisé par la STM est le système OPUS. Il regroupe trois types de cartes différentes. La première est la carte OPUS commune de transport (CPCT). C'est la carte à puce traditionnelle qui a donné son nom au système. Les deux autres sont la carte à puce occasionnelle (CPO) et la carte magnétique jetable (CMJ). La carte OPUS peut être utilisée dans tous les réseaux de transport de la communauté OPUS (exo, MRCJT, RTC, RTL, STL, ST LÉVIS et STM), ce qui inclut également des réseaux à l'extérieur du territoire de la région de Montréal : le Réseau de transport de la Capitale (RTC) et Société de transport de Lévis (ST LÉVIS). Elle est rechargeable et sa fabrication en plastique lui confère une longue durée de vie. La carte OPUS est utilisée par une très grande majorité des voyageurs réguliers. À l'inverse, les CPO et CMJ sont respectivement en carton et en papier. La CPO est utilisable dans la majorité des réseaux tandis que la CMJ est réservée à la STM, la STL et une partie du réseau exo (secteur Laurentides). Les deux cartes ne sont pas rechargeables et n'acceptent qu'un seul titre de transport à la fois. La différence entre ces deux cartes réside dans leur fonctionnement. La CPO a une carte à puce, ce qui lui permet de recevoir plus de types de produits tarifaires, alors que la CMJ a une bande magnétique. Le tableau 3.1 montre la différence d'utilisation entre les trois types de cartes. Les CPCT représentent seulement 6,6 % des cartes, mais 89,1 % des transactions. L'annexe C illustre les trois types de cartes et liste les titres de transport existant pour chacune d'entre elles.

Tableau 3.1 Utilisation du réseau par type de support en 2016

	CPCT	CPO	CMJ	Total
Nombre de cartes	1 984 040	9 276 571	19 000 394	30 261 005
Nombre de validations	429 137 757	25 825 024	26 787 303	481 750 084
Validations / carte	216,29	2,78	1,41	
Part des validations	89,1	5,4	5,6	

En plus de vérifier la validité des titres de transport, le système de perception de la carte OPUS permet de générer un grand nombre de données de type SCFCD (de l'anglais *smart card fare collection data*). Chaque enregistrement de transaction est sauvegardé dans la machine de perception puis une centralisation est faite. Cette centralisation est quotidienne pour

les équipements du métro et s'échelonnent sur quelques jours pour les boites de perception des bus, car ils doivent être sondés au garage. La composition de ces données est explicitée à la section 3.2.3.

3.2.2 Échantillon des données de cartes à puce

L'objectif du projet est de développer une méthodologie d'analyse de l'utilisation des correspondances par un groupe d'usagers du transport en commun. L'originalité de cette étude repose sur deux paramètres :

- une grande échelle temporelle : un an de données;
- une grande échelle spatiale : tout le réseau de bus et de métro de la STM.

Cependant, le réseau de transport de la STM enregistre environ 480 millions de transactions par an. Il n'est pas avantageux de travailler avec autant de données, car cela augmente considérablement le temps d'exécution des algorithmes et complique la gestion de l'information. Les données sont alors échantillonnées en suivant le processus schématisé à la figure 3.3. Il faut préciser que l'échantillon de données est utilisé afin de valider les méthodes, il n'est pas nécessairement représentatif de l'utilisation du réseau de la STM.

Premièrement, seules les données provenant des cartes OPUS sont sélectionnées. Les autres supports de titres ne fournissant pas suffisamment d'informations longitudinales ; ils ne permettent pas d'analyser le comportement des usagers. Ensuite, pour être assuré de modéliser le comportement d'une carte qui existe (elle pourrait être perdue ou expirée), seules les données provenant des cartes à puce qui réalisent au moins une validation par mois durant toute la période de l'étude sont gardées. Enfin, comme l'étude porte davantage sur les usagers réguliers du transport en commun, seules les données de cartes ayant un abonnement mensuel ou annuel sont sélectionnées. L'application de ces trois filtres permet le passage de plus de 30 millions à environ 2 millions de cartes. Finalement, un tirage aléatoire de 10 000 cartes est réalisé à l'intérieur de cet ensemble. Pour respecter les deux paramètres énoncés précédemment, aucun filtre sur la position ou la période d'utilisation des cartes n'est appliqué.

Sur les 366 jours de la période d'étude (du 1er janvier au 31 décembre 2016), les 10 000 cartes OPUS sélectionnées constituent 9 550 725 transactions. L'échantillon en question représente ainsi 0,50 % de toutes les CPCT et 2,22 % des transactions de l'année 2016.

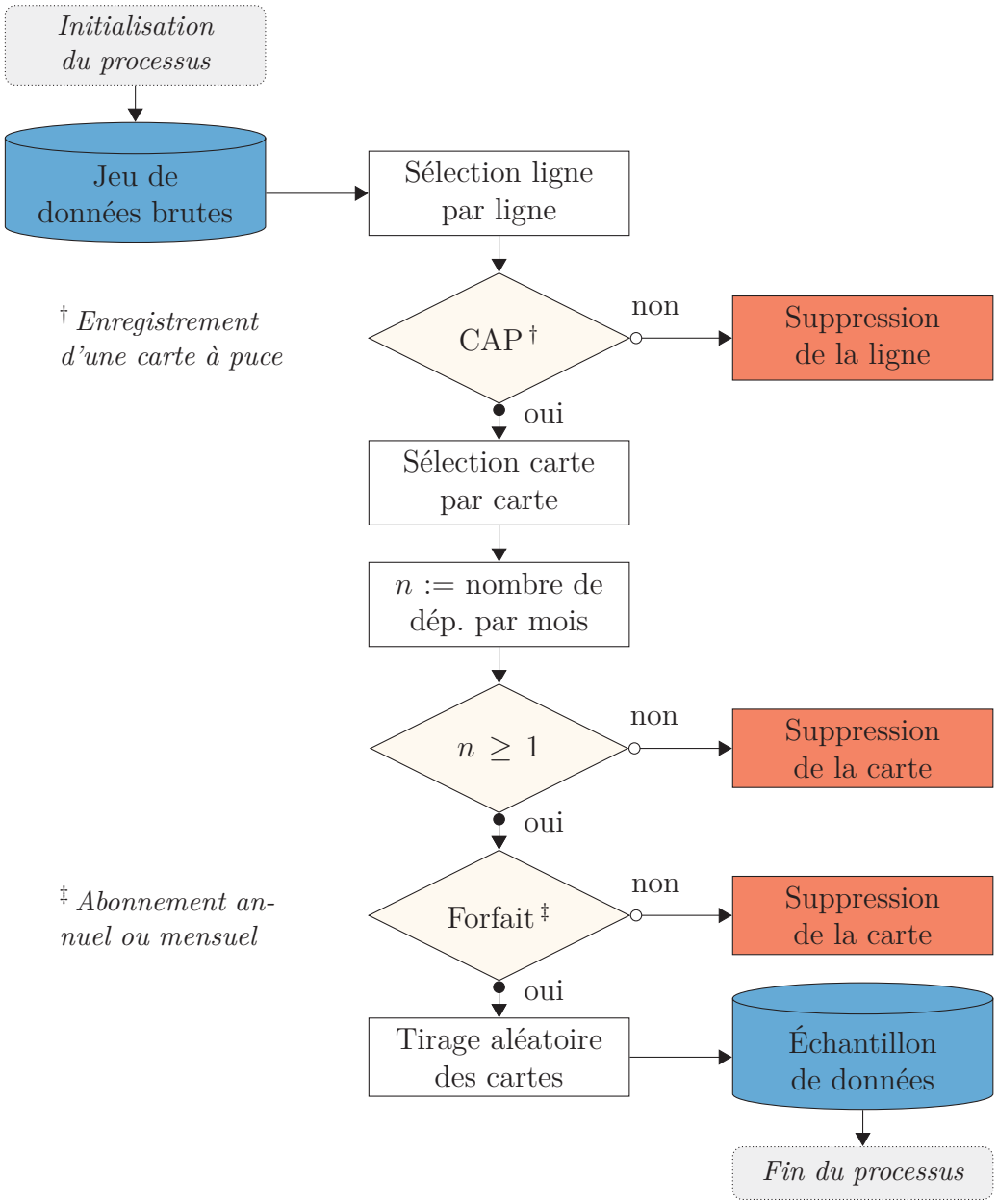


Figure 3.3 Processus de sélection de l'échantillon de données

3.2.3 Informations disponibles dans la table des données brutes

Sur le réseau de la STM, les voyageurs n'ont besoin de valider leur carte à puce que lors de la montée dans un bus, ou lors de l'entrée dans une station de métro. Il n'y a pas de trace de leur sortie d'un véhicule. Ainsi, les informations contenues dans un enregistrement de transaction sont un mélange des informations constituant la carte à puce et des informations fournies par la borne de perception. Le tableau 3.2 recense l'ensemble des champs d'information contenus dans les 9 550 725 données brutes.

Tableau 3.2 Informations recueillies lors de chaque transaction

Attribut	Format	Description	Exemple
id_validation	Entier	Identifiant unique de la transaction	284 545
date_heure	Balise temporelle	Date et heure de la transaction	2016-08-05 10:52:30
code_type_support	Entier	Type de carte	1 (OPUS)
code_produit	Caractère	Code du produit (titre de transport)	CAMOP
id_carte	Entier	Identifiant unique de la carte à puce	196
code_ceb	Facteur	Code d'emplacement billettique	B–03
code_ligne	Entier	Numéro de la ligne utilisée	165
code_direction	Entier	Direction de la ligne utilisée	5
act_cal	Caractère	Champ d'activité recalculée	D1

Tout d'abord, l'attribut "id_validation" est un numéro unique donné à chaque enregistrement d'une transaction afin de pouvoir l'identifier. De même, chaque carte à un numéro unique qui permet de la repérer pour son contrôle et son suivi, c'est l'attribut "id_carte". Toutefois, le numéro unique de la carte physique est anonymisé et toutes les informations de personnalisation de la carte OPUS (carte avec photo ou enregistrée) ne peuvent pas être croisées avec les données de validations. Les informations sur le profil des usagers et les données sociodémographiques sont inconnues. De plus, chaque transaction est positionnée dans le temps avec l'attribut "date_heure". La mesure de l'heure de validation est précise à la seconde.

L'attribut "code_type_support" définit le type de carte sur lequel le billet est enregistré. Comme à l'étape d'échantillonnage seules les données provenant d'une carte OPUS sont sélectionnées. Un seul élément compose cet attribut.

Le code du produit "code_produit" fait référence au titre de transport qui est utilisé lors de la transaction. La STM propose un grand nombre de types de produits tarifaires afin de répondre aux besoins des voyageurs. En particulier, le billet va dépendre de la durée de validité souhaitée (un ou plusieurs passages, une journée, une semaine, un mois, etc.) et du tarif (ordinaire, réduit, étudiant, gratuit, spécial). Dans ce projet, seules les validations provenant d'abonnements mensuels et annuels sont utilisées.

Comme son nom l'indique, le "code_ligne" informe sur la ligne empruntée lors de la validation. Cependant, cet attribut ne peut pas être utilisé lors d'une transaction dans une station de métro car il est possible de prendre plusieurs lignes différentes. Le Code d'emplacement billettique (CEB) "code_ceb" permet alors de compléter le manque d'information. C'est un

élément permettant d'identifier sur quelle machine de perception une carte a été validée. Le CEB apporte ainsi l'information du mode utilisé (bus ou métro) et localise la station d'entrée dans le métro pour les transactions de ce mode. Autrement dit, dans le cas d'une transaction en bus, c'est le "code_ligne" qui est utilisé et lors d'une transaction en métro, c'est le CEB. Il est important de noter que la table des données brutes provenant des cartes à puce de la STM ne donne aucune information sur le lieu d'embarquement (l'arrêt) dans un bus.

Le "code_direction" donne une indication sur la direction de la ligne de bus empruntée. Pour maintenir cette information à jour, les chauffeurs doivent régler manuellement le système de perception du bus à chaque début de voyage. Par conséquent, les erreurs de manipulation et les oublis conduisent à une mauvaise précision de l'attribut. Il n'est pas utilisé dans l'étude.

Enfin, le champ "act_cal" représente l'activité recalculée de la carte. Il s'agit d'un moyen pour la STM de lier deux transactions entre elles, de sorte à différencier les correspondances des nouveaux déplacements. L'activité recalculée de la carte est définie par l'application des règles d'affaires de la STM, régissant l'opération et l'utilisation du réseau (définition 3.1).

Définition 3.1. Activité recalculée [62]
Les activités de validation doivent être triées par support ("id_carte"), date et heure. Un nouveau déplacement débute (1re montée) lorsque l'une des conditions suivantes est rencontrée :

- c'est la première validation de la carte sur la journée;
- le temps écoulé depuis la première validation du déplacement est égal ou supérieur au temps de durée de correspondance (120 minutes pour un titre de la STM);
- la ligne courante est identique à la précédente;
- la validation actuelle est faite dans une station de métro, ainsi que la précédente.

Dans les autres cas, la validation est une correspondance et fait partie du déplacement.
Une activité de validation recalculée est donc créée pour tenir compte de ces critères.
L'activité de validation recalculée prend les valeurs suivantes :

- D1 : 1re montée d'un déplacement;
- DC : correspondance.

3.3 Algorithme de transformation des transactions

Les données de cartes à puce générées à partir du système OPUS de la STM manquent d'information sur les déplacements puisqu'en raison des correspondances, une transaction n'est pas toujours équivalente à un déplacement complet. Lorsqu'une correspondance a lieu, le déplacement peut être constitué de plusieurs transactions successives.

Pour étudier les correspondances, il est nécessaire de fusionner les transactions qui sont associées à un même déplacement. C'est dans cette vision qu'est développé l'algorithme VALIDEP (des validations au déplacement). Son objectif est de détecter et de dénombrer l'ensemble des correspondances qui sont faites par les usagers possédant une carte à puce, tout en essayant de les localiser. La suite de cette section décrit la succession d'étapes de l'algorithme, dont l'ensemble est schématisé à la figure 3.4. VALIDEP est programmé dans le langage `R 3.5.1` (`https://www.r-project.org`) avec le logiciel `RStudio` (`https://rstudio.com`).

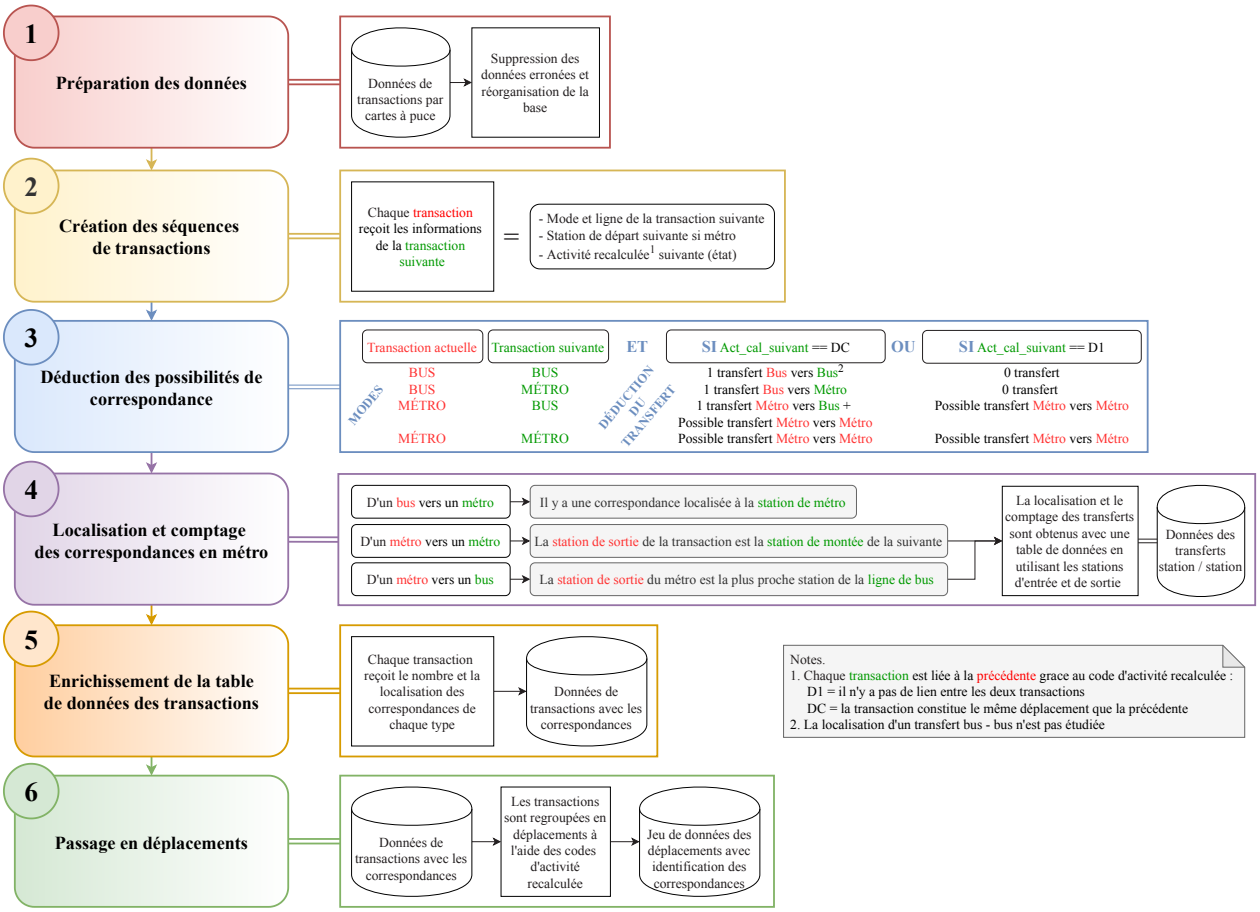


Figure 3.4 Algorithme de transformation des transactions en déplacements VALIDEP

3.3.1 Étape 1 : Préparation des données

L'échantillon des données brutes de la STM est initialement importé en format .csv (de l'anglais *comma-separated values*) puis est immédiatement converti dans le format .rds (*R data serialization*). C'est un format de fichier propre à `R` permettant un encodage des données performant avec une conservation des types des variables. Une fois les données converties, chaque attribut est renommé en suivant les appellations du tableau 3.2, afin de garantir une bonne lisibilité et une unicité des entrées de la table. Dans l'objectif de rendre cet algorithme applicable à tout jeu de données de transactions par cartes à puce, les normalisations suivantes sont réalisées :

- découpage du champ "date_heure" en deux champs "date" et "heure";
- déduction du mode de transport utilisé ("mode") à partir de l'attribut "code_ceb";
- création du champ "code_station" qui indique la station de la validation de la carte; il est déduit du "code_ceb" pour le métro et prend la valeur 0 pour le bus;
- nettoyage des champs pour ne garder que l'information essentielle.

À la fin de cette étape, les champs disponibles dans la table sont ceux présentés dans le tableau D.1 de l'annexe D. Les données sont ensuite filtrées afin de supprimer les erreurs présentes dans la table. Les erreurs sont typiquement :

- un enregistrement sans numéro de carte;
- un enregistrement sur une ligne de transport inexistante;
- un enregistrement dont la ligne utilisée est incompatible avec le mode déduit;
- un enregistrement avec une donnée manquante.

Finalement, la table est réorganisée pour offrir une meilleure compréhension. Les attributs sont triés des plus généraux aux plus spécifiques (ordre du tableau D.1) et les transactions sont classées par date, carte et heure.

3.3.2 Étape 2 : Création des séquences de transactions

Il est nécessaire de comparer les transactions deux à deux pour déterminer si elles sont associées à un même déplacement ou non. Cependant, pour gagner en rapidité d'exécution sur `R`, il est conseillé de réaliser des calculs entre les colonnes plutôt qu'entre les lignes. Ainsi, cette étape vise à préparer le jeu de données en ajoutant à chaque transaction les informations de la suivante. Cela augmente la masse de données enregistrées dans la table, mais diminue le nombre d'itérations de calculs, ce qui offre un meilleur rendement aux étapes 3 et 4.

À l'aide d'une copie décalée de la table, chaque transaction d'une carte chaque jour reçoit le mode, la ligne, la station et l'activité recalculée de la transaction suivante. Les attributs copiés sont renommés avec le suffixe "_suivant". Cette méthode est inspirée des travaux de Hofmann et O'Mahony (2005) [22].

Une modification est apportée à la dernière transaction quotidienne de chaque carte. Afin d'appliquer la méthode développée par He et Trépanier (2015) [19], la dernière transaction quotidienne de chaque carte reçoit, dans les attributs "_suivant", les informations de la première transaction de cette même carte ce même jour. Cette action est justifiée au paragraphe 3.3.4.

Cependant, il est à noter que dans cette étude, l'algorithme développé par He et Trépanier (2015) [19] ne fonctionne pas si une seule transaction en métro est réalisée dans la journée. Comme elles sont peu nombreuses (0,97 %), ces transactions sont supprimées des données.

3.3.3 Étape 3 : Déduction des possibilités de correspondance

Deux modes sont à l'étude, le bus et le métro. Par conséquent, quatre types de correspondances existent : bus – bus (BB), bus – métro (BM), métro – bus (MB) et métro – métro (MM). Comme les usagers doivent valider leur carte à puce à l'embarquement d'un bus et à l'entrée d'une station de métro, les correspondances BB, BM et MB sont facilement identifiables. En effet, elles sont visibles dans le jeu de données comme deux transactions successives dont l'activité recalculée de la seconde indique une correspondance. Cependant, comme il n'y a pas de validation de carte lors d'un changement de métro, les correspondances MM ne sont pas directement visibles dans les données. Chaque fois qu'une transaction a lieu dans le métro, il peut y avoir une ou plusieurs correspondances MM.

Ainsi, une première lecture de la table est faite en utilisant uniquement les attributs "mode", "mode_suivant" et "act_cal_suivant". La confrontation des informations va permettre de déterminer si les deux transactions sont liées par une correspondance ou non :

Bus – bus Si le bus est le mode des deux transactions et que le champ d'activité recalculée de la seconde transaction indique une suite de déplacement ("DC"), alors il y a une correspondance BB. Sinon, il n'y a pas de correspondance BB.

Bus – métro Si les modes de la première et de la seconde transaction sont respectivement bus et métro et que le deuxième champ d'activité recalculée est "DC", alors il y a une correspondance BM. Dans ce cas, le lieu de la correspondance est la station d'entrée dans le métro. Sinon, il n'y a pas de correspondance BM.

Métro – bus Si les modes de la première et de la seconde transaction sont respectivement métro et bus et que le deuxième champ d'activité recalculée est "DC", alors il y a une correspondance MB. Dans ce cas, le lieu de la correspondance doit être déterminé. Sinon, il n'y a pas de correspondance MB. Comme la première transaction est en métro, il peut y avoir des correspondances MM. Il faut donc étudier ce cas plus en profondeur.

Métro – métro Si le métro est le mode des deux transactions, alors le deuxième champ d'activité recalculé est toujours "D1", mais il peut y avoir des correspondances de type MM. Il faut donc étudier ce cas plus en profondeur.

Ainsi, grâce à une première lecture de la table est détectée la présence de trois des quatre types de correspondances existantes (BB, BM et MB). Il reste à localiser les correspondances MB et à détecter les correspondances MM.

3.3.4 Étape 4 : Localisation et comptage des correspondances en métro

C'est à cette étape qu'il est défini précisément le nombre et le lieu des correspondances. L'objectif est de déterminer la station de sortie du métro pour chaque transaction concernée. Cela est possible grâce à une adaptation de l'algorithme développé par He et Trépanier (2015) [19]. Cette méthode repose sur le principe suivant : le lieu de descente d'un passager est l'arrêt de sa ligne qui est le plus proche du lieu d'embarquement de sa prochaine transaction. De même, la dernière transaction de la journée est comparée avec la première puisque la méthode suppose un retour au lieu d'origine en fin de journée. Deux situations distinctes sont alors à traiter.

Une transaction en métro suivie d'une transaction en métro La station de sortie de la première transaction en métro est considérée comme étant la station d'entrée de la transaction suivante.

Une transaction en métro suivie d'une transaction en bus La station de sortie du métro est considérée comme étant la station la plus proche de la ligne de bus utilisée après. Cela est déterminé grâce à un programme `QGIS` (`https://www.qgis.org`) qui indique pour chaque ligne de bus, la station de métro la plus proche. Si ces deux transactions sont liées par une correspondance MB, alors celle-ci est localisée dans la station de sortie du métro.

Une fois la station de sortie déterminée, il est possible de déduire la présence d'une ou plusieurs correspondances MM. Comme le réseau de métro de la STM comporte 68 stations, une matrice des correspondances station / station est utilisée. Cette matrice créée manuellement indique pour chaque couple de stations entrée / sortie le nombre de correspondances

nécessaires et les stations de changement. Sa création repose sur les deux hypothèses suivantes : l'usager minimise le nombre de correspondances et d'arrêts à faire. Finalement, grâce à cette étape, les correspondances en métro sont déduites et localisées. Comme les lieux d'embarquements en bus sont inconnus, il n'est pas possible de déduire la localisation des correspondances BB.

3.3.5 Étape 5 : Enrichissement de la table de données

Les étapes 2, 3 et 4 ont permis de détecter pour chaque transaction l'ensemble des correspondances et de localiser toutes celles qui impliquent un métro. L'étape 5 vise à réorganiser la table afin de la rendre utilisable pour toute la suite de l'étude. Plusieurs attributs sont ajoutés à la table originale :

- "id_dep" : l'identifiant du déplacement. Toutes les transactions impliquées dans un même déplacement ont la même valeur d'attribut. Cela permet d'identifier chaque déplacement et d'avoir accès à toutes les transactions dont il est composé.
- "BB" : 1 s'il y a une correspondance BB, 0 sinon.
- "BM" : 1 s'il y a une correspondance BM, 0 sinon.
- "MB" : 1 s'il y a une correspondance MB, 0 sinon.
- "MM" : le nombre de correspondances MM ($0 \leq$ MM ≤ 2).
- "BM_loc" : le code de la station de métro pour la correspondance BM (0 si BM $= 0$).
- "MB_loc" : le code de la station de métro pour la correspondance MB (0 si MB $= 0$).
- "MM_loc_1" : code de la première station des correspondances MM (0 si MM $= 0$).
- "MM_loc_2" : code de la deuxième station des correspondances MM (0 si MM ≤ 1).

Finalement, l'ensemble des attributs composant la table des transactions enrichie avec les correspondances est ordonné, comme l'illustre par un exemple le tableau D.2. Ce nouveau jeu de données de transactions est enregistré au format .rds pour pouvoir être utilisé dans toute la suite de l'étude.

3.3.6 Étape 6 : Passage en déplacements

Le jeu de données des transactions est maintenant enrichi avec toutes les informations concernant les correspondances. La dernière étape consiste alors à regrouper les transactions qui ont le même identifiant pour obtenir la table des déplacements. Lors du regroupement, l'enregistrement du déplacement combine : l'identifiant, le nombre de transactions impliquées, le

numéro de la carte, la date et l'heure de la première transaction, le nombre total de correspondances, le nombre de correspondances de chaque type, les localisations des correspondances de chaque type et la succession des modes utilisés.

Avant d'être enregistrée, la nouvelle table des déplacements est nettoyée, triée et le format de chaque variable est vérifié, comme l'illustre le tableau D.3 avec le déplacement 684. Cette table est enregistrée au format .rds pour pouvoir être utilisée dans toute la suite de l'étude. Comme le jeu de données des transactions et le jeu de données des déplacements intègrent tous les deux le champ "id_dep", il est aisé de communiquer de l'un à l'autre pour accéder à toutes les informations souhaitées.

3.3.7 Exemple explicatif des étapes de l'algorithme

Afin d'aider à la compréhension des étapes de l'algorithme VALIDEP, le tableau 3.3 présente un cas d'étude simplifié composé de quatre transactions successives d'une même carte. L'ensemble de la démarche est schématisé sur la figure 3.5.

Tableau 3.3 Transactions du cas d'étude

id_validation	id_carte	date	heure	mode	code_ligne	code_station	act_cal
A	1	2016-05-14	10:11:12	B	43	0	D1
B	1	2016-05-14	10:25:32	B	33	0	DC
C	1	2016-05-14	10:58:47	M	0	Langelier	DC
D	1	2016-05-14	11:40:01	B	72	0	DC

À l'étape 2, chaque transaction reçoit les informations de la transaction suivante. B est alors ajoutée à A, C à B, etc. Comme la transaction D est la dernière, elle reçoit les informations de la transaction A. Cette étape est illustrée par le tableau 3.4.

L'étape 3 utilise les informations de l'étape 2 pour déterminer les possibilités de correspondances. Les transactions A et B étant en bus et l'activité recalculée de B étant "DC", il y a une correspondance BB entre les lignes 43 et 33. Elle n'est pas localisée. L'activité recalculée de la transaction C étant "DC", il y a une correspondance BM entre B et C. Celle-ci est localisée à la station de la transaction C, soit Langelier. La suite de transactions C et D indique une possibilité de correspondance MM (puisque C est du mode M) et l'assurance d'une correspondance MB car l'activité recalculée de la transaction D est "DC".

Il faut déduire la station de sortie du métro pour connaître le nombre de correspondances MM et la position de la correspondance MB. La transaction en métro (C) est suivie d'une transaction en bus (D). Par conséquent, la station de sortie du métro est celle qui est la plus proche de la ligne de bus utilisée en D, soit la ligne 72. Grâce au programme `QGIS`, il est connu que la station la plus proche de la ligne 72 est la station Du Collège. Ainsi, la sortie de la transaction C est Du Collège. D'après la matrice des correspondances station / station, pour rejoindre les stations Langelier et Du Collège, il faut faire un changement de métro par la station Lionel-Groulx. Toutes les correspondances sont donc dénombrées et localisées. Le résultat est donné au tableau 3.5. Finalement, le regroupement de ces transactions permet d'obtenir le résultat présenté dans le tableau 3.6.

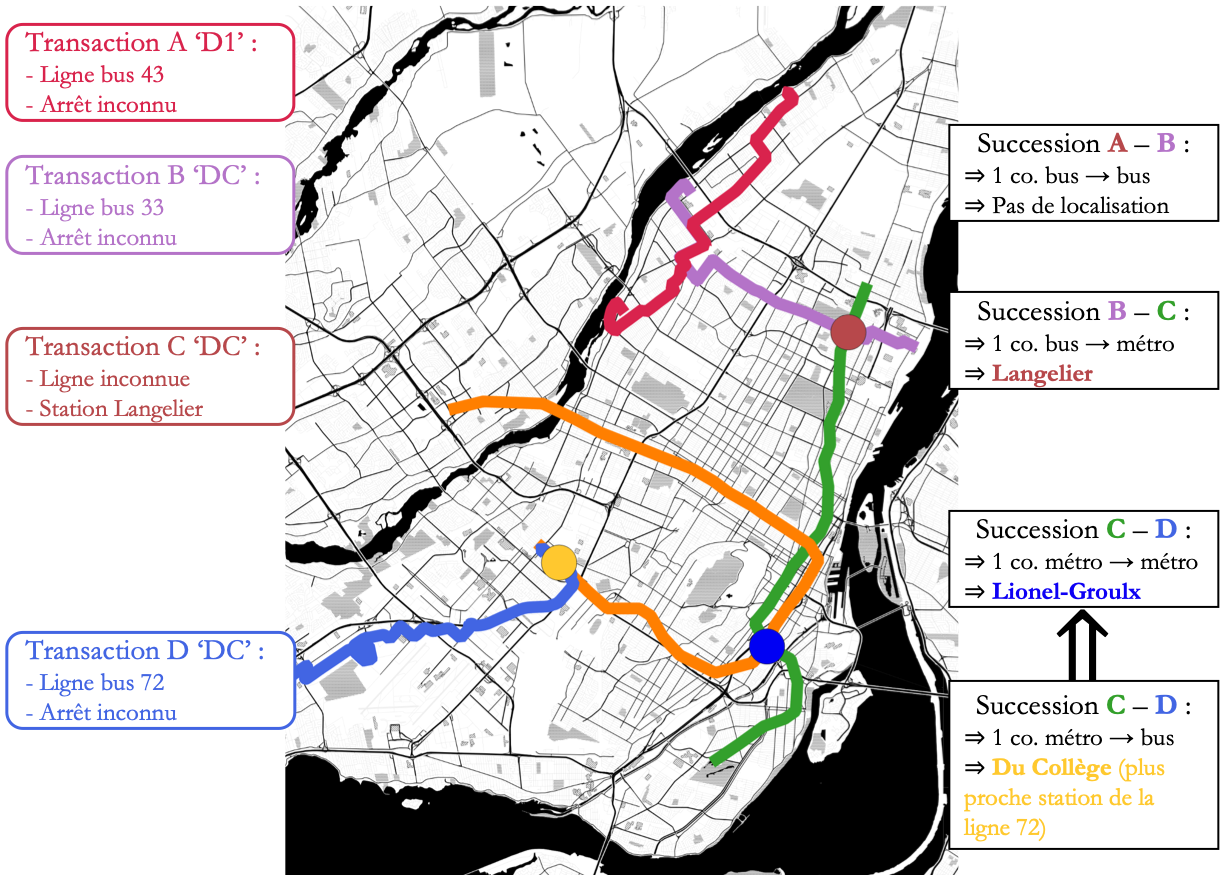


Figure 3.5 Succession des transactions du cas d'étude

Tableau 3.4 Table des transactions du cas d'étude après l'étape 2

id_validation	id_carte	date	heure	mode	mode_suivant	code_ligne	code_ligne_suivant	code_station	code_station_suivant	act_cal	act_cal_suivant
A	1	2016-05-14	10:11:12	B	B	43	33	0	0	D1	DC
B	1	2016-05-14	10:25:32	B	M	33	0	0	Langelier	DC	DC
C	1	2016-05-14	10:58:47	M	B	0	72	Langelier	0	DC	DC
D	1	2016-05-14	11:40:01	B	B	72	43	0	0	DC	D1

Tableau 3.5 Table des transactions du cas d'étude après l'étape 4

id_dep	id_val	id_carte	date	heure	mode	code_ligne	code_station	act_cal	BB	BM	BM_loc	MB	MB_loc	MM	MM_loc_1	MM_loc_2
1	A	1	2016-05-14	10:11:12	B	43	0	D1	1	0	0	0	0	0	0	0
1	B	1	2016-05-14	10:25:32	B	33	0	DC	0	1	Langelier	0	0	0	0	0
1	C	1	2016-05-14	10:58:47	M	0	Langelier	DC	0	0	0	1	Du Collège	1	Lionel-Groulx	0
1	D	1	2016-05-14	11:40:01	B	72	0	DC	1	0	0	0	0	0	0	0

Tableau 3.6 Déplacement obtenu à partir des transactions du cas d'étude

id_dep	nb_val	id_carte	date	heure	modes	nb_corresp	BB	BM	BM_loc	MB	MB_loc	MM	MM_loc_1	MM_loc_2
1	4	1	2016-05-14	10:11:12	B,B,M,M,B	4	1	1	Langelier	1	Du Collège	1	Lionel-Groulx	0

3.4 Typologie des lignes de bus

La section 3.3 a décrit le fonctionnement de l'algorithme VALIDEP en explicitant ces résultats : un enrichissement des données brutes par l'ajout d'informations sur les correspondances réalisées sur le réseau. L'obtention de ces nouvelles tables permet de poursuivre le projet en étudiant la répartition des correspondances, le comportement des usagers de l'échantillon et les interactions entre les lignes. Cependant, puisque les correspondances impliquent nécessairement plusieurs lignes de transport, il est indispensable de définir à l'avance une typologie des lignes. Cela pour identifier les lignes similaires afin d'enrichir les modèles.

Nombreux sont les facteurs qui peuvent caractériser une ligne de transport : son mode, sa catégorie, sa capacité, sa fréquence, etc. Dans le cadre de ce projet, l'objectif est de définir les lignes de bus selon des caractéristiques pouvant avoir un impact sur leur utilisation. Il s'agit d'avoir une table où chaque ligne est spécifiée. Les attributs suivants sont retenus :

Classification de la STM Les lignes de bus sont classifiées par la Société de transport de Montréal en fonction du service offert. La répartition des 6 types de lignes (présentées à la section 3.2.1) est prise en compte pour élaborer la table. Cette classification intègre les données opérationnelles sur les lignes, ce qui permet d'alléger la typologie.

Positionnement géographique Le positionnement d'une ligne sur le territoire impacte son utilisation. Pour visualiser cette influence, la présence ou non de chaque ligne sur les régions de l'Île de Montréal est renseignée. Comme l'illustre la figure 3.6, il y a quatre régions sur l'Île : le centre-ville, le centre, l'Est et l'Ouest.

Nombre de correspondances potentielles avec le bus La possibilité de faire une correspondance augmente avec le nombre d'arrêts communs entre deux lignes. Ainsi, un programme `QGIS` est développé pour compter, pour chaque ligne, le nombre d'arrêts communs avec d'autres lignes de bus.

Nombre de lignes de bus croisées Avec l'attribut précédent, il est également possible de compter le nombre de lignes de bus qui sont accessibles depuis la ligne considérée.

Nombre de correspondances potentielles avec le métro Les lignes de bus sont régulièrement utilisées pour faire du rabattement vers un mode plus lourd. Autrement dit, les lignes sont conçues pour transporter des usagers vers et depuis le métro. Ainsi, le programme `QGIS` compte le nombre de fois qu'une station de métro est accessible depuis la ligne (en prenant un rayon de 150 mètres autour de la station). Pour chaque ligne de bus, la liste des stations accessibles est également disponible.

Nombre de lignes de métro croisées Avec la liste des stations accessibles depuis chaque ligne de bus, il est également possible de savoir le nombre de lignes de métro croisées.

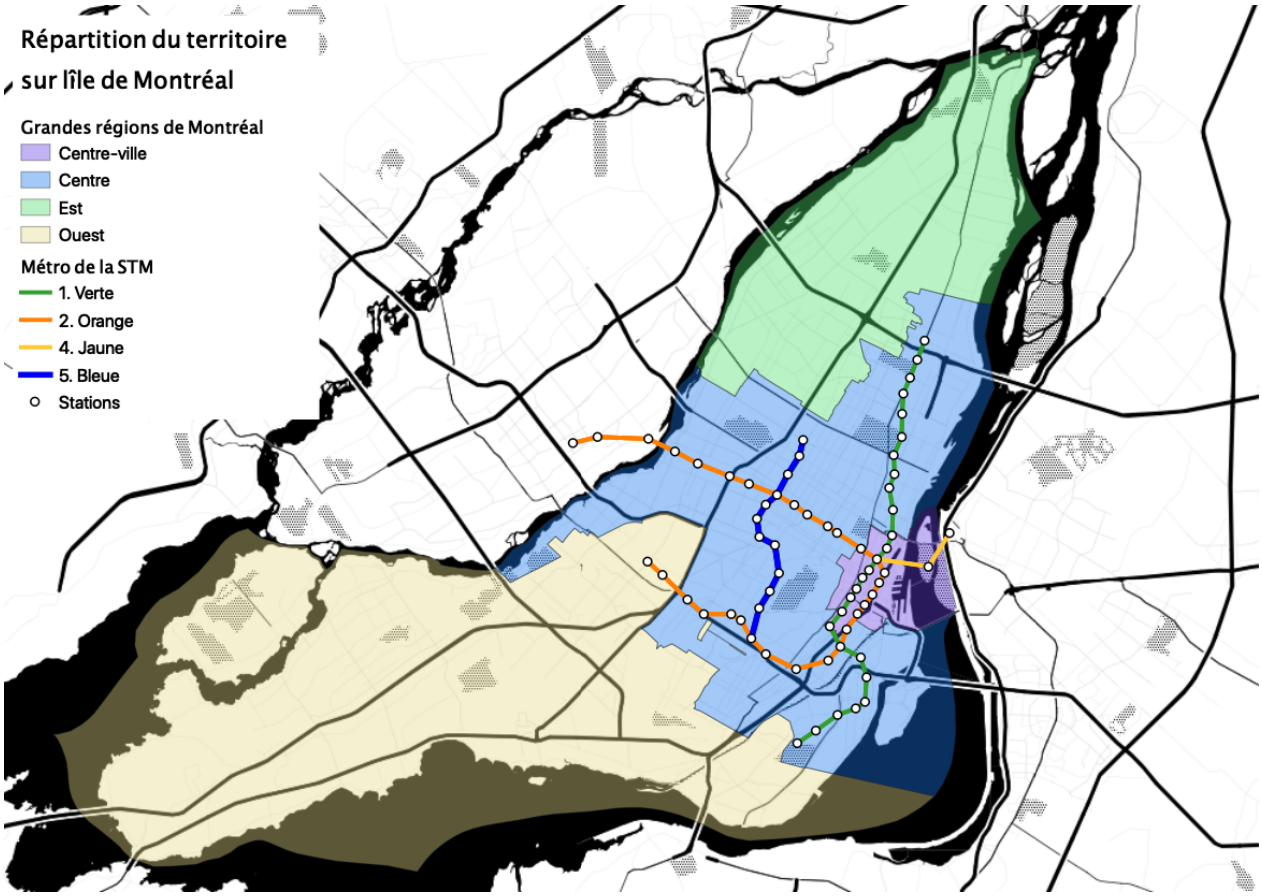


Figure 3.6 Régions de l'Île de Montréal

Finalement, une table permettant de décrire les lignes du réseau de bus de la STM est obtenue. Un exemple est donné dans le tableau 3.7. La prise en compte de cette typologie dans la suite du projet vise à mettre en avant les attributs qui ont le plus d'influence sur le taux de correspondances d'une ligne ainsi que sur le comportement des usagers. Comme le réseau de la STM comporte seulement quatre lignes de métro, l'intérêt de développer une typologie pour ces lignes est limité. Cette typologie est centrée sur les lignes de bus.

La typologie des lignes proposée permet de caractériser chaque ligne du réseau suivant son implantation sur l'Île de Montréal. Il est alors possible de regrouper entre elles les lignes qui ont le même positionnement géographique, c'est-à-dire les mêmes régions traversées. Il vient ainsi dix combinaisons représentées visuellement dans l'annexe E. Ce regroupement des lignes est utile afin de partager les usagers en fonction des régions qu'ils parcourent. Dans cet objectif, chaque combinaison de régions est identifiée à l'aide d'un repère (de A à J). L'analyse de l'utilisation des lignes de chaque combinaison de régions est effectuée à la partie 4.3.4.

Tableau 3.7 Typologie des lignes de bus

Ligne	Classification	Région de l'Île				N. lignes		N. arrêts		Croisement du métro			
	STM	CV	Centre	Est	Ouest	bus	métro	bus	métro	verte	orange	jaune	bleue
10	local	1	1	0	0	6	1	6	1	1	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
51	10' max	0	1	0	1	9	2	60	4	0	1	0	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
440	express	0	1	1	0	15	1	29	1	0	1	0	0

3.5 Classification des déplacements : séries de modes

À la suite des étapes précédentes, les correspondances réalisées sont identifiées et il est possible de continuer l'étude en se concentrant parallèlement sur le réseau et sur ses utilisateurs. Cependant, il faut définir préalablement la variable à utiliser pour les mesures. En effet, son choix impacte l'élaboration de la méthodologie et peut avoir un effet sur la véracité des résultats obtenus. La question est de savoir quel est l'attribut (ou la combinaison d'attributs) du jeu de données qu'il est pertinent d'utiliser pour caractériser l'utilisation des correspondances. Il faut également définir l'unité à employer. S'agit-il de la correspondance ? Du déplacement ? De la transaction ?

L'unité de mesure choisie pour l'étude doit répondre aux critères suivants :

- prendre en compte le nombre de correspondances de chaque déplacement;
- prendre en compte le type des correspondances de chaque déplacement;
- prendre en compte l'ordre des correspondances de chaque déplacement;
- être indépendante du nombre de déplacements et de correspondances.

Une solution combinant tous ces critères consiste à classifier les déplacements selon la série de modes utilisée. C'est l'attribut "modes" de la table des déplacements qui résulte de l'algorithme VALIDEP. Chaque changement de ligne est enregistré, ce qui permet de suivre l'évolution du déplacement. Ainsi, en lisant l'attribut "modes", l'analyste détermine combien de correspondances sont faites, de quels types et dans quel ordre. Par exemple, la série de modes du déplacement du tableau 3.6 est B,B,M,M,B. C'est-à-dire que l'usager a pris un bus, puis un bus, puis un métro, puis un métro, puis un bus. Il y a donc quatre correspondances, une BB, puis une BM suivie d'une MM et enfin d'une MB.

Cette solution a plusieurs avantages. Tout d'abord, elle est simple à comprendre tout en conservant un grand nombre d'informations. Elle permet aussi bien d'analyser un déplacement direct (B ou M) qu'un déplacement avec beaucoup de correspondances. De plus, en utilisant cet attribut, l'unité de mesure est le nombre de déplacements. En effet, comme chaque déplacement est associé à une seule série de modes, la somme totale de l'attribut sera toujours égale au nombre de déplacements enregistrés dans la table. Cela apporte de la clarté au modèle et permet de travailler en unité absolue : le rapport entre le nombre de déplacements de la série et le nombre total de déplacements.

L'analyse des déplacements obtenus avec l'algorithme VALIDEP montre qu'il y a une grande diversité de séries de modes. En effet, sur l'échantillon de 10 000 cartes avec un an de transactions, presque 500 combinaisons différentes sont révélées. Il est donc difficile de les étudier individuellement. Néanmoins, le dénombrement des déplacements de chaque série montre que les huit combinaisons de modes les plus utilisées représentent 87,4 % de tous les déplacements. En conséquence, pour simplifier l'étude, les séries les moins représentées sont regroupées dans une seule série notée "Autre". Le tableau 3.8 permet de caractériser les séries de modes finalement conservées pour classifier les déplacements. Dans la suite de ce mémoire, le terme "série" fait toujours référence à cette classification, qui est utilisée dans toutes les analyses.

Tableau 3.8 Classification des déplacements : séries de modes

Série	Définition	Correspondance	Proportion des déplacements*	Proportion cumulée
B	1 bus	0	22,00 %	22,00 %
M	1 métro	0	21,34 %	43,34 %
M,M	2 métros	1 (MM)	10,23 %	53,57 %
B,M	bus - métro	1 (BM)	8,91 %	62,48 %
B,B	2 bus	1 (BB)	8,80 %	71,28 %
M,B	métro - bus	1 (MB)	8,04 %	79,32 %
M,M,B	2 métros - bus	2 (1 MM / 1 MB)	4,20 %	83,52 %
B,M,M	bus - 2 métros	2 (1 BM / 1 MM)	3,85 %	87,37 %
Autre	Tout le reste	Non défini	12,64 %	100,00 %

* Proportion des 5 887 074 déplacements enregistrés sur l'année 2016 par les 10 000 cartes à puce sélectionnées pour l'étude.

3.6 Analyse des déplacements quotidiens

L'objectif est de développer une méthode pour repérer les séries de modes favorisées par les usagers de l'échantillon étudié. Après avoir caractérisé la répartition générale des déplacements obtenus avec l'algorithme VALIDEP, la recherche se divise en deux parties. En premier, la répartition des correspondances sur une journée est étudiée pour montrer la variation de l'utilisation du réseau. Ensuite, un indicateur de symétrie des déplacements est proposé pour expliquer comment les voyageurs de l'échantillon ont recours aux correspondances lors de leurs déplacements journaliers.

3.6.1 Répartition des correspondances

Les habitudes de déplacements des voyageurs dépendent de plusieurs paramètres. En particulier, les données montrent que l'utilisation du réseau est différente selon l'heure de la journée et le jour de la semaine (figure 4.3). Entre les contraintes opérationnelles et personnelles, la charge à bord de chaque ligne varie continuellement. Cette étude veut mettre en évidence et quantifier la variation horaire de l'usage des correspondances.

Pour cela, les données de déplacements sont agrégées par heure et série de modes. Le programme compte, pour chaque heure, le nombre total de déplacements ainsi que le nombre de déplacements de chaque série. Les jours de la semaine sont décomposés en deux groupes, les jours ouvrés (lundi à vendredi) et les jours de fin de semaine (samedi et dimanche). Afin de ne pas biaiser l'étude, les jours fériés et les semaines de vacances scolaires sont retirés de la table. En 2016, le premier jour de l'année était un vendredi et le dernier était un samedi, les deux semaines concernées sont donc incomplètes. Par conséquent, elles sont supprimées des données. Enfin, le métro de la STM était ouvert durant 24 heures pendant la nuit blanche de Montréal, le 28 février 2016. Cette journée est également supprimée. Les jours atypiques de 2016 sont listés en annexe F.

En additionnant le nombre de déplacements total et de chaque série, en fonction du groupe d'heure et du type de jour, il est possible de calculer la proportion moyenne de déplacements de chaque série de modes. Sa représentation graphique avec un histogramme permet de mettre en avant la variation horaire de l'utilisation des correspondances. L'avantage de cette méthode de calcul est qu'elle ne prend pas en compte le nombre de déplacements. Elle est centrée sur la répartition des correspondances faites. Cette étude apporte des connaissances sur les séries de modes favorisées par les usagers de l'échantillon en fonction du moment de la journée.

3.6.2 Symétrie des déplacements pendulaires

L'étude de la répartition des correspondances met en avant une inversion de l'ordre des séries de modes faites durant la journée. Si cette transposition provient des mêmes usagers de l'échantillon, alors elle peut traduire des allers-retours pendulaires. Ainsi, cette partie s'intéresse à expliquer comment les voyageurs modifient leur utilisation des correspondances entre l'aller et le retour d'une activité.

Pour cela, la structure des déplacements d'une même personne sur une journée est analysée. Premièrement, la répartition des écarts de temps entre deux déplacements est représentée graphiquement, de même que la distribution du nombre de déplacements quotidiens des passagers. Cela permet de décrire les séquences d'activités générées par le TC.

La symétrie des déplacements allers-retours est ensuite explorée. Pour ce faire, les déplacements sont agrégés en cartes-jours. C'est-à-dire que tous les déplacements d'une carte sur une journée sont regroupés dans le même enregistrement. Pour simplifier l'étude, les déplacements en fin de semaine et la nuit (entre 21:00 et 5:00) sont retirés et seules les cartes-jours composées de deux déplacements sont conservées. Trois critères sont proposés pour mesurer la symétrie des déplacements :

Horaires Comment les deux déplacements sont positionnés dans la journée ? Une symétrie existe-t-elle entre le déplacement de l'aller et celui du retour ? Autour de quelle heure se situe le point de symétrie ?

Séries de modes Existe-t-il un lien entre l'ordre de la série de modes du matin et celle du soir ? Une typologie de similarité des séries est proposée pour mesurer la part des cartes-jours avec un aller-retour symétrique. Elle est expliquée à la figure 3.7.

Lignes Est-ce que les lignes utilisées à l'aller et au retour sont identiques ? La proportion des cartes-jours qui utilisent les mêmes lignes à l'aller et au retour en ordre inverse est calculée en suivant le processus de la figure 3.8.

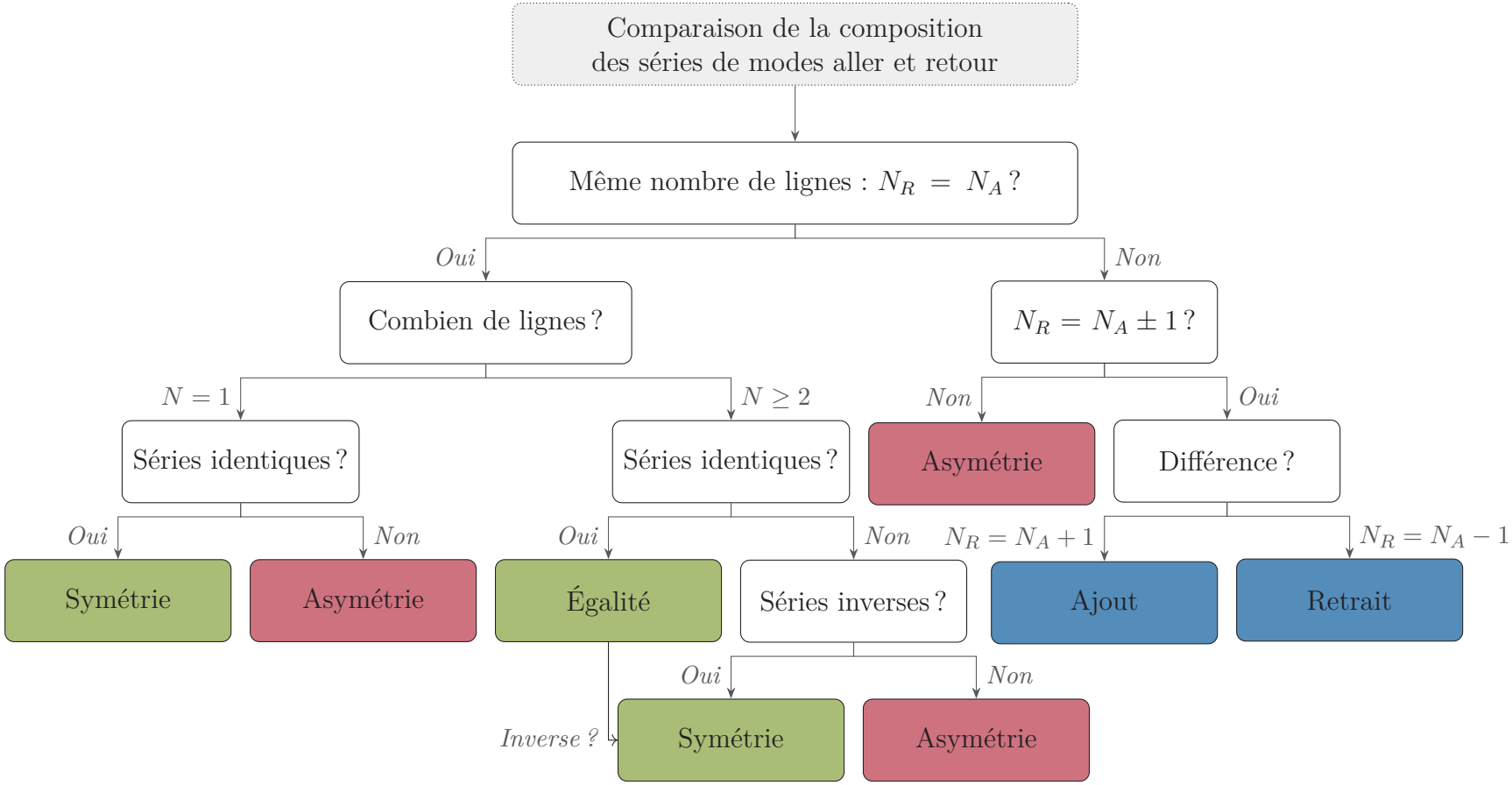


Figure 3.7 Arbre de décision pour évaluer la symétrie des séries de modes

N_A : Nombre de lignes dont est composée la série de modes de l'aller.

N_R : Nombre de lignes dont est composée la série de modes du retour.

Si $N_A = N_R$ alors $N = N_A = N_R$.

Appliquée à deux déplacements successifs d'une même carte-jour répondant aux exigences de l'étude.

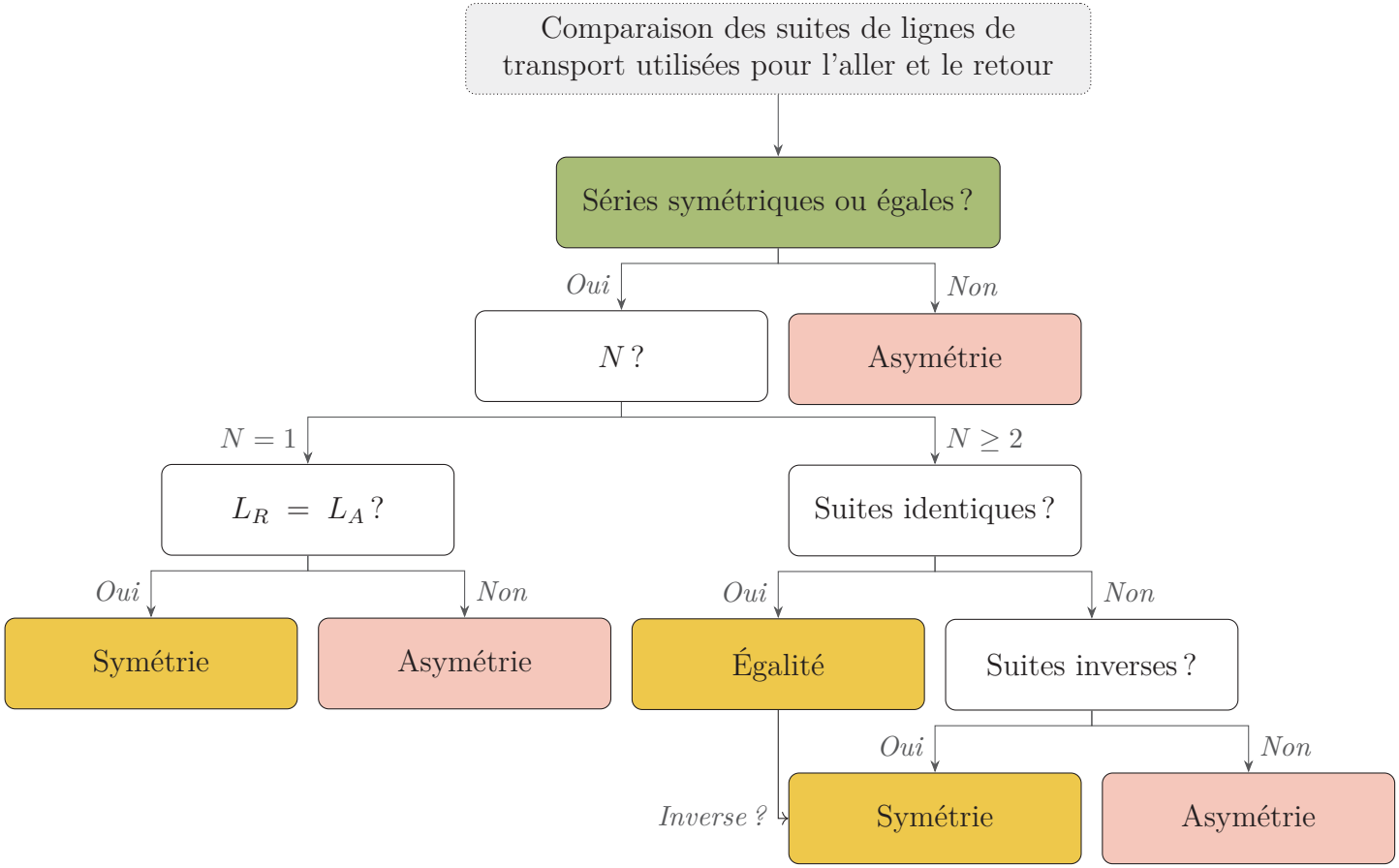


Figure 3.8 Arbre de décision pour évaluer la symétrie des lignes utilisées

N : Nombre de lignes dont sont composées les séries de modes de l'aller et du retour.

L_A : Ligne utilisée à l'aller si $N = 1$. L_R : Ligne utilisée au retour si $N = 1$.

Suites identiques : 51-2-1 et 51-2-1. Symétrie égalité : 51-5-51 et 51-5-51. Symétrie : 51-5-2 et 2-5-51.

Appliquée à deux déplacements successifs d'une même carte-jour répondant aux exigences de l'étude.

3.7 Analyse du comportement des usagers de l'échantillon

L'étude du comportement des utilisateurs du transport en commun ayant une carte à puce s'intègre parmi les quatre objectifs principaux de ce projet. Sa vocation est d'expliquer la façon dont les usagers de l'échantillon ont recours aux correspondances. Tous les voyageurs ne sont pas égaux dans leur manière de se déplacer et cette recherche vise à définir si la présence des correspondances dans le réseau peut expliquer (en partie) cette différence. L'objectif de cette étude est de déterminer s'il existe plusieurs formes de comportements basés sur l'usage des correspondances et si ces agissements sont constants pendant l'année.

La méthode retenue est la segmentation. Il s'agit d'une technique d'exploration de données (de l'anglais *data mining*) visant à partitionner un ensemble d'enregistrements en plus petits groupes d'objets similaires, appelés *clusters*, afin de regrouper ensemble des individus semblables. Il existe une multitude d'algorithmes permettant de segmenter des jeux de données. Ce projet utilise une segmentation par la méthode des k-moyennes (en anglais *k-means*). Il s'agit d'un partitionnement en k groupes de la table. Initialement, l'algorithme sélectionne k points aléatoirement qui deviennent les premiers centres des groupes. Grâce à une mesure de distance, chaque vecteur est alors associé au centre le plus proche. À chaque itération, l'algorithme recalcule les centres des groupes et recommence l'association des points. Tant qu'il y a une évolution dans les groupes et que la limite n'est pas atteinte, le processus itératif continue. Ainsi, l'algorithme des k-moyennes cherche à maximiser autant l'homogénéité à l'intérieur de chaque groupe que l'hétérogénéité entre tous les groupes. C'est la méthode des k-moyennes qui est retenue car elle est particulièrement bien adaptée pour travailler sur des ensembles de données massives [63].

La démarche de la segmentation est schématisée à la figure 3.9. Le choix du vecteur, la détermination du nombre de groupes, leur création et le déroulement des analyses sont exposés dans la suite de cette section.

3.7.1 Création des vecteurs de la segmentation

La pertinence de la segmentation repose sur les attributs à segmenter. En effet, il faut réorganiser la table des déplacements pour qu'elle soit représentative de l'utilisation faite du réseau par les voyageurs. Les informations doivent être centrées sur les passagers plutôt que sur les déplacements. De la même manière, puisque la segmentation doit refléter le comportement des usagers vis-à-vis des correspondances, il faut connaître la composition des déplacements. Enfin, pour que l'étude des variations des comportements soit réalisable, l'agrégation des données doit être suffisamment fine.

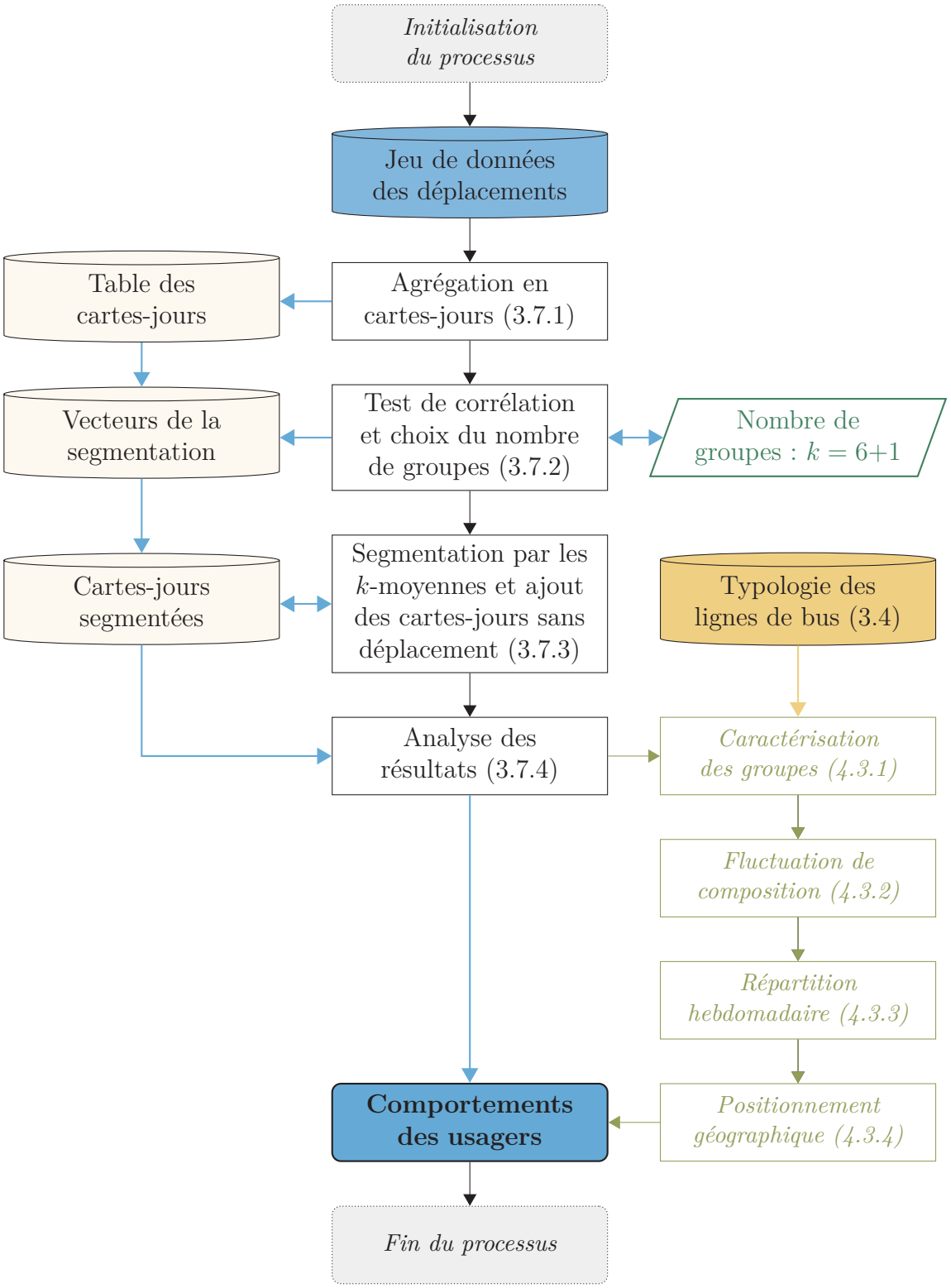


Figure 3.9 Méthodologie utilisée pour la segmentation des usagers

Les données de déplacements sont alors agrégées en cartes-jours. Autrement dit, tous les déplacements quotidiens d'une carte sont regroupés dans un unique enregistrement. Celui-ci est constitué du nombre total de déplacements et de la proportion des déplacements de chaque série de modes réalisés par la carte au jour concerné.

En agrégeant les données en cartes-jours, la composition quotidienne des déplacements de chaque carte est décrite aisément, traduisant l'usage des correspondances. Ainsi, le comportement quotidien d'un usager de l'échantillon est connu, et son évolution jour après jour est visible. De plus, comme la répartition quotidienne des correspondances est examinée dans l'étude précédente, il n'est pas nécessaire de mieux décrire la distribution journalière des déplacements. Le tableau 3.9 illustre les vecteurs d'un échantillon de la table des cartes-jours.

Tableau 3.9 Vecteurs des cartes-jours pour la segmentation

carte-jour	déplacements	norm*	B	M	B,B	B,M	M,B	M,M	B,M,M	M,M,B	Autre
506 – 1	1	0,06	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
506 – 4	2	0,12	0,00	0,00	0,00	0,50	0,50	0,00	0,00	0,00	0,00
506 – 18	7	0,41	0,00	0,29	0,14	0,14	0,14	0,00	0,00	0,00	0,29
506 – 22	6	0,35	0,00	0,33	0,00	0,00	0,00	0,17	0,17	0,17	0,17
506 – 26	2	0,12	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00
506 – 34	5	0,29	0,40	0,40	0,00	0,00	0,00	0,00	0,00	0,00	0,20

* Nombre de déplacements après la normalisation min / max.

Il y a $10\,000 \times 366 = 3\,660\,000$ cartes-jours, ce qui est tout à fait acceptable pour l'utilisation de l'algorithme des k-moyennes. Toutefois, par souci d'allégement et de simplification, les cartes-jours sans déplacement (vecteur nul) sont traitées séparément. De plus, hormis le nombre de déplacements quotidiens, tous les attributs sont naturellement normalisés. Ainsi, une normalisation min / max est appliquée aux nombre de déplacements. Finalement, la table des vecteurs ne comporte pas de trous et toutes les données ont le même format : $\forall\ x \in T_{cartes-jours},\ x \in \mathbb{R}\ /\ x \in [0\,;\,1]$. Par conséquent, une distance euclidienne est utilisée pour faire la segmentation.

3.7.2 Choix du nombre de groupes

La difficulté principale de la méthode des k-moyennes repose sur le choix du nombre de groupes à utiliser, puisque la qualité de la segmentation en dépend. S'il n'y a pas assez de groupes, les résultats ne sont pas consistants et s'il y a trop de groupes l'analyse devient très

complexe et l'intérêt de la segmentation disparaît. Le nombre de groupes optimal est celui qui maximise l'hétérogénéité entre les groupes et l'homogénéité à l'intérieur des groupes, sans surcharger l'analyse de trop de détails.

En premier lieu, un test de corrélation multiple est réalisé sur les vecteurs pour s'assurer que les données ne sont pas liées les unes aux autres. Une corrélation entre plusieurs attributs peut impacter le résultat de la segmentation car l'indépendance entre les données n'est pas respectée. Comme le vecteur contient la part des déplacements de chaque série de modes par jour, la somme des séries est toujours égale à 1. Pour éviter cette corrélation, l'attribut "autre" est retiré lors de la segmentation. Aucun autre lien n'est observé dans les données.

Trois tests sont effectués pour choisir le bon nombre de groupes. Le premier est l'analyse de la distorsion par la méthode du coude (de l'anglais *elbow method*). Il s'agit de faire plusieurs segmentations avec un nombre de groupes k variable. Pour chaque valeur, la distorsion est calculée. C'est la somme des distances au carré entre les points et leur centre le plus proche (équation 3.1). Par définition, les k-moyennes cherchent à minimiser cette valeur. Plus il y a de groupes, moins la distorsion est grande. Le tracé de la distorsion en fonction du nombre de groupes laisse apparaître une rupture, ce qui traduit la valeur de k à partir duquel l'homogénéité dans les groupes est bonne. C'est le coude et il est recommandé de choisir un nombre de groupes après son début.

$$D = \sum_{m=1}^{k} \sum_{x_i \in C_m} (x_i - \mu_m)^2 \tag{3.1}$$

Avec D la distorsion générale
k le nombre de groupes
C_m le groupe numero m
x_i un point du groupe C_m
μ_m le centre du groupe C_m

Sur les données de l'étude (figure 3.10), le coude apparaît entre $k = 5$ et $k = 9$. En conséquence, il est bon de choisir entre six et neuf groupes. En dessous, les groupes risquent de manquer de propretés et au-dessus, il y en aura trop.

Les deuxième et troisième tests sont respectivement la mesure des sauts d'inertie et la représentation du dendrogramme. Pour mettre en place ces deux essais, une segmentation k-moyennes est faite avec cinquante groupes et une classification hiérarchique ascendante (CHA) est appliquée sur les centres. Il s'agit de regrouper les éléments les plus proches deux

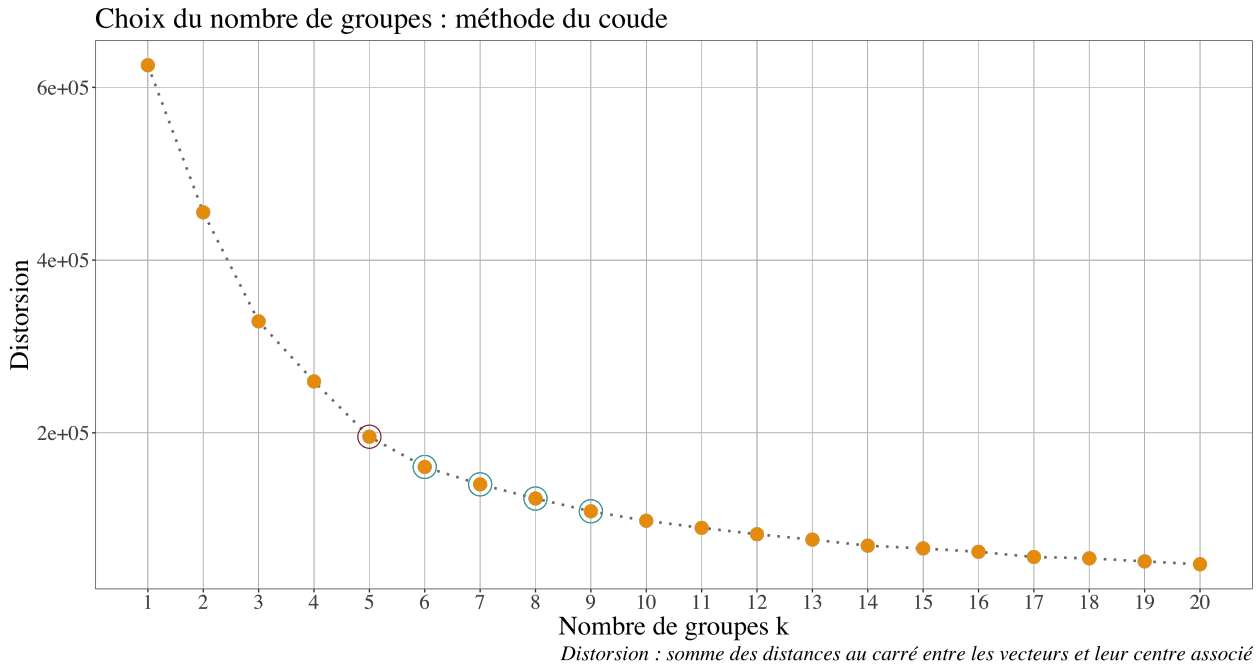


Figure 3.10 Analyse de la distorsion des groupes

à deux jusqu'à obtenir un seul groupe contenant toutes les données. Le CHA est appliqué en suivant la méthode de Ward [64]. La distorsion du partitionnement est alors calculée pour chaque regroupement et son tracé en fonction du nombre de groupes permet d'observer les sauts d'inertie. La méthode conseille de choisir le nombre de groupes offrant la plus grande chute d'inertie. En regardant la figure 3.11, il semble intéressant de faire une segmentation entre cinq et sept groupes.

Les deux premières méthodes combinées conseillent de choisir entre six et sept groupes. En regardant la forme des regroupements sur le dendrogramme (figure 3.12), il semble plus intéressant de choisir une segmentation en six ensembles. En effet, avec six groupes, les ensembles ont une taille approximativement identique. À l'inverse, l'ajout d'un septième groupe crée une différence de taille conséquente entre eux.

Pour résumer, le choix du nombre de groupes de la segmentation se fait par la vérification de l'absence de corrélation dans les données, suivie de l'application de la méthode du coude et des sauts d'inertie. Le choix définitif est pris à l'aide du dendrogramme. Finalement, la méthode des k-moyennes est appliquée à la table pour réaliser une segmentation des cartes-jours en six groupes ($k = 6$).

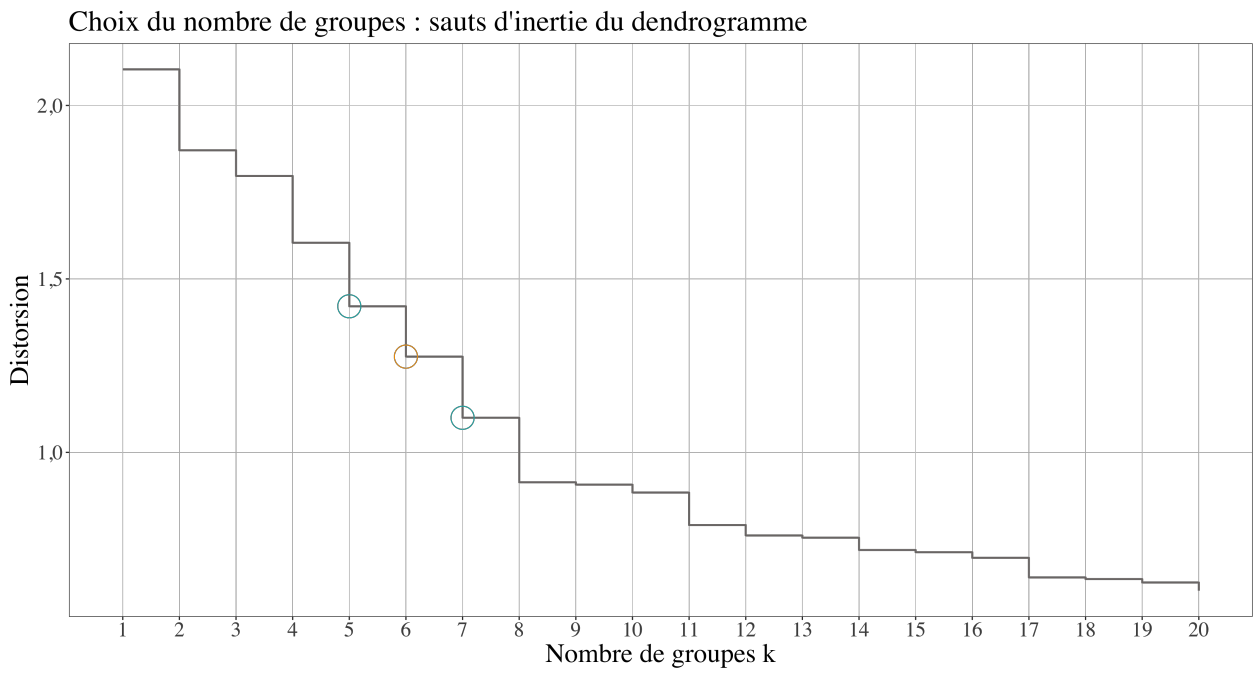


Figure 3.11 Analyse des sauts d'inertie du dendrogramme

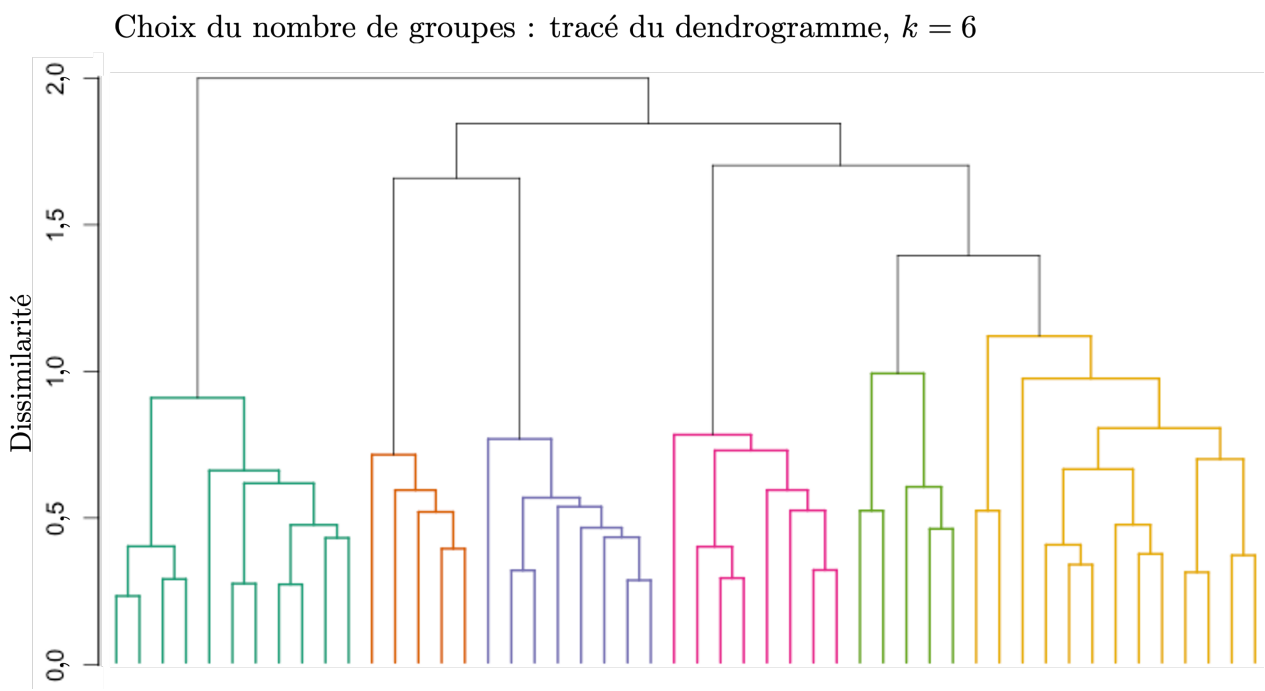


Figure 3.12 Dendrogramme avec $k = 6$ groupes

3.7.3 Segmentation des cartes-jours

Le processus de segmentation commence par le retrait des cartes-jours sans déplacement de la table. Le `seed` (`http://rfunction.com/archives/62`) est ensuite fixé afin de garantir la reproductibilité des résultats. Les cartes-jours avec au moins un déplacement sont ainsi segmentées par la méthode des k-moyennes, avec $k = 6$ groupes, 10 tirages aléatoires des points initiaux et 200 itérations de calcul. Il faut suffisamment d'itérations de calcul pour obtenir une convergence de la composition des groupes et le tirage répété des points de départ permet d'éviter que cette convergence soit vers un optimum local.

Après la segmentation, la table est enrichie en ajoutant à chaque carte-jour le groupe qui lui est associé (numéroté de 1 à 6). Pour plus de lisibilité dans les données, les numéros de groupes sont attribués en fonction de leur centre, de sorte à respecter l'ordre des séries de modes. Pour finir, un septième ensemble portant le numéro 0 est associé aux cartes-jours sans déplacement. Ainsi, les 3 660 000 cartes-jours sont segmentées.

Pour gagner du temps lors des analyses, la table des cartes-jours segmentées est nettoyée et des informations telles que la semaine de l'année et les jours atypiques (annexe F) sont ajoutées. Finalement, la table des cartes-jours segmentées est enregistrée au format .rds et l'analyse des comportements peut commencer.

3.7.4 Analyse des résultats de la segmentation des cartes-jours

La segmentation par les méthodes des k-moyennes fournit une table dans laquelle chaque carte-jour est associée à un groupe. Plusieurs analyses sont alors possibles. En premier, les groupes sont caractérisés. Il s'agit de calculer leurs centres afin de comprendre quel comportement caractéristique est associé à un groupe. La distance entre les vecteurs d'un groupe et leur centre permet également de détailler la variabilité de comportement à l'intérieur de cet ensemble. De plus, la taille des groupes, c'est-à-dire le nombre de cartes-jours contenues dans chaque ensemble, révèle les comportements les plus fréquents chez les voyageurs de l'échantillon.

Ensuite, puisque la segmentation est réalisée sur les cartes-jours, il est possible de décomposer les données en fonction de la date. Cela permet d'observer d'un côté, la fluctuation de composition des groupes et de l'autre, les changements de comportements d'un même usager entre deux jours consécutifs. La variation de composition des groupes est mise en évidence à travers un diagramme à moustaches (de l'anglais *box-plot*). Ce diagramme illustre la taille moyenne d'un ensemble sur une journée ainsi que ses fluctuations habituelles. De même, une

matrice de probabilité de passage d'un groupe à un autre est calculée. Elle traduit l'attraction entre deux ensembles, ce qui permet de décrire les changements de comportements typiques des usagers de l'échantillon.

Pour donner suite à l'analyse de la fluctuation de taille des groupes, un regroupement des données en fonction des jours de la semaine est réalisé. Il permet d'étudier la répartition quotidienne des cartes dans les groupes afin d'observer si le comportement des usagers diffère d'un jour à l'autre. L'analyse continue avec le tracé d'un diagramme de Sankey sur une semaine. Il s'agit d'un histogramme représentant à la fois le nombre de cartes dans chaque groupe par jour et les flux de cartes d'un groupe à un autre sur deux jours consécutifs. En plus de décrire la composition des groupes, ce diagramme illustre de quel ensemble proviennent et vers quel ensemble s'en vont les cartes. Ainsi, le Sankey met en évidence la régularité ou la variation de comportement des usagers de l'échantillon sur une semaine.

La dernière analyse consiste à exprimer la façon dont les usagers des groupes utilisent les lignes de bus du réseau. Il s'agit de positionner géographiquement les groupes sur l'Île de Montréal. Pour ce faire, les lignes empruntées par les passagers sont identifiées à l'aide de leur repère géographique défini à la section 3.4. Ensuite, un regroupement des informations permet de compter le nombre de lignes de chaque combinaison utilisées par un groupe. Un passage en proportion des lignes empruntées par les usagers du groupe est alors réalisé afin de tracer un histogramme. Celui-ci met en avant les différentes utilisations des lignes. Comme les lignes favorisées par les passagers varient, il devient possible de déduire quelles sont les régions que privilégient les usagers de chaque ensemble.

3.8 Analyse de l'utilisation des lignes du réseau

L'étude de l'utilisation des lignes du réseau s'inscrit comme un des quatre principaux objectifs du projet. Il s'agit de décrire la façon dont les lignes sont empruntées, à l'échelle du réseau et indépendamment du comportement des usagers. Caractériser l'utilisation des lignes est un moyen de mesurer l'impact de chacune d'entre elles sur le réseau et de détecter celles qui nécessitent une amélioration.

L'étude est décomposée en trois parties. La première étape consiste à estimer l'autonomie de chaque ligne par la mesure de son taux de correspondances (3.8.1). La deuxième vise à identifier les couples de lignes les plus régulièrement utilisés lors des changements (3.8.2) et permet dans la dernière partie de localiser les principaux arrêts et stations où les passagers correspondent (3.8.3).

3.8.1 Indice d'autonomie d'une ligne

Certaines lignes sont impliquées dans beaucoup de correspondances alors que d'autres sont peu utilisées. Il est proposé d'attribuer un indice d'autonomie à chaque ligne. Comme son nom le laisse entendre, il s'agit d'un indicateur qui reflète l'indépendance d'une ligne sur le réseau. Plus la ligne génère de correspondances, moins son autonomie est grande. Autrement dit, cet indice permet de détecter les lignes qui causent le plus de correspondances.

L'indice d'autonomie repose sur le calcul du taux de correspondances d'une ligne, c'est-à-dire la proportion des déplacements enregistrés sur la ligne qui incluent une correspondance. Pour son calcul, un programme basé sur la confrontation des validations successives compte simultanément le nombre total de déplacements (N) empruntant la ligne et le nombre de déplacements qui se font avec une correspondance. Cependant, le type varie selon les modes et la ligne peut à la fois générer et recevoir des correspondances. Ainsi, neuf formes de correspondances sont utilisées pour le comptage :

- Direct (N_d) : la ligne est utilisée sans correspondance;
- BB génératrice (N_{bb_g}) : la ligne de bus génère une correspondance BB;
- BB réceptrice (N_{bb_r}) : la ligne de bus reçoit une correspondance BB;
- BM génératrice (N_{bm_g}) : la ligne de bus génère une correspondance BM;
- BM réceptrice (N_{bm_r}) : la ligne de métro reçoit une correspondance BM;
- MB génératrice (N_{mb_g}) : la ligne de métro génère une correspondance MB;
- MB réceptrice (N_{mb_r}) : la ligne de bus reçoit une correspondance MB;
- MM génératrice (N_{mm_g}) : la ligne de métro génère une correspondance MM;
- MM réceptrice (N_{mm_r}) : la ligne de métro reçoit une correspondance MM.

Le taux de chaque forme est obtenu en divisant le nombre de déplacements de celle-ci par le nombre total de déplacements (3.2). Il est possible de calculer le taux de déplacements avec une correspondance (3.3) ainsi que le taux de correspondances générées (3.4) et le taux de correspondances réceptionnées (3.5) par la ligne. Finalement, le taux de chaque type de correspondance est calculé (3.6). Il faut noter que sur une ligne de bus $N_{bm_r} = N_{mb_g} = N_{mm} = 0$. À l'inverse, sur une ligne de métro $N_{bb} = N_{bm_g} = N_{mb_r} = 0$. Si lors d'un déplacement la ligne est à la fois génératrice et réceptrice avec des correspondances du même mode, alors ce déplacement est comptabilisé deux fois.

$$\text{Taux de la forme de correspondance } i : \quad \tau_i = \frac{N_i}{N} \tag{3.2}$$

$$\text{Taux de correspondances de la ligne : } \quad \tau_c = 1 - \frac{N_{td}}{N} \tag{3.3}$$

$$\text{Taux de correspondances générées : } \quad \tau_g = \frac{N_{bb_g} + N_{bm_g} + N_{mb_g} + N_{mm_g}}{N} \tag{3.4}$$

$$\text{Taux de correspondances réceptionnées : } \quad \tau_r = \frac{N_{bb_r} + N_{bm_r} + N_{mb_r} + N_{mm_r}}{N} \tag{3.5}$$

$$\text{Taux de correspondances de type } j : \quad \tau_j = \frac{N_{j_g} + N_{j_r}}{N} \tag{3.6}$$

Finalement, il est proposé de calculer le taux de correspondances généralisé de la ligne (τ_l), qui est la moyenne des taux de chaque type de correspondances (3.7). La moyenne prend une division par cinq et non par six puisqu'en fonction de la ligne, soit $\tau_{bb} = 0$, soit $\tau_{mm} = 0$. Pour terminer, l'indice d'autonomie de la ligne I_l est déduit du taux de correspondances généralisé (3.8). Ainsi, moins la ligne génère de correspondances, plus son autonomie est grande.

$$\text{Taux de correspondances généralisé : } \quad \tau_l = \frac{\tau_g + \tau_r + \tau_{bb} + \tau_{bm} + \tau_{mb} + \tau_{mm}}{5} \tag{3.7}$$

$$\text{Indice d'autonomie de la ligne } l : \quad I_l = 1 - \tau_l \quad ; \quad I_l \in [\,0\,;1\,] \tag{3.8}$$

Puisque le taux de correspondances généralisé d'une ligne est calculé à l'aide d'une division par le nombre de déplacements effectués, l'indice d'autonomie est un indicateur indépendant de l'achalandage des lignes. Il est basé seulement sur la présence des correspondances.

Une fois le calcul de l'indice d'autonomie effectué, trois analyses sont réalisées. La première consiste à étudier la répartition de l'indice en fonction de la classification et de la géolocalisation des lignes. Il s'agit de montrer les impacts qu'ont le service offert et la couverture d'une ligne sur son taux de correspondances. Ensuite, une matrice de corrélation est calculée afin d'identifier quelles sont les variables de la typologie des lignes qui ont un impact significatif sur l'indice d'autonomie. Enfin, les variables identifiées par la matrice de corrélation sont utilisées pour faire une régression linéaire multiple. Celle-ci vise à prouver que l'indice d'autonomie d'une ligne peut être prédit à partir de sa typologie, sans utilisation de données opérationnelles. Il est important de préciser que l'indice d'autonomie d'une ligne est calculé à partir des données de l'échantillon de cartes à puce. Il n'est donc pas nécessairement représentatif de l'utilisation réelle du réseau de la STM et l'application de la méthode sur un autre échantillon pourrait donner des résultats différents.

3.8.2 Couples de lignes liées par une correspondance

L'indicateur d'autonomie d'une ligne exprime la façon dont elle est utilisée pour correspondre. Néanmoins, cette représentation n'apporte pas d'information sur les autres lignes impliquées. Comme une correspondance inclue toujours deux lignes distinctes, un outil visuel est développé pour identifier les couples de lignes qui sont les plus souvent jumelés. L'objectif est de détecter les liens du réseau qui nécessitent un intérêt particulier.

En utilisant de nouveau une comparaison des transactions successives, les deux lignes résultant d'une correspondance sont repérées puis une agrégation de la table permet de compter le nombre de changements qui se produisent entre chaque paire, selon l'échantillon. Une table comme celle du tableau 3.10 est alors générée.

Tableau 3.10 Table des correspondances entre chaque paire de lignes

Date	Ligne 1	Ligne 2	Nombre de correspondances	Type de correspondance
2016-01-15	1	2	1805	mm
2016-01-15	1	67	88	mb
2016-01-15	51	2	91	mb
2016-01-15	165	161	34	bb

Des représentations visuelles sous forme de graphes sont réalisées. Chaque ligne est associée à un sommet et les interactions entre les lignes sont représentées par des arêtes. Pour améliorer la lisibilité, les graphes sont séparés en fonction du type de correspondance. De même, afin de rendre l'outil évolutif, le programme permet de sélectionner diverses périodes d'agrégation. Par exemple, un graphe peut tracer toutes les interactions de l'année, la moyenne des interactions quotidiennes, une journée ou une semaine en particulier, etc. Malgré cela, les graphes sont illisibles si les 224 lignes de la STM ont un sommet. C'est pourquoi par défaut le programme trace uniquement les arêtes les plus denses et leurs sommets associés. Ainsi, une ligne très achalandée mais avec un faible taux de correspondances n'est pas affichée dans les graphes. L'utilisateur du programme sélectionne manuellement le nombre d'arêtes à afficher et il peut si nécessaire désactiver la sélection des liens majoritaires.

Les graphes suivent une nomenclature rigoureuse afin de permettre une analyse plus fine. Les sommets sont définis par un numéro, une couleur et une taille. Le numéro est celui de la ligne et la couleur s'accorde avec celle attribuée par la STM pour chaque typologie (figure 3.13). La taille du sommet réfère au taux de correspondances de la ligne en fonction des types

étudiés. Si le graphe décrit les interactions BB, alors la taille du sommet est définie par le taux de correspondances BB τ_{bb} de la ligne, etc. Si plusieurs types de correspondances sont représentés dans le graphe, alors la taille des sommets se fait en fonction du taux de correspondances généralisé τ_l. De même, les arêtes sont définies par une couleur, un nombre et une taille. Chaque type de correspondance est associé à une couleur distincte (figure 3.13), afin d'être différencié facilement sur les graphes. Le nombre et la taille réfèrent à la quantité de correspondances entre les deux sommets, selon l'agrégation mise en place. La taille des arêtes permet de repérer rapidement les liens denses.

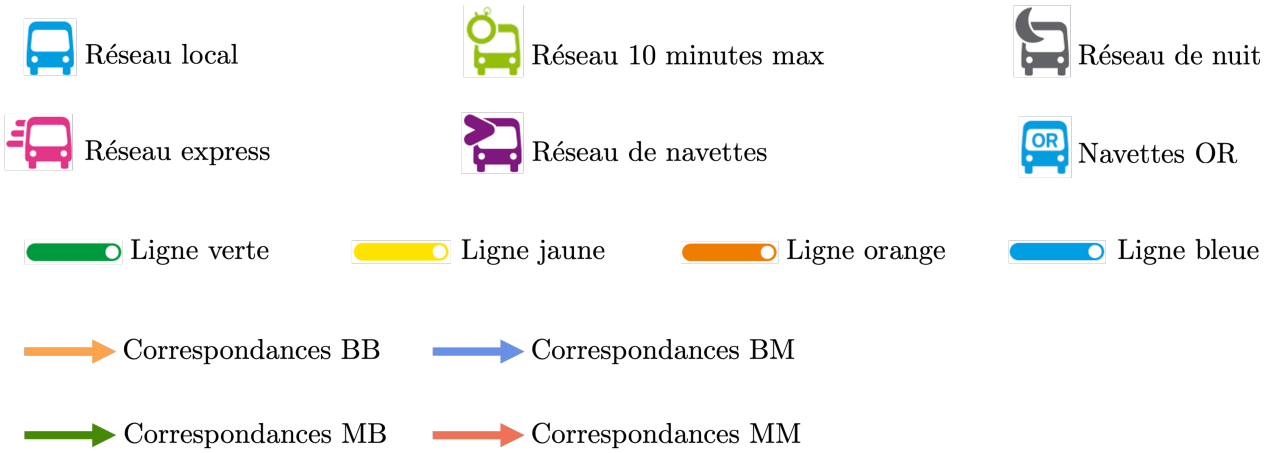


Figure 3.13 Nomenclature des graphes d'interactions des lignes de la STM

Finalement, les graphes permettent d'identifier les lignes les plus fréquemment utilisées par les usagers de l'échantillon pour faire des correspondances. En utilisant les couples principaux, il est alors possible de repérer les stations de métro et les arrêts de bus qui sont impliqués dans les changements.

3.8.3 Principaux arrêts et stations de correspondance

L'objectif de cette dernière partie est de localiser les nœuds du réseau où il se produit le plus de correspondances. Cela permet simultanément de comprendre où les usagers favorisent des changements et d'identifier les lieux du réseau qui nécessitent une attention particulière. Effectivement, repérer les arrêts où les correspondances sont nombreuses sert à déterminer où ajouter des infrastructures plus sécurisantes et facilitant les transferts. De même, cela donne des indications pour améliorer l'offre de transport en commun.

Grâce aux successions d'étapes de l'algorithme VALIDEP, les correspondances BM, MB et MM peuvent être localisées. Elles ont toutes lieu dans une station de métro et par conséquent, il est possible de connaître le nombre de transferts se produisant dans chacune d'entre elles. Un regroupement des informations fournit alors la proportion quotidienne moyenne de correspondances de chaque type ayant lieu dans toutes les stations du réseau. Afin de faciliter l'analyse, ces proportions sont représentées sur une carte. Les stations les plus utilisées pour correspondre, selon l'échantillon initial de données, sont ainsi révélées.

Néanmoins, l'algorithme VALIDEP ne peut pas de localiser les correspondances BB, à la fois car les possibilités de lieux sont trop nombreuses et, car il n'y a pas toujours un seul croisement entre deux lignes de bus. L'identification des arrêts où il y a le plus de correspondances BB se fait alors manuellement, à l'aide du graphe obtenu précédemment. Sur une carte, le croisement des principaux couples de lignes est étudié et l'arrêt est marqué. Cette technique sert à représenter les lieux des principaux échanges, mais ne permet pas d'afficher la proportion de correspondances sur chaque arrêt du réseau. Malgré tout, c'est un moyen d'identifier les arrêts nécessitant une attention particulière.

CHAPITRE 4 EXPÉRIMENTATIONS ET RÉSULTATS

Le chapitre 3 définit le cadre théorique du projet. Pour y donner suite, les expérimentations sont réalisées. Il s'agit de l'application de la méthodologie sur un échantillon de données de cartes à puce fournies par la Société de transport de Montréal (STM) sur une année. Ce chapitre commence par l'affichage des résultats de l'algorithme VALIDEP et poursuit par l'analyse descriptive des déplacements déduits. Ensuite, la démarche méthodologique est suivie pour effectuer successivement les analyses des déplacements quotidiens, du comportement des usagers et de l'utilisation des lignes. Les expérimentations sont appliquées à un échantillon de données de cartes à puce dans l'objectif de valider les méthodes proposées. Elles ne permettent pas de qualifier de manière représentative l'utilisation du réseau de la STM.

4.1 Analyse descriptive des données

4.1.1 Résultats de l'algorithme VALIDEP

L'échantillon des données de transactions par cartes à puce fournies par la STM est composé de 10 000 cartes OPUS avec des abonnements mensuels et annuels entre le 1er janvier et le 31 décembre 2016. Ces cartes sont sélectionnées aléatoirement et par conséquent, elles ne sont pas nécessairement représentatives des usagers de la STM. L'étude couvre une période de 366 jours puisque 2016 est une année bissextile. Cet échantillon compte 9 550 725 transactions.

Pour commencer l'expérimentation, l'algorithme VALIDEP est exécuté sur le jeu de données. L'ensemble du projet est compilé sur un unique ordinateur doté d'un processeur `Intel Core i5` 6360 `U CPU` 2,3 `GHz` et `16 Go` de mémoire vive. Avec cette machine, le temps d'exécution de l'algorithme VALIDEP sur le jeu de données est inférieur à deux heures. L'étape la plus consommatrice est le regroupement des transactions en déplacements, puisque beaucoup d'attributs sont générés. Les autres programmes compilent en quelques minutes chacun.

L'algorithme VALIDEP réussit à traiter 97,48 % des données, fournissant une nouvelle table composée de 9 309 572 transactions. Les données qui ne sont pas prises en compte par le modèle (2,52 %) sont réparties entre des transactions sur des lignes inexistantes et des transactions en métro dont la sortie ne peut pas être déduite, respectivement 1,55 % et 0,97 %.

Le regroupement des 9 309 572 transactions par VALIDEP produit une table de 5 887 074 déplacements. Cela constitue une moyenne de 1,61 déplacement par carte et par jour, soit

environ 589 déplacements par an et par usager. De plus, 56,64 % des déplacements sont faits avec au moins une correspondance. L'algorithme VALIDEP détecte un total de 4 983 278 correspondances : 988 853 bus – bus (BB), 1 215 452 bus – métro (BM), 1 215 459 métro – bus (MB) et 1 563 514 métro – métro (MM).

4.1.2 Distribution annuelle des déplacements

L'analyse descriptive des données de l'échantillon commence par le tracé de l'achalandage. La figure 4.1 représente ainsi la fluctuation du nombre de déplacements effectués chaque jour par les usagers de l'échantillon étudié. La répétition des semaines est bien visible, marquée par une baisse du nombre de déplacements les samedis et les dimanches. Cette diminution de l'achalandage en fin de semaine s'explique par l'absence de travail pour la majorité des usagers. Les déplacements sont alors motivés par du loisir et du magasinage, ce qui en génère moins. À l'inverse, les jours ouvrés sont similaires et avec beaucoup de déplacements. Toutefois, il y a moins d'achalandage les lundis et les vendredis que les autres jours de semaines. Finalement, une saisonnalité s'observe sur l'ensemble de l'année. Il y a un fort achalandage entre janvier et mai, puis une diminution jusqu'à un minimum en août, suivi d'une hausse entre septembre et décembre.

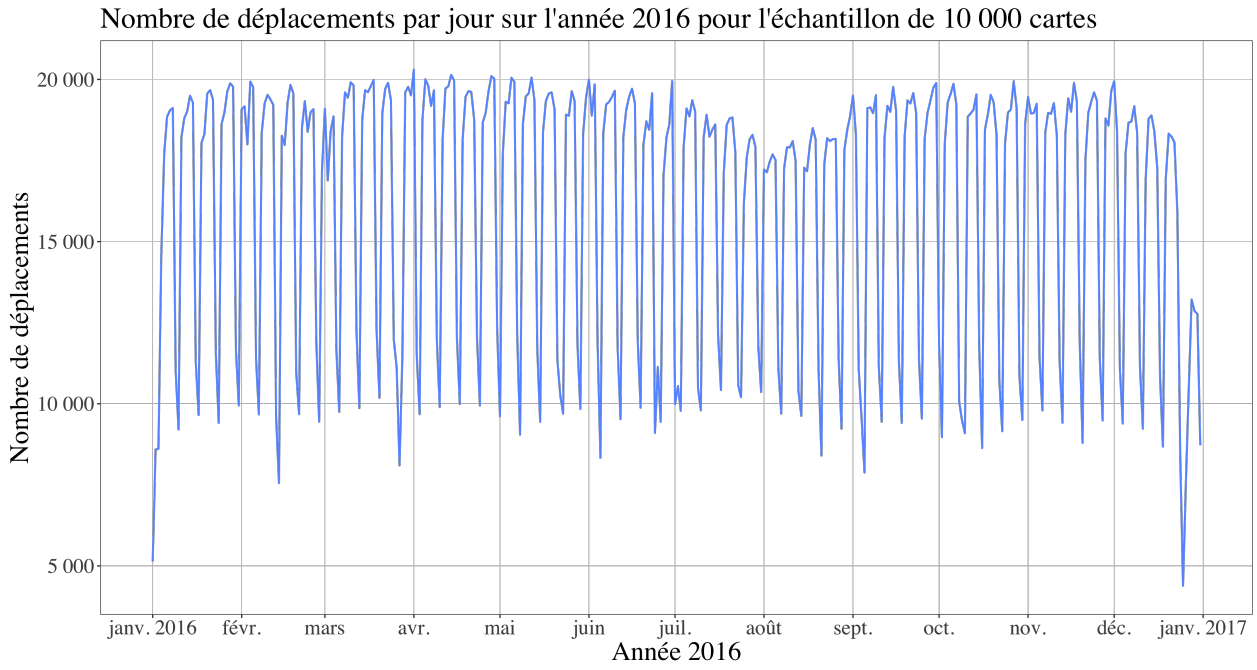


Figure 4.1 Distribution annuelle des déplacements

La fluctuation de l'achalandage obtenue avec l'algorithme VALIDEP s'accorde avec le calendrier de travail de la population active et étudiante de Montréal. Cependant, pour continuer les analyses, il est essentiel de vérifier la cohérence des résultats de l'algorithme en comparaison avec l'intégralité des déplacements du réseau. Pour ce faire, la figure 4.3 représente la proportion de tous les déplacements des cartes à puce enregistrés par la STM en 2016 qui sont contenus dans l'échantillon de données. Premièrement, l'échantillon de données inclut en moyenne 2,10 % de tous les déplacements. Sur l'ensemble de l'année, l'algorithme VALIDEP interprète au minimum 1,67 % et au maximum 2,65 % des déplacements quotidiens. Cela justifie le bon fonctionnement de l'algorithme puisque l'échantillon de départ est constitué de 2,17 % des transactions. Ensuite, la proportion des déplacements constituant l'échantillon est plus grande les fins de semaine et durant l'été. Cela s'explique par le fait que les cartes à puce choisies pour l'étude sont celles qui ont un abonnement. Ce sont les cartes les plus régulières. Dans les périodes creuses, elles enregistrent plus de validations que les autres cartes.

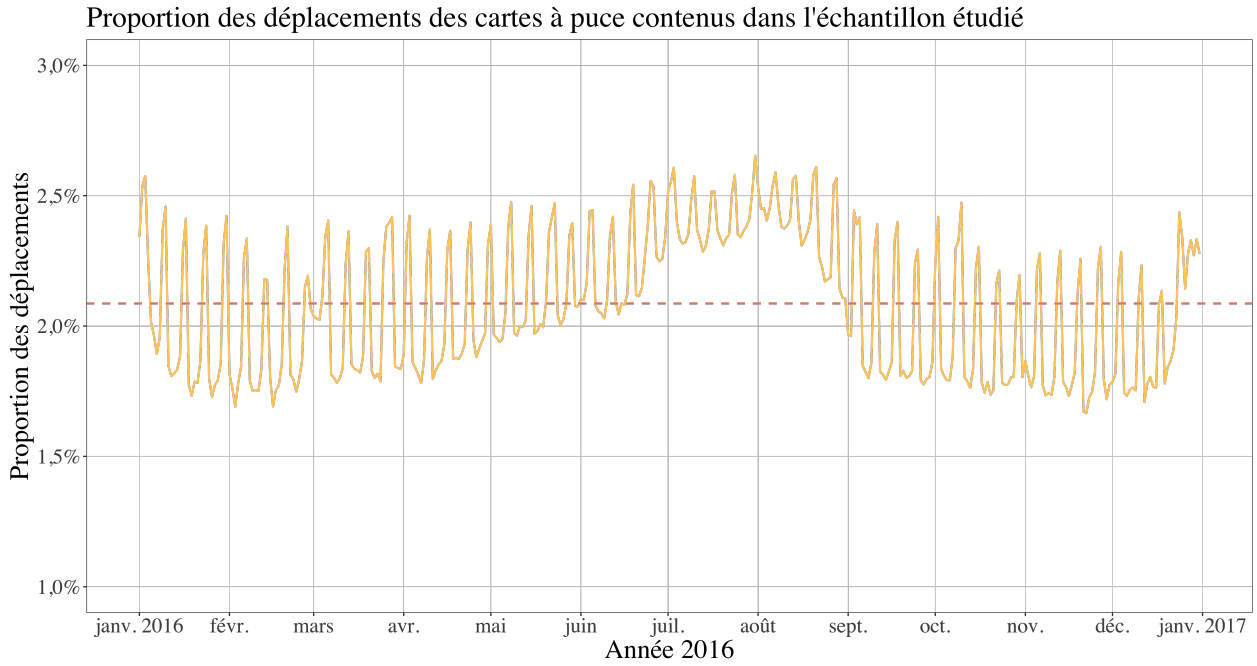


Figure 4.2 Proportion des déplacements contenus dans l'échantillon étudié

4.1.3 Distribution quotidienne des déplacements

De même que pour la distribution annuelle des déplacements, la figure 4.3 représente l'achalandage moyen chaque heure de la journée en fonction du type de jour. Avant tout, il faut noter que le calcul de ces moyennes ne prend pas en compte les jours atypiques de l'annexe F. La distribution observée suit la tendance communiquée par la STM [65], ce qui confirme la cohérence des résultats fournis par l'algorithme VALIDEP.

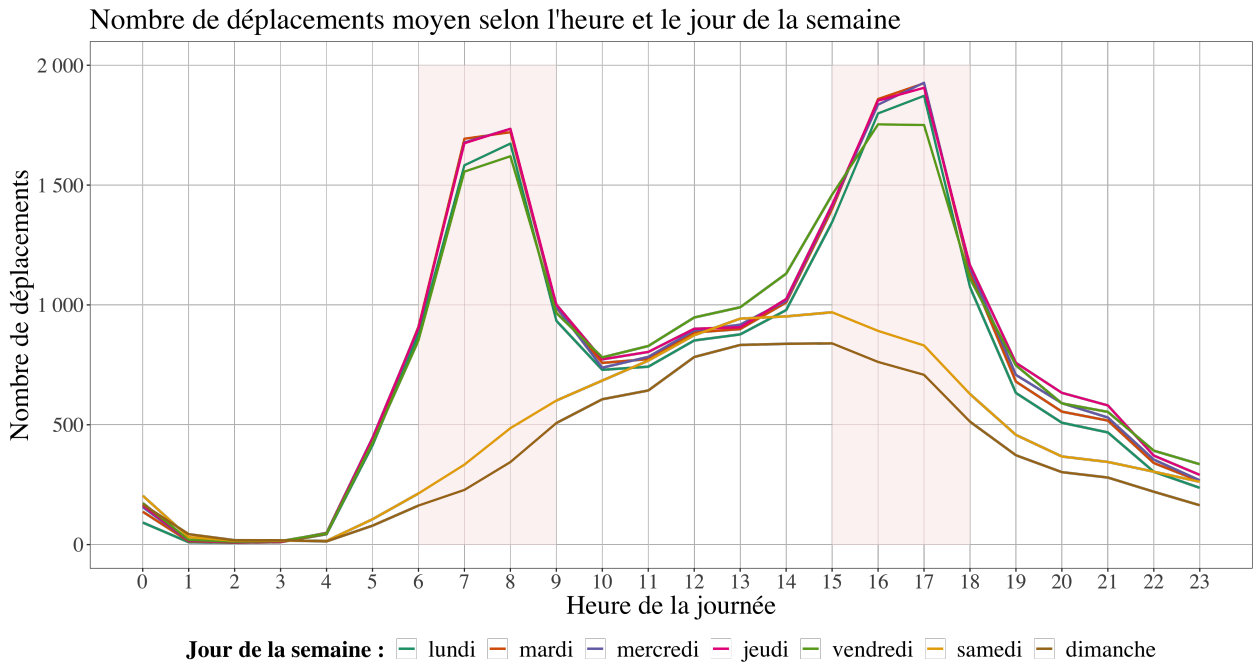


Figure 4.3 Distribution quotidienne des déplacements

Premièrement, la répartition de l'achalandage met en avant une différence majeure entre les jours ouvrés et ceux de fins de semaine. Lors des jours ouvrés, deux pointes sont visibles, traduites par une forte augmentation du nombre de déplacements. La pointe du matin est observée entre 6:00 et 9:00 tandis que celle du soir l'est entre 15:00 et 18:00. Ces pointes traduisent des déplacements pour le travail et les études. Le soir, la pointe est légèrement plus intense que le matin, avec un maximum d'achalandage atteint aux alentours de 17:00. À l'inverse, pendant les fins de semaine il n'y a pas de pointe visible, l'achalandage est croissant jusqu'à 15:00 puis diminue ensuite. Toutefois, la tendance d'achalandage observée le samedi suit celles des jours ouvrés entre 10:00 et 14:00. Le dimanche, il y a moins de déplacements.

Une différence s'observe également entre les jours ouvrés. Les pointes du lundi et du vendredi sont moins intenses que celles des autres jours. De plus, entre 10:00 et 14:00, l'achalandage

du lundi est le plus faible alors que celui du vendredi est le plus élevé. Cette différence peut s'expliquer par la proximité du lundi et du vendredi avec la fin de semaine. En effet, ne pas travailler ces jours permet d'avoir un week-end prolongé. De même, l'achalandage élevé plus tôt le vendredi peut s'expliquer par des usagers qui ne travaillent pas le vendredi après-midi.

4.2 Composition des déplacements quotidiens

La première analyse que propose la méthodologie est l'étude de la composition des déplacements. L'objectif est de décrire comment les usagers de l'échantillon ont recours aux correspondances pour leurs trajets quotidiens. Pour cela, la répartition des séries de modes est d'abord étudiée afin d'observer comment sont utilisées les correspondances pendant la journée. Elle est suivie par l'observation des écarts de temps entre deux activités générées par le transport en commun (TC) et termine par l'analyse de la symétrie des déplacements pendulaires.

4.2.1 Répartition des correspondances

L'étude débute par le tracé de la répartition des séries de modes selon l'heure, visible à la figure 4.4. Pour cette représentation, les journées sont regroupées en deux groupes découlant de la figure 4.3, les jours ouvrés et les jours de fins de semaine. La répartition sert à décrire quelles sont les séries de modes favorisées par les usagers de l'échantillon selon les heures de la journée. De manière générale, il y a 22,00 % de déplacements directs en bus et 21,34 % en métro. Les déplacements avec un seul mode sont répartis comme suit : 33,04 % en bus et 32,51 % en métro. Ainsi, sur l'échantillon étudié, le métro est autant utilisé que le bus. Toutefois, il est important de noter que cette recherche ne permet pas de déduire la préférence des usagers. L'expression « favoriser » fait référence à une utilisation du réseau orientée vers un type de déplacement en particulier, mais il n'est pas possible de déduire si les usagers le font par choix pour par obligation.

Plusieurs observations sont possibles à partir de la figure 4.4. Premièrement, la répartition des séries de modes diffère entre la nuit et le reste de la journée. Ceci est normal puisque le métro n'est pas nocturne à Montréal. La nuit, les usagers ne peuvent pas utiliser le métro donc seules les séries de modes avec du bus sont présentes. La part des trajets directs est élevée à la fois, car les alternatives de transport sont moins nombreuses et, car la fréquence des bus est plus faible qu'en journée. Certains usagers de l'échantillon font tout de même des correspondances. Comme le métro est inaccessible, la part de la série “B,B” augmente

et la série “Autre” traduit des déplacements avec aux moins deux correspondances en bus. De manière générale, entre 23:00 et 4:00, les déplacements directs sont majoritaires. À partir de 4:00, la part des déplacements des séries “B,M”, “B,M,M” et “Autre” croît. Ce sont les usagers qui prennent le bus pour correspondre avec le métro qui ouvre un peu plus tard.

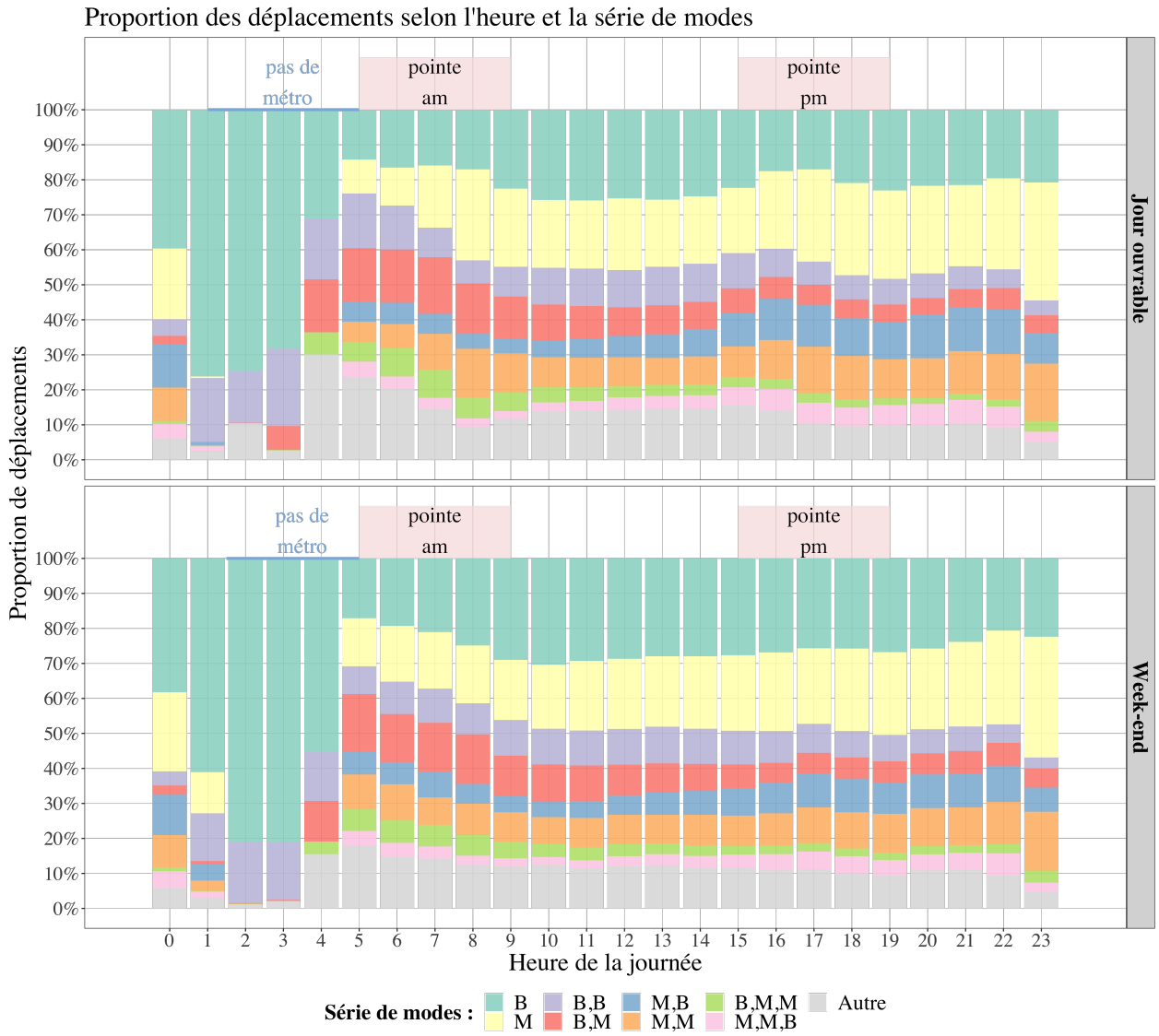


Figure 4.4 Répartition des déplacements selon la série de modes

Au cours de la journée, les déplacements avec des correspondances sont majoritaires, en moyenne 55,53 % des déplacements de l’échantillon. Cependant, il s’agit principalement de trajets ayant une seule correspondance, 35,52 %. La faible part de déplacements avec au moins deux correspondances laisse supposer que cet ensemble d’usagers défavorise les itinéraires qui nécessitent plusieurs changements. La part des déplacements de la série de modes “Autre”

augmente pendant la pointe du matin puis se stabilise durant la journée aux alentours de 12 %. Cela montre que dans cet échantillon, les usagers favorisent des déplacements avec plusieurs correspondances surtout pendant les heures de pointe.

La répartition des séries de modes montre que l'utilisation du réseau change pendant les pointes. En effet, durant les heures de pointe, il y a une augmentation de la part des déplacements qui incluent une correspondance. Inversement, hors des pointes les trajets directs sont favorisés. La proportion des déplacements "B" et "B,B" est forte tant que le métro est fermé. Pendant la pointe du matin, elle diminue puis croît à nouveau entre 10:00 et 14:00. Pendant la pointe du soir, elle diminue encore et elle se stabilise après 19:00. La proportion des déplacements "M" et "M,M" suit la tendance inverse. Il y a une augmentation de la part de ces déplacements pendant les pointes et une diminution en dehors. Autrement dit, dans l'échantillon utilisé, le métro est favorisé pendant les heures de pointe alors que le bus est favorisé pendant les heures creuses.

Les déplacements "B,M" et "B,M,M" voient leur proportion augmenter pendant la pointe du matin, avec un maximum atteint aux environs de 7:00. Leur proportion baisse continuellement après la pointe du matin. Les déplacements "M,B" et "M,M,B" suivent la tendance inverse, avec une augmentation continuelle de leur proportion jusqu'à atteindre le maximum pendant la pointe du soir, aux alentours de 17:00. L'inversion des tendances de ces séries de modes est illustrée par les figures 4.5 et 4.6.

L'inversion des proportions des séries de modes pendant la journée laisse penser qu'il s'agit de déplacements pendulaires. Pour se rendre à leurs activités, les usagers de l'échantillon suivent un itinéraire le matin et font l'itinéraire inverse le soir. L'étude des durées entre les déplacements et de la symétrie des séries vise à confirmer cette hypothèse.

Pour terminer, la répartition des déplacements de la fin de semaine ressemble à celle des jours ouvrés, mais les périodes de pointe sont plus étendues et avec moins de différences. La fin de semaine, le bus est plus utilisé et la proportion des déplacements avec une correspondance est plus faible. En conclusion, dans l'échantillon étudié, l'analyse révèle que le bus est un mode de transport favorisé principalement hors pointe alors que le métro est favorisé principalement en pointe. De même, elle souligne que les usagers de cet ensemble ont tendance à faire plus de correspondances pendant les heures de pointe, ce qui pourrait traduire des déplacements contraints pour le travail.

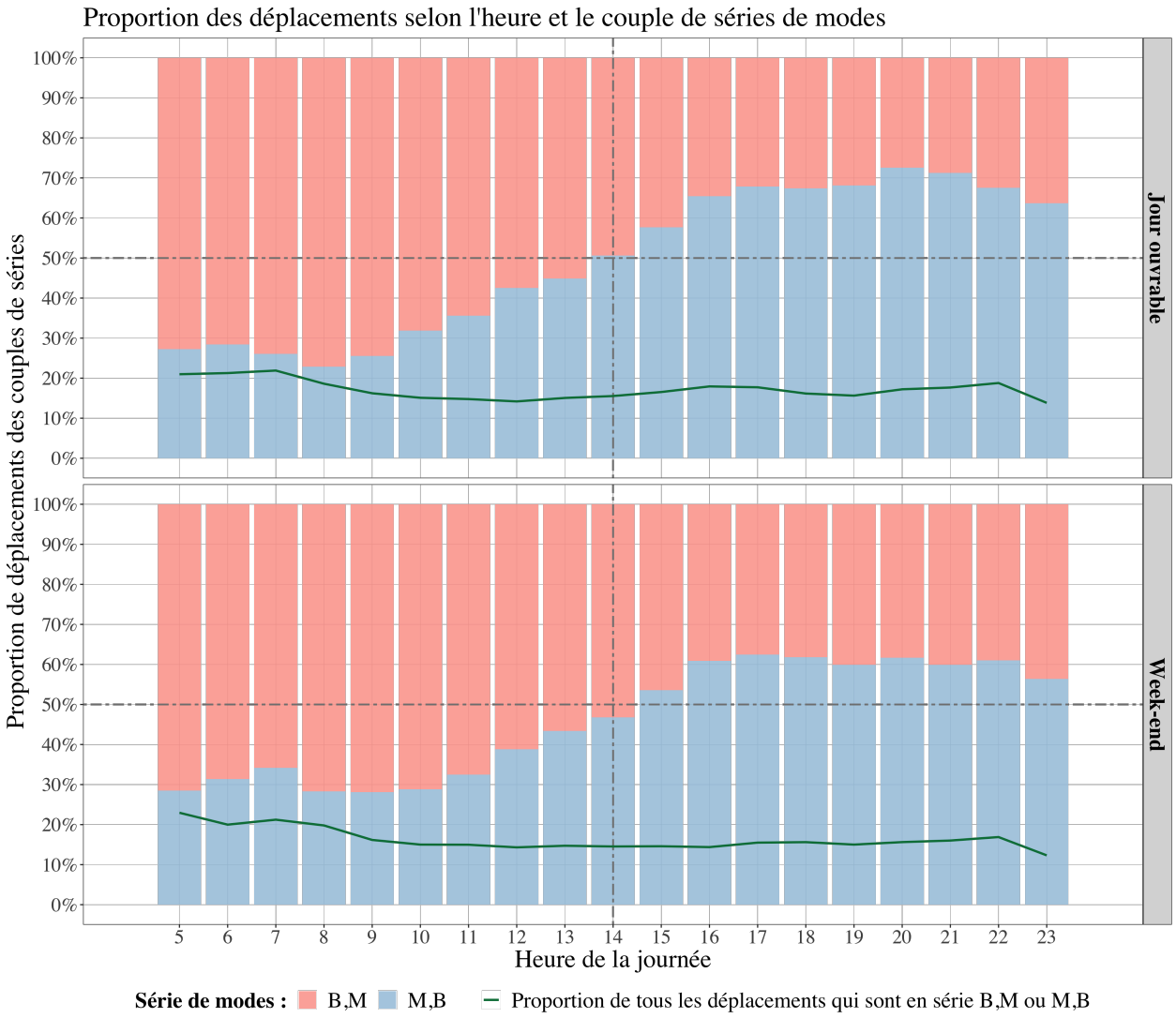


Figure 4.5 Répartition des déplacements des séries “B,M” et “M,B”

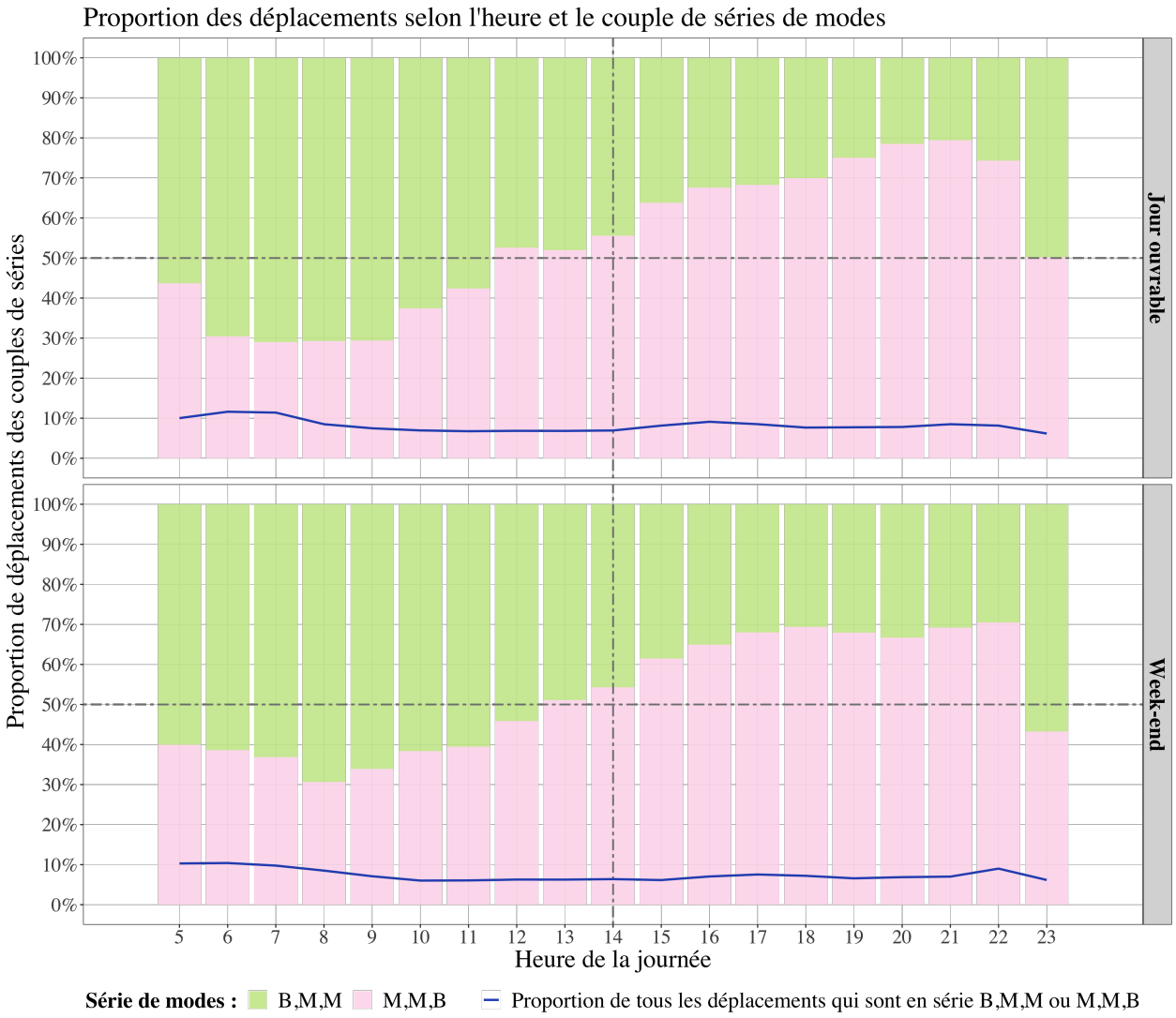


Figure 4.6 Répartition des déplacements des séries “B,M,M” et “M,M,B”

4.2.2 Écart de temps entre deux déplacements

Les données de déplacements sont maintenant agrégées en cartes-jours. Comme l'étude porte sur un échantillon aléatoire de 10 000 cartes à puce et 366 jours, il y a 3 660 000 cartes-jours. Certaines cartes-jours enregistrent plusieurs déplacements alors que d'autres n'en font pas. La figure 4.7 représente la proportion des cartes-jours en fonction du nombre de déplacements quotidiens.

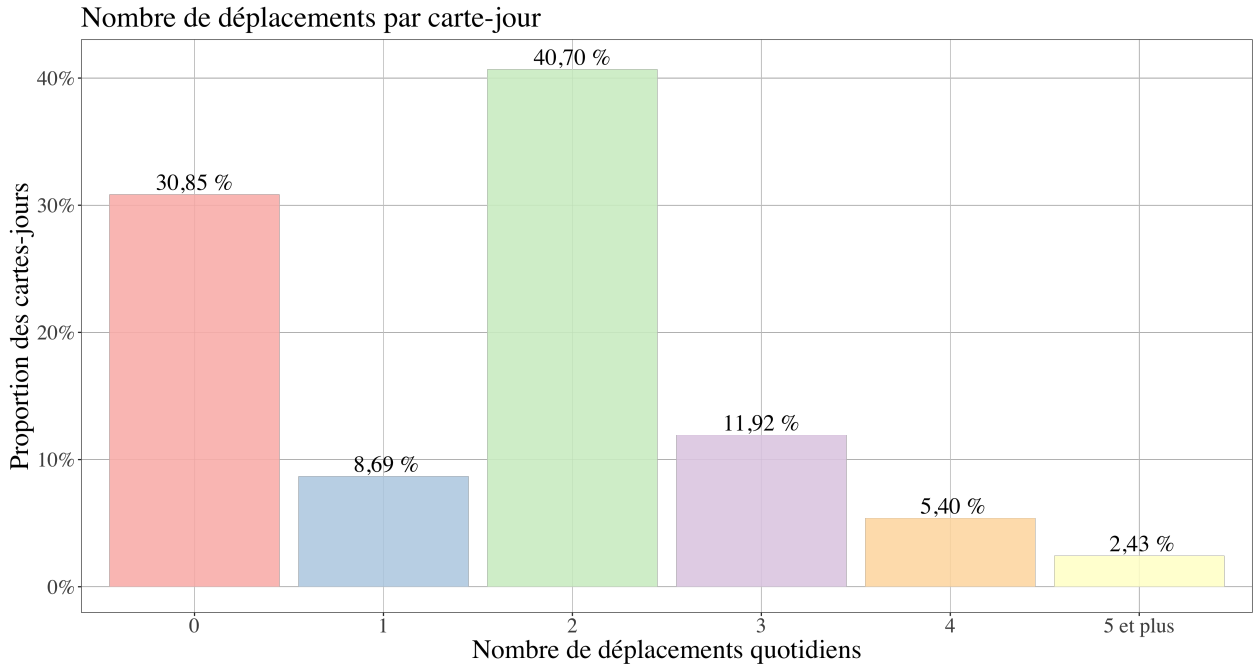


Figure 4.7 Nombre de déplacements par carte-jour

Tout d'abord, la répartition montre que pour 30,85 % des cartes-jours de l'échantillon, il n'y a pas de déplacement en transport en commun. Cette valeur est en adéquation avec la proportion des jours de fin de semaine dans l'année (28,56 %). L'analyse des données de l'échantillon montre que plus de la moitié des cartes-jours sans déplacement est pendant des jours de fin de semaine. Il s'agit de 15,61 % des cartes-jours.

La plus grande proportion de cartes-jours (40,70 %) concerne les journées avec deux déplacements. Cela peut signifier que le TC est utilisé à l'aller et au retour. Une supposition est que l'usager réalise un premier déplacement pour se rendre sur son lieu d'activité, puis un second pour en revenir. À l'inverse, la proportion de cartes-jours faisant un seul déplacement est faible, 8,69 %. Pour ces usagers, il n'y a pas d'aller-retour en TC. Si l'utilisateur fait un aller-retour, alors un de ses déplacements se fait avec un autre mode de transport.

La proportion des cartes-jours faisant trois déplacements est de 11,92 %, soit plus que les cartes-jours avec un seul déplacement. Il peut s'agir des passagers faisant une chaîne de déplacements ou un aller-retour avec une étape vers une autre activité. Finalement, la proportion d'usagers de l'échantillon faisant quatre déplacements et plus en une journée est faible, 7,83 %. Il pourrait s'agir de voyageurs faisant deux allers-retours ou plus.

Pour mieux visualiser comment sont répartis les déplacements quotidiens, la figure 4.8 représente l'écart de temps écoulé entre deux déplacements consécutifs d'une même carte-jour. Pour ce calcul, seules les cartes-jours avec au moins deux déplacements sont conservées. Cela représente 60,46 % des données, soit 2 212 617 cartes-jours.

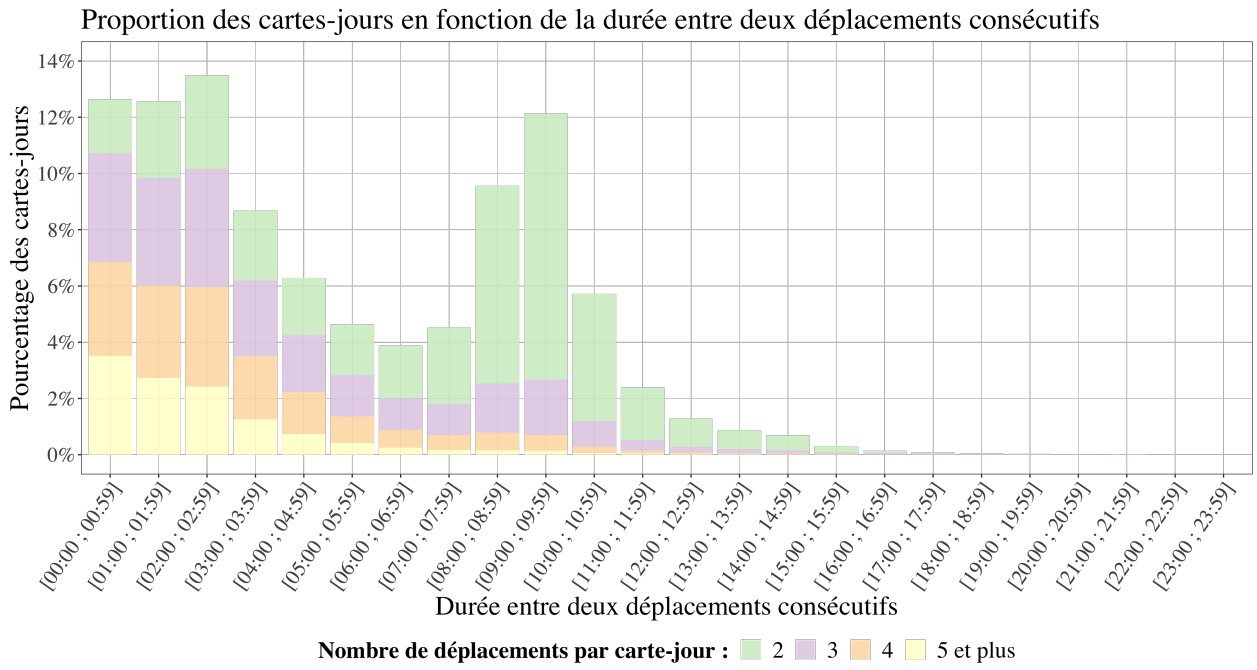


Figure 4.8 Écarts de temps entre deux déplacements consécutifs

Cette représentation permet de décrire comment le TC génère des activités au sein de l'échantillon d'usagers. Deux pics sont visibles sur la figure 4.8. Le premier montre qu'il y a beaucoup de paires de déplacements séparées par moins de 3h. Le second quant à lui met en avant des durées comprises entre 8h et 10h. Le premier cas signifie que des activités de courtes durées sont générées par le TC. Cependant, à partir des informations disponibles, il est impossible de déterminer le motif des déplacements. Au vu des durées, il peut s'agir d'activités de loisir ou de magasinage. Le deuxième cas signifie que le TC génère également des activités de longues durées. Au Québec, une journée de travail moyenne dure 8h [66]. Par conséquent, le second pic pourrait décrire des déplacements liés à des activités professionnelles. La forte proportion

de cartes-jours avec deux déplacements dans ce pic vient appuyer cette hypothèse. De plus, le nombre de déplacements quotidiens diminue avec l'augmentation de la durée des activités. En moyenne, dans cet échantillon, le temps entre deux déplacements est de 5h20.

Pour confirmer ces observations, la figure 4.9 représente sous forme matricielle la répartition des couples de déplacements. Ainsi, il est possible de visualiser à la fois la durée et la période de la journée de l'activité générée par le TC dans l'échantillon. Pour les activités de longues durées, les départs se produisent majoritairement entre 5:00 et 9:00 et les sorties entre 15:00 et 19:00, ce qui concorde avec les horaires de travail conventionnels. Les activités de courtes durées sont quant à elles principalement réparties entre 9:00 et 17:00.

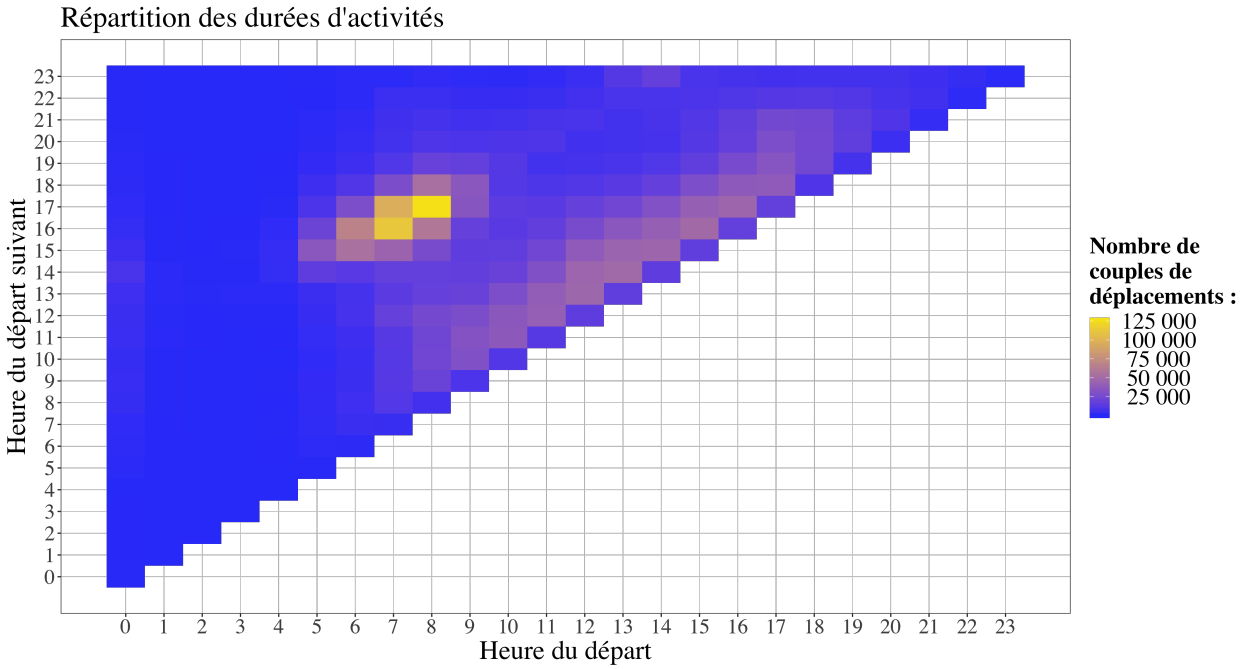


Figure 4.9 Matrice de répartition des durées d'activités

En conclusion, les figures 4.7, 4.8 et 4.9 montrent que dans l'échantillon étudié, les voyageurs utilisent le TC majoritairement pour réaliser deux ou trois déplacements par jour. Ils l'empruntent pour rejoindre à la fois des activités de courtes et de longues durées, réparties tout au long de la journée. Ces observations laissent penser que les usagers de l'échantillon se servent du TC en partie pour faire des allers-retours entre leurs différents lieux d'activités. L'étude de la symétrie des déplacements vise à vérifier cette hypothèse.

4.2.3 Symétrie des déplacements pendulaires

Pour analyser la symétrie des déplacements, seules les cartes-jours de l'échantillon ayant deux trajets compris entre 5:00 et 21:00 les jours ouvrés sont sélectionnées. Il y a ainsi 1 069 054 cartes-jours, soit 29,21 % de la table. L'étude porte ainsi sur 2 138 108 déplacements, c'est-à-dire 39,95 % des données. Cette partie s'intéresse en premier à la symétrie des horaires, puis celles des séries de modes pour finir par la symétrie des lignes utilisées.

Symétrie des horaires de déplacements

L'étude vise à décrire comment les deux déplacements sont positionnés dans la journée. Au sein de l'échantillon, la médiane des heures de départ des déplacements est à 13:23. Il y a autant de départs avant cette heure qu'après. Puisqu'elle coupe la journée en deux parties d'une même durée (8h), elle est utilisée pour distinguer les déplacements du matin et ceux du soir. Ainsi les déplacements qui se produisent avant 13:23 sont des déplacements "am", et les déplacements qui se produisent après 13:23 sont des déplacements "pm".

Il est alors possible de comparer l'horaire du premier et du second déplacement de chaque carte-jour. Pour cela, trois catégories sont possibles :

- am / am : les deux déplacements de la carte-jour sont avant 13:23;
- am / pm : le premier déplacement de la carte-jour est avant 13:23 et le second est après;
- pm / pm : les deux déplacements de la carte-jour sont après 13:23.

La figure 4.10 présente la répartition des durées entre les deux déplacements en fonction de leurs horaires. La plus grande proportion de cartes-jours a une durée entre les déplacements comprise entre 8h et 10h, ce qui est en accord avec les observations apportées par la figure 4.8. La majorité des cartes-jours dont il y a un déplacement le matin et un second l'après-midi est générée par des activités de longues durées. Il peut s'agir de déplacements au motif travail ou étude. Les paires de déplacements qui sont seulement le matin ou seulement l'après-midi sont générées par des activités de courtes durées, moins de 4h pour la majorité d'entre elles.

La proportion des cartes-jours ayant un déplacement dans chaque tranche horaire est calculée et les résultats sont présentés dans le tableau 4.1. Il s'observe une symétrie des horaires des déplacements pour 81,16 % des cartes-jours. Cependant, cette analyse ne permet pas de conclure sur la présence d'allers-retours, c'est pourquoi il faut analyser les séries de modes. Dans l'échantillon, la proportion de cartes-jours faisant deux déplacements le matin est identique à celle des cartes-jours faisant deux déplacements l'après-midi.

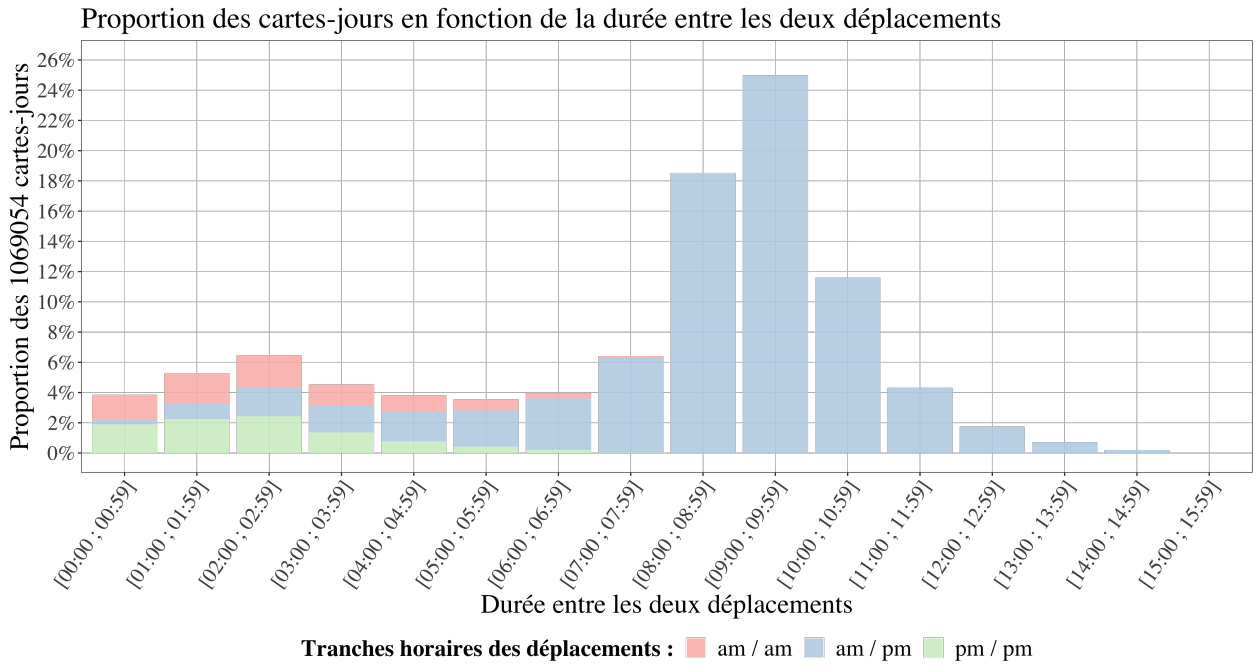


Figure 4.10 Écarts de temps entre deux déplacements en fonction des tranches horaires

Tableau 4.1 Proportion des cartes-jours selon l'heure du premier et du second déplacement

1er déplacement	2e déplacement	Nombre de cartes-jours	Proportion de cartes-jours
am	am	100 756	9,42 %
am	pm	867 658	81,16 %
pm	pm	100 640	9,41 %

Symétrie des séries de modes utilisées

L'application de l'arbre de décision 3.7 permet d'associer chaque carte-jour à une typologie de symétrie puis de comptabiliser l'ensemble des séries de modes de chaque classe. Le résultat de son application sur l'échantillon est donné en fonction des tranches horaires des deux déplacements dans le tableau 4.2.

Tout d'abord, il y a un lien entre la série de modes du premier et du second déplacement pour 87,36 % des cartes-jours. Seulement 12,64 % des cartes-jours n'ont aucun lien entre le premier et le second déplacement. Une symétrie entre la série de modes des deux déplacements est observée dans plus de la majorité des cas, 61,11 %. Deux facteurs peuvent expliquer cette

valeur. La première est qu'il s'agit d'un aller-retour pour une activité où les usagers suivent le même itinéraire dans un sens et dans l'autre. La deuxième est que, dans cet échantillon, les usagers favorisent une série en particulier, peut être à cause de leur connaissance du réseau. Parmi les voyageurs qui utilisent une seule ligne du même mode pour l'aller et le retour (28,66 %), 11,89 % utilisent le bus et 16,77 % le métro. Ainsi, dans l'échantillon étudié, le métro est plus utilisé que le bus pour les déplacements symétriques.

Tableau 4.2 Symétries des séries de modes

Type de symétrie	am / am	am / pm	pm / pm	Part des cartes-jours	Exemple
Asymétrie	1,37 %	9,88 %	1,40 %	12,64 %	
Direct	*1,31 %*	*8,99 %*	*1,28 %*	*11,57 %*	B – M
Correspondance.s	*0,06 %*	*0,89 %*	*0,12 %*	*1,07 %*	B,M – B,B
Ajout	1,25 %	10,34 %	1,13 %	12,72 %	B – B,M
Retrait	1,12 %	10,80 %	1,23 %	13,14 %	B,M – B
Égalité	0,02 %	0,35 %	0,02 %	0,39 %	B,M – B,M
Symétrie	5,65 %	49,79 %	5,65 %	61,11 %	
Direct	*3,23 %*	*21,84 %*	*3,58 %*	*28,66 %*	B – B
Correspondance.s	*2,42 %*	*27,95 %*	*2,07 %*	*32,45 %*	B,M – M,B
Total	9,42 %	81,16 %	9,41 %	100,00 %	

L'étude des symétries montre également qu'une grande partie des usagers de l'échantillon change de série en ajoutant ou en enlevant une ligne. Il y a 12,72 % des cartes-jours qui font une correspondance de plus sur le deuxième déplacement et 13,14 % qui en font une de moins. Cette différence peut être expliquée par un aller-retour utilisant deux itinéraires différents, ou par une succession d'activités. Le tableau 4.3 met en avant les couples de séries de modes les plus utilisés dans le cas des ajouts et des retraits d'une correspondance.

Le tableau 4.3 montre que certains changements sont préférés à d'autres. L'échange entre un déplacement direct et un déplacement avec une correspondance est ce qu'il y a de plus fréquent. En particulier, dans l'échantillon étudié, les usagers semblent favoriser l'ajout d'un bus dans leur série, en faisant ainsi une correspondance vers ou depuis le bus. À l'inverse, lors des retraits, ils favorisent en priorité le retrait d'une correspondance avec un bus. Autrement dit, cette étude montre que sur l'échantillon analysé, il y a plus de variabilité dans l'utilisation du bus que du métro pour les déplacements quotidiens.

Tableau 4.3 Proportion des cartes-jours avec un ajout ou un retrait d'une correspondance

Type d'ajout	Part des ajouts	Part des cartes-jours
B → B,B	16,68 %	2,12 %
M → M,B	16,66 %	2,12 %
B → M,B	3,98 %	0,51 %
M → B,M	3,78 %	0,48 %
B → B,M	3,25 %	0,41 %
B → M,M	2,76 %	0,35 %
M → M,M	2,10 %	0,27 %
M → B,B	0,68 %	0,09 %
Autre	50,10 %	6,37 %
Total	100,00 %	12,72 %

Légende : série du premier déplacement → série du second.

Type de retrait	Proportion des retraits	Part des cartes-jours
B,M → M	22,61 %	2,97 %
B,B → B	16,14 %	2,12 %
M,B → B	5,10 %	0,66 %
B,M → B	3,35 %	0,44 %
M,B → B	2,37 %	0,31 %
M,M → M	2,05 %	0,27 %
M,M → B	1,84 %	0,24 %
B,B → M	0,85 %	0,11 %
Autre	45,71 %	6,01 %
Total	100,00 %	13,14 %

Légende : série du premier déplacement → série du second.

Pour terminer, la répartition des symétries ne semble pas être impactée par les horaires des deux déplacements, puisque la proportion observée au tableau 4.1 est respectée pour chaque typologie de symétrie. L'horaire des déplacements a peu d'impacts sur les choix d'itinéraires des usagers.

Symétrie des lignes de transport empruntées

Pour étudier la symétrie des lignes utilisées, seules les cartes-jours dont les séries de modes sont symétriques ou égales sont sélectionnées. En effet, dans le cas contraire, il ne peut pas y avoir de lien entre les lignes. Ainsi, par déduction de l'étude de la symétrie des séries de modes, 38,89 % des cartes-jours n'ont pas de symétrie dans l'utilisation des lignes alors que 61,11 % des cartes-jours ont une symétrie potentielle.

En appliquant l'arbre de décision 3.8 pour comparer les successions de lignes du premier et du second déplacement de chaque carte-jour, la symétrie de chacune peut être déduite. Les résultats sont formalisés dans le tableau 4.4.

Le tableau 4.4 montre que pour 50,09 % des cartes-jours de l'échantillon, la succession de lignes du deuxième déplacement est l'inverse de celle du premier. En particulier, il y a 26,40 % des cartes-jours qui utilisent une seule ligne pour les deux déplacements et 23,69 % qui en utilisent plusieurs. Cela représente 81,97 % des cartes-jours ayant une symétrie des séries de modes. Cette observation permet de confirmer que dans la majorité des cas observés, les usagers de l'échantillon utilisent le transport en commun pour faire un aller-retour vers leur lieu d'activité.

Tableau 4.4 Symétries des lignes utilisées pour les deux déplacements

Type de symétrie	Nombre de cartes-jours	Part des cartes-jours
Asymétrie	533 488	49,90 %
Direct	*509 315*	*47,64 %*
Correspondance.s	*24 173*	*2,26 %*
Égalité	162	0,02 %
Symétrie	535 404	50,09 %
Direct	*282 183*	*26,40 %*
Correspondance.s	*253 221*	*23,69 %*

Toutefois, il n'y a pas de symétrie observée pour 49,91 % des cartes-jours. Plusieurs suppositions sont possibles. Il peut s'agir de deux déplacements indépendants, avec des objectifs et des lieux différents. Il n'y a donc pas de lien entre le premier et le second itinéraire. Autrement, il peut s'agir d'un aller-retour avec deux itinéraires différents. Puisque les données de cartes à puce ne permettent pas de déterminer l'origine et la destination des déplacements, il n'est pas possible de valider ou de réfuter ces hypothèses.

Finalement, il est possible de regarder quelles sont les lignes les plus utilisées par les usagers de l'échantillon faisant des allers-retours symétriques. C'est la liste que présente le tableau 4.5. Les lignes les plus fréquemment utilisées pour des allers-retours sont des lignes de métro. Ceci est cohérent puisqu'une ligne de métro transporte une plus grande proportion d'usagers et couvre un territoire plus vaste qu'une ligne de bus. Dans l'échantillon utilisé, l'analyse révèle que les lignes de bus les plus utilisées dans les allers-retours sont des lignes du réseau haute fréquence (10 minutes max.). La figure 4.11 illustre la position des lignes du tableau 4.5 sur le territoire de Montréal. Il s'agit de lignes d'axe nord – sud et est – ouest, principalement proches du centre de Montréal et ayant des interactions avec le métro.

Résumé et conclusion de l'analyse des déplacements quotidiens

L'étude de la symétrie d'utilisation des correspondances dans les déplacements quotidiens met en avant des déplacements pendulaires de plusieurs sortes. Tout d'abord, l'exploration des données montre que, dans l'échantillon étudié, la médiane des horaires des déplacements est à 13:23. De ce point, l'analyse des tranches horaires des déplacements rapporte que 81,16 % des cartes-jours font leur premier déplacement le matin et le second l'après-midi. La durée moyenne entre deux activités est de 8h.

Tableau 4.5 Lignes majoritairement impliquées dans les symétries

Ligne.s aller	Ligne.s retour	Part des cartes-jours
Orange	Orange	16,09 %
Verte	Verte	15,56 %
Orange → Verte	Verte → Orange	6,69 %
Verte → Orange	Orange → Verte	6,32 %
Bleue → Orange	Orange → Bleue	2,36 %
Bleue	Bleue	1,65 %
Orange → Bleue	Bleue → Orange	1,57 %
161	161	1,05 %
139	139	0,76 %
141	141	0,73 %
51	51	0,65 %
18	18	0,64 %
24	24	0,62 %
69	69	0,59 %
165	165	0,57 %
Bleue → Orange → Verte	Verte → Orange → Bleue	0,53 %

Ensuite, sur l'ensemble des données étudiées, l'analyse révèle qu'il n'y a pas de lien entre le premier et le second déplacement dans 12,64 % des cas. L'ajout d'une correspondance pour le deuxième déplacement représente 12,72 % des cas et le retrait d'une correspondance 13,14 %. Regarder la composition des ajouts et des retraits permet de conclure que dans cet échantillon, l'usage du bus est plus variable que celui du métro. Enfin, une symétrie d'utilisation des séries est observée dans 61,11 % des cas.

Finalement, l'utilisation des lignes prises par les voyageurs symétriques est analysée. Cette étude démontre que pour 50,09 % des cartes-jours, la succession de lignes du soir est l'inverse de celle du matin. Il s'agit alors d'allers-retours. Dans l'échantillon étudié, les usagers faisant des allers-retours avec le même itinéraire ont tendance à favoriser l'utilisation du métro et des lignes à haute fréquence. Toutefois, cette étude ne permet pas de conclure sur les préférences des usagers. Seules des informations sur la composition des déplacements pendulaires au sein du groupe d'usagers étudié peuvent être obtenues.

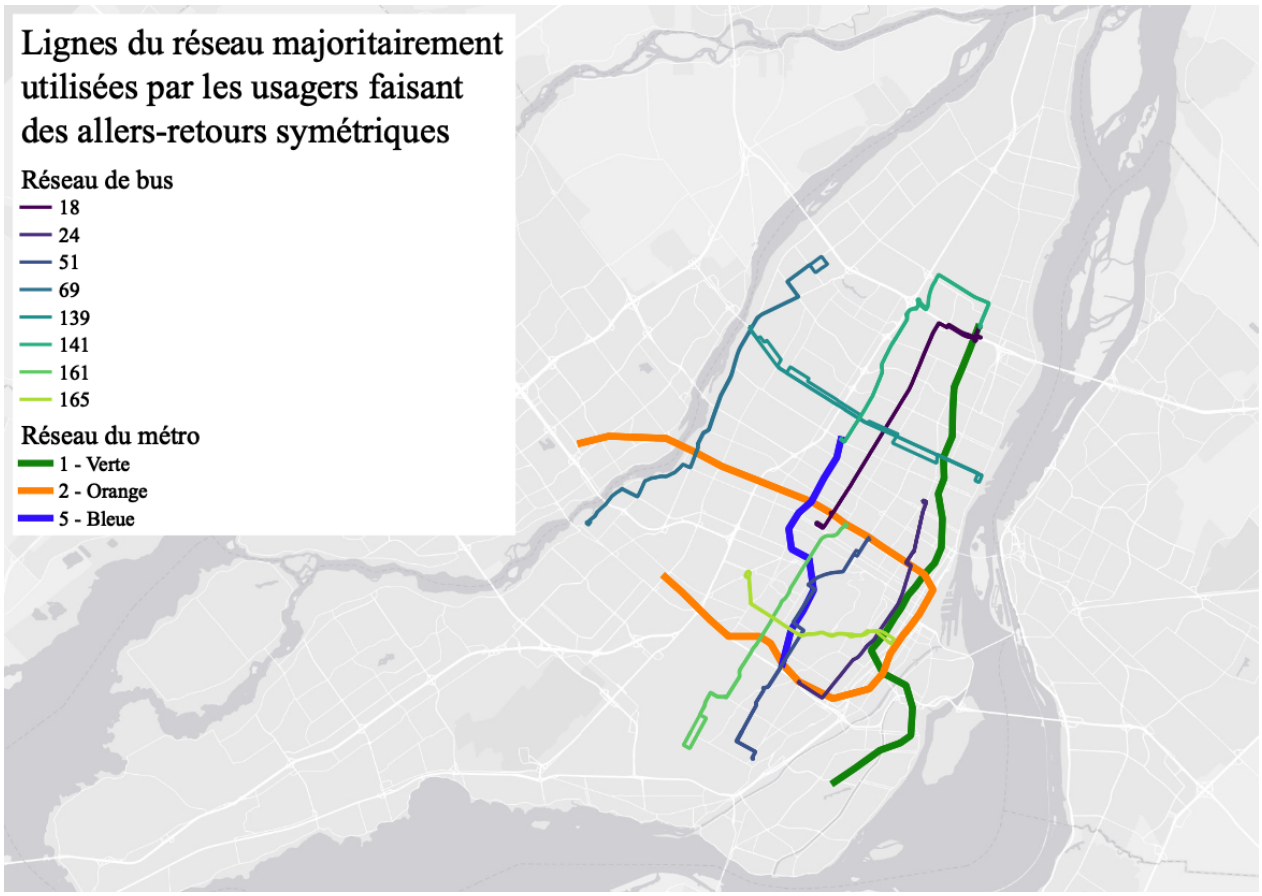


Figure 4.11 Lignes fréquemment utilisées par les usagers faisant des allers-retours symétriques

4.3 Comportement des usagers de l'échantillon étudié

Une segmentation des cartes-jours basée sur la composition des déplacements quotidiens est réalisée afin de mieux comprendre comment les usagers de l'échantillon étudié ont recours aux correspondances. Cette partie présente en premier la caractérisation des groupes de la segmentation (4.3.1), puis poursuit par l'analyse de l'évolution de leur remplissage (4.3.2). La répartition hebdomadaire des cartes-jours dans les groupes est ensuite présentée (4.3.3) et finalement, l'étude termine par le positionnement géographique des groupes (4.3.4). Il faut noter que tout au long de cette étude, les jours de la première et de la dernière semaine de 2016 ne sont pas sélectionnés. Tous les autres jours atypiques sont conservés. Ainsi, 3 570 000 cartes-jours sont segmentées. De même, cette expérimentation est appliquée sur un échantillon de données qui n'est pas nécessairement représentatif de tous les usagers de la STM.

4.3.1 Groupes de la segmentation

La segmentation des cartes-jours comptabilisant au moins un déplacement est réalisée avec $k = 6$ groupes par la méthode des k-moyennes. Le *cluster* 0 est ajouté a posteriori, composé de toutes les cartes-jours sans déplacement. Le tableau 4.6 présente le centre de chaque groupe obtenu après la segmentation.

Tableau 4.6 Centres des groupes de la segmentation des cartes-jours

Groupe	Dép.†	B [%]‡	M [%]	B,B [%]	B,M [%]	M,B [%]	M,M [%]	B,M,M [%]	M,M,B [%]	Autre [%]
0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
1	2,15	90,94	0,32	2,22	0,46	0,31	0,18	0,12	0,11	5,33
2	2,53	0,77	92,64	0,16	0,98	0,94	1,56	0,15	0,24	2,57
3	2,17	15,88	0,71	69,74	1,63	1,17	0,46	0,57	0,45	9,39
4	2,20	4,34	11,94	0,72	39,49	33,14	0,57	0,18	1,70	7,93
5	2,41	2,27	6,79	0,10	1,62	1,62	77,16	4,23	3,86	2,34
6	2,52	9,46	7,91	1,55	2,42	3,74	3,81	14,72	15,85	40,54

† Nombre de déplacements moyen par carte-jour. / ‡ Pourcentage des déplacements de la carte-jour faits dans la série de modes.

Les sept groupes ont une composition différente basée sur l'utilisation des correspondances. En effet, mis à part le *cluster* 0 qui est composé des cartes-jours sans déplacement, chaque groupe favorise une série de modes en particulier. Le nombre de déplacements moyen par carte-jour ne varie pas significativement d'un groupe à l'autre. Une particularité de comportement peut alors être donnée à chacun des groupes :

Groupe 0. Il n'y a pas de déplacement sur la journée.

Groupe 1. Les déplacements directs en bus sont favorisés.

Groupe 2. Les déplacements directs en métro sont favorisés.

Groupe 3. Les déplacements avec une correspondance entre deux bus sont favorisés.

Groupe 4. Les déplacements avec des correspondances entre bus et métros sont favorisés.

Groupe 5. Les déplacements avec une correspondance entre deux métros sont favorisés.

Groupe 6. Les déplacements avec plus d'une correspondance sont favorisés.

L'écart de composition des groupes démontre qu'il existe une différence de comportement entre les usagers de l'échantillon. En effet, tous les utilisateurs réguliers du TC ne favorisent pas les mêmes modes de transport pour se déplacer. De même, la taille des groupes est variable ce qui signifie que certains comportements s'appliquent à plus de cartes-jours que

d'autres. La figure 4.12 propose une visualisation des caractéristiques de chaque groupe ainsi que la proportion de cartes-jours qui les composent.

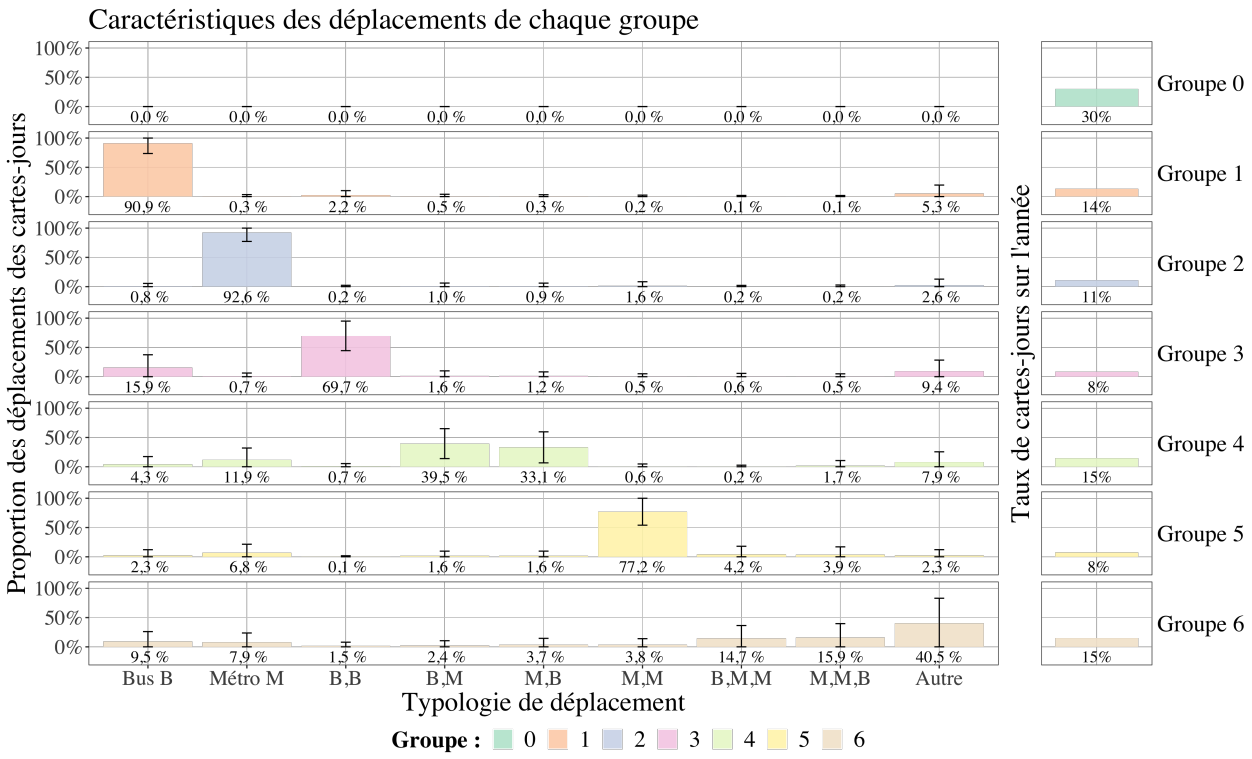


Figure 4.12 Caractéristiques des déplacements de chaque groupe et taille

Par définition, le *cluster* 0 est caractérisé par l'absence de déplacement. Ainsi, il n'y a pas de comportement à déduire. Comme ce groupe n'est pas issu de la segmentation par les k-moyennes, tous les vecteurs qui le composent sont identiques. Par conséquent, il n'y a pas de variabilité entre les éléments. Bien que ce groupe soit un supplément de la segmentation, il est composé de 30 % des cartes-jours.

Les *clusters* 1 et 2 sont constitués des cartes-jours faisant presque exclusivement des déplacements directs, respectivement en bus et en métro. Dans ces groupes, la variabilité entre les éléments de l'ensemble est basse (respectivement $\pm 6{,}1$ % et $\pm 6{,}0$ %) ce qui traduit un comportement stable. La proportion de composition de l'attribut "Autre" peut s'expliquer comme l'inclusion d'un déplacement inhabituel parmi tous les déplacements quotidiens. Également, le *cluster* 1 favorise, en plus des trajets directs en bus, des déplacements avec une correspondance BB. À l'inverse, le *cluster* 2 favorise, en plus des trajets directs en métro, des déplacements avec une correspondance MM. Cela montre que dans cet échantillon, les

usagers favorisent un mode de transport en particulier. Enfin, il y a plus de cartes-jours dans le *cluster* 1 que le *cluster* 2 (14 % et 11 % des cartes-jours). Il y a davantage d'usagers faisant des déplacements directs en bus que d'usagers faisant des déplacements directs en métro.

Pour continuer, les *clusters* 3, 4 et 5 sont constitués des cartes-jours favorisant une correspondance pour leurs déplacements. Pour ces trois groupes, la variabilité entre les éléments de l'ensemble est plus élevée, respectivement $\pm 11{,}2$ %, $\pm 13{,}7$ % et $\pm 11{,}3$ %. Ce sont des comportements moins stables. Le *cluster* 4 est constitué des cartes-jours mélangeant des déplacements avec une correspondance BM et MB. Ce mélange découle de la symétrie des déplacements, observée à la partie 4.2.3. De plus, les *clusters* 4 et 5 vont avoir tendance à favoriser des trajets directs en métro en complément, alors que le *cluster* 3 favorise davantage les trajets directs en bus. Cela montre une nouvelle fois que, dans l'échantillon étudié, les voyageurs favorisent un mode de transport en particulier. Il y a plus de cartes-jours dans le *cluster* 4, ce qui peut être expliqué par sa plus grande polyvalence.

Pour terminer, le *cluster* 6 est celui où il s'observe le plus de variabilité ($\pm 17{,}3$ %). C'est un groupe où les comportements sont moins précis donc les usagers à l'intérieur sont moins stables. Toutefois, 15 % des cartes-jours appartiennent à cet ensemble. Il s'agit d'un groupe où les cartes-jours font majoritairement des déplacements avec plusieurs correspondances. Cependant, les trajets directs et avec une correspondance sont aussi représentés, ce qui montre que dans l'échantillon étudié, les usagers qui ne font que des déplacements avec plusieurs correspondances sont rares.

4.3.2 Fluctuation de composition des groupes

Puisque ce sont les données de cartes-jours qui sont segmentées, une même carte peut changer de groupe pendant l'année et il est possible de suivre son évolution jour après jour. De même, la composition des groupes change selon la date. Cette partie vise à expliquer comment évoluent les groupes et les usagers au cours de l'année. Tout d'abord, la figure 4.13 met en avant la fluctuation de taille des groupes pendant la période étudiée.

En premier, le *cluster* 0 montre une proportion de remplissage et des fluctuations plus élevées que tous les autres groupes. Comme il représente les cartes-jours sans déplacement, cette proportion s'explique par la part d'usagers de l'échantillon qui n'utilisent pas le TC tous les jours. De même, la variabilité élevée dans ce groupe peut s'expliquer par l'alternance des jours ouvrés et des jours de fins de semaine.

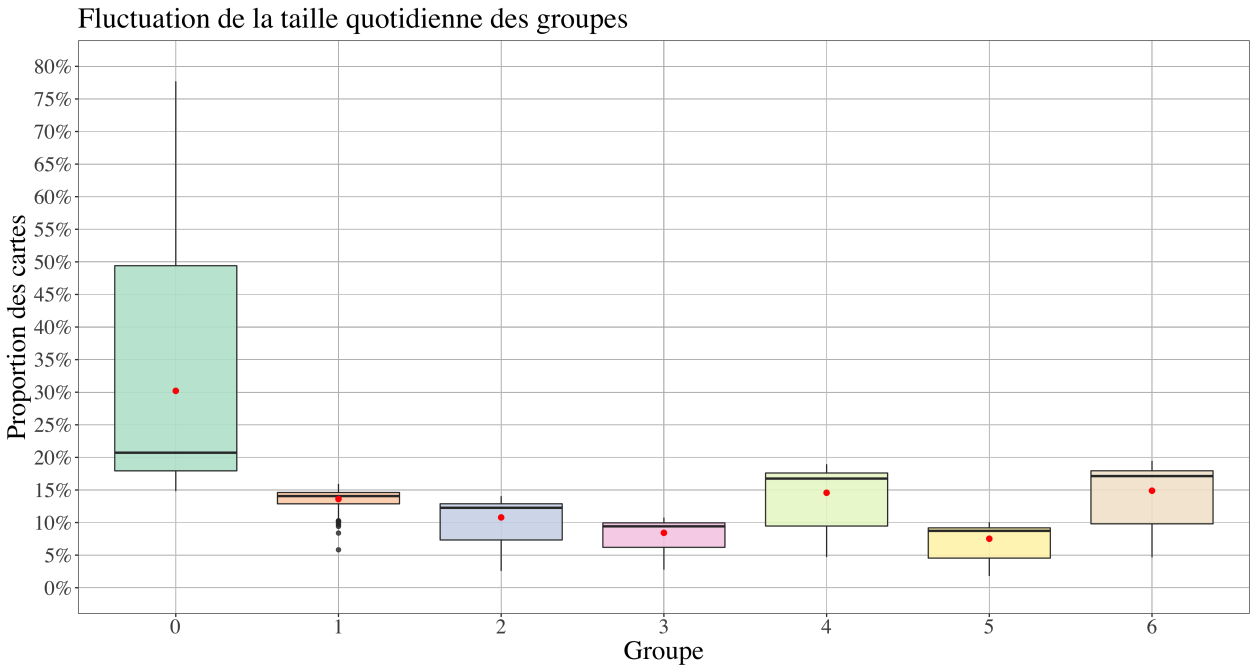


Figure 4.13 Variation de la taille des groupes

Ensuite, les groupes impliquant majoritairement des déplacements en bus (*clusters* 1 et 3) ont une fluctuation de composition plus faible que ceux impliquant majoritairement du métro (2 et 5). Ceci traduit, pour l'échantillon étudié, une plus grande variabilité de comportement des usagers en métro. De même, les fluctuations à l'intérieur des *clusters* 4 et 6 sont plus élevées. Pour expliquer d'où viennent ces variations, il faut étudier quand et entre quels couples de groupes ont lieu les changements.

Probabilité de passage d'un groupe à un autre

Une méthode pour connaître la tendance de comportement des usagers de l'échantillon est le calcul de la matrice de probabilité de changement d'un groupe à l'autre. Chaque nouveau jour, les cartes qui quittent un groupe pour se rendre dans un autre sont comptabilisées et la moyenne annuelle permet de donner la probabilité de passage entre deux ensembles. En d'autres termes, cette probabilité traduit l'attraction entre les groupes. La matrice de probabilité de passage d'un groupe à un autre est donnée au tableau 4.7.

Quel que soit le groupe, il est autant probable pour une carte de rester dans celui-ci que d'en changer. En moyenne, la moitié des cartes (50,96 %) se maintiennent dans le même ensemble deux journées successives. Cela signifie qu'au sein de l'échantillon d'usagers, certains passagers

sont réguliers tout en soulignant que les comportements sont variables d'un jour à l'autre. Ceci peut être due à l'alternance des jours ouvrés et des fins de semaine.

Tableau 4.7 Matrice de probabilité de passage d'un groupe à un autre

	0	1	2	3	4	5	6	Total
0	44,97	10,35	9,25	5,71	12,54	6,72	10,46	100,00
1	15,59	45,36	2,40	12,85	9,12	1,74	12,94	100,00
2	12,10	2,07	60,81	0,62	10,24	8,69	5,48	100,00
3	11,66	18,38	1,14	47,28	6,59	1,52	13,43	100,00
4	9,67	5,82	5,82	3,64	55,27	3,20	16,58	100,00
5	11,43	2,90	11,11	0,64	7,41	53,14	13,37	100,00
6	10,77	7,00	3,90	7,27	15,55	5,65	49,87	100,00

Valeurs en pourcentage des cartes-jours du groupe de départ [%].
Ligne : groupe de départ / Colonne : groupe d'arrivée.

Le *cluster* 0 est le moins régulier. Chaque nouveau jour, il perd la majorité des cartes qui le composaient le jour d'avant et il reçoit de nouvelles cartes provenant d'autres groupes. Cette variabilité est cohérente puisque les cartes étudiées proviennent d'un échantillon d'utilisateurs réguliers du réseau. Ce sont des usagers quotidiens du TC pour lesquels les jours consécutifs sans déplacement sont peu fréquents. De même, la définition des comportements du *cluster* 6 est vague, ce qui explique sa variabilité. Cependant, c'est un groupe qui attire principalement les cartes en provenance des *clusters* 3, 4 et 5. À partir de l'échantillon étudié, l'analyse révèle que les usagers favorisant déjà une correspondance en font plus facilement d'autres.

Les *clusters* 2, 4 et 5 montrent la plus grande stabilité de composition, avec plus de la majorité des cartes persistant dans ces groupes d'un jour à l'autre. Ce sont les groupes associés principalement à l'usage du métro. À l'inverse, les *clusters* 1 et 3, associés à l'usage du bus, indiquent une stabilité plus faible. Dans l'échantillon, les usagers favorisant principalement le métro sont donc plus constant que ceux favorisant le bus. Toutefois, beaucoup de cartes échangent chaque jour entre les *clusters* 1 et 3. Ce sont des usagers qui favorisent le bus, certaines fois avec des correspondances et d'autres fois sans. Les échanges entre les *clusters* 2 et 5 sont moins nombreux. Enfin, le *cluster* 4 traduit par les correspondances entre métro et bus, attire autant les usagers du bus que ceux du métro.

En conclusion, dans l'échantillon étudié, les usagers favorisant le bus ont une fluctuation de comportement plus marquée que les usagers favorisant le métro. Cependant, les échanges sont

principalement centrés entre les *clusters* 1 et 3, ce qui montre que malgré les changements de comportement, ces usagers favorisent le mode bus. Pour ces utilisateurs, le choix du mode de transport paraît plus important que le choix de la correspondance. Les passagers du métro sont plus réguliers, mais leurs changements de comportements sont moins prévisibles. Ils semblent favoriser autant le mode métro que le nombre de correspondances. La couverture spatiale du réseau peut être une source d'explication. En effet, avoir accès ou non au métro joue un rôle dans le choix d'itinéraire des usagers.

Composition quotidienne des groupes

La figure 4.14 présente la fluctuation de composition des groupes chaque jour de l'année. Il ne fait aucun doute que le comportement des usagers de l'échantillon varie quotidiennement. Cependant, une régularité hebdomadaire s'observe. La représentation du remplissage des groupes chaque jour montre une démarcation de comportement entre les jours ouvrés et les fins de semaine, ce qui permet de les différencier. Toutefois, même si les comportements varient quotidiennement, il semble y avoir une tendance qui se répète de manière hebdomadaire. Pour l'expliquer, la composition des groupes sur une semaine est étudiée à la partie 4.3.3.

Certaines périodes de l'année ont une composition des groupes non habituelle, marquée par une augmentation du nombre de cartes dans le *cluster* 0. Il s'agit de début mars, fin juin, juillet, août et fin décembre. La méthode de caractérisation des comportements permet donc d'identifier les périodes de vacances et confirme que ces dernières génèrent moins de déplacements de la part des usagers étudiés.

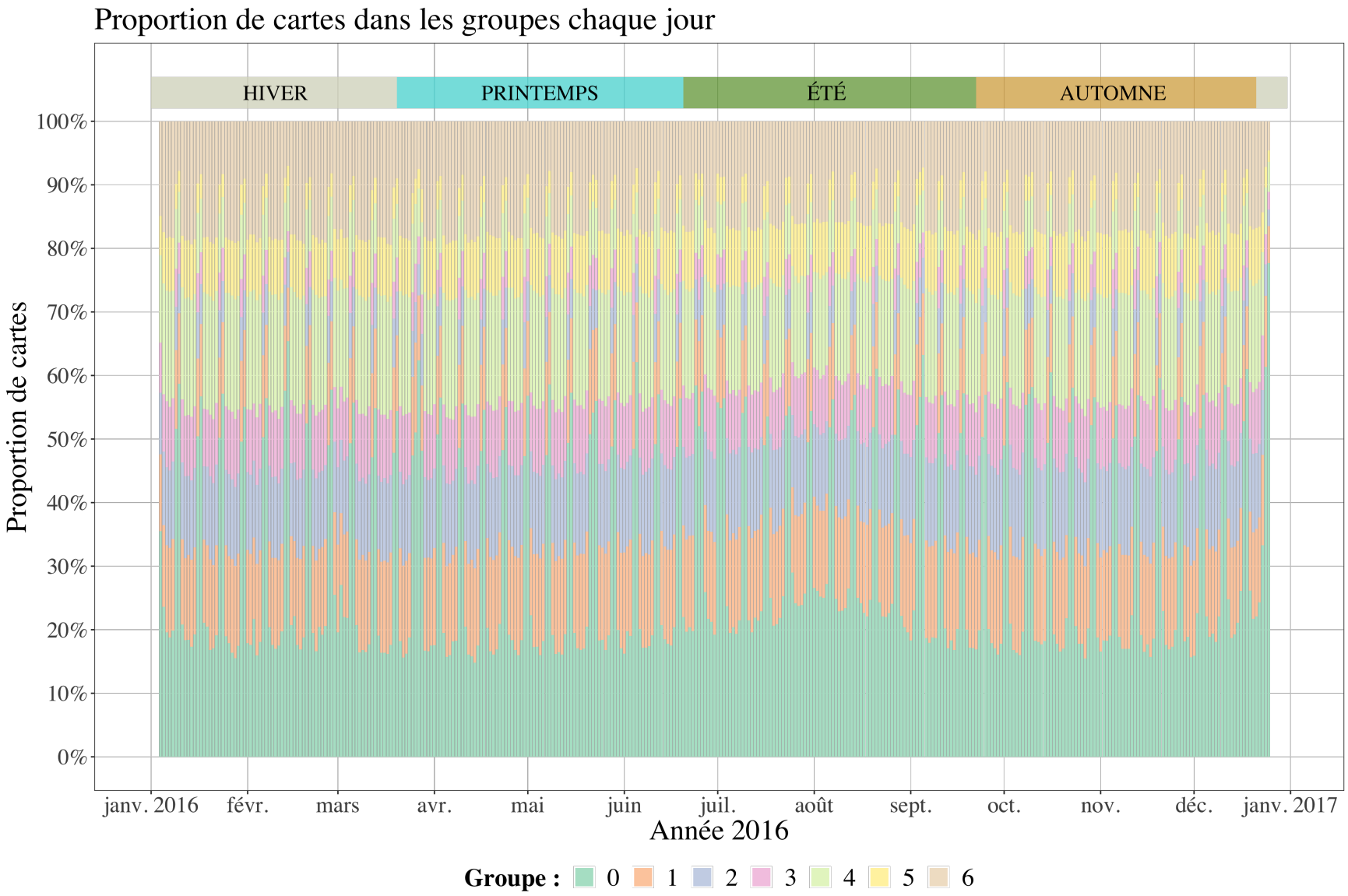


Figure 4.14 Proportion de cartes dans les groupes chaque jour

4.3.3 Répartition hebdomadaire des cartes-jours

La partie précédente explique que la composition des groupes varie jour après jour tout en montrant qu'une régularité hebdomadaire s'observe. Pour y donner suite, l'étude de la composition des groupes en fonction du jour de la semaine est réalisée. Elle vise à mettre en évidence la variation de comportement des usagers de l'échantillon au cours de la semaine. Pour cela, la figure 4.15 présente la répartition moyenne des cartes-jours dans les groupes en fonction du jour de la semaine.

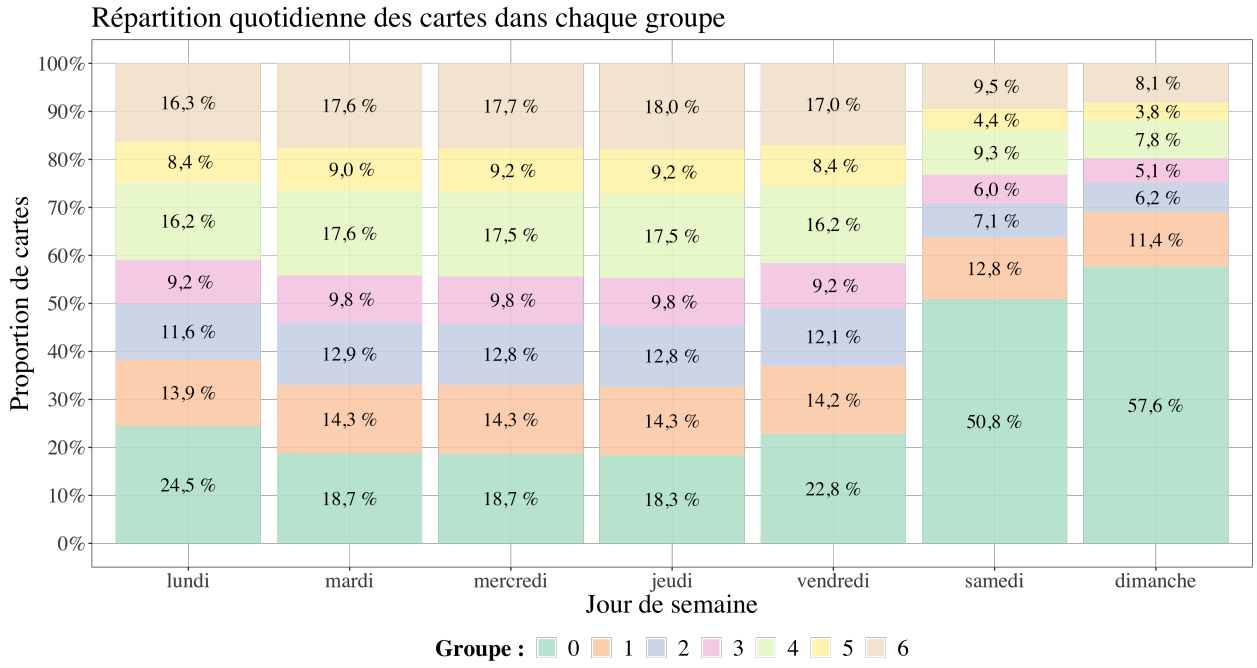


Figure 4.15 Répartition quotidienne des cartes dans les groupes

Pendant les jours ouvrés, la composition des groupes varie peu. Cette stabilité met en avant la régularité des usagers étudiés au court de la semaine. Toutefois, une différence de comportement s'observe le lundi et le vendredi. En effet, ces jours sont marqués par un nombre plus important de cartes dans le *cluster* 0. Autrement dit, moins d'usagers se déplacent le premier et le dernier jour de la semaine ouvrable. Cependant, la proportion d'usagers dans les autres groupes garde le même prorata. L'absence de déplacement le lundi et le vendredi concerne tous les ensembles, donc tous les types de comportements. Ces observations coïncident avec celles apportées par la figure 4.3.

Il y a une fracture dans la composition des groupes entre les jours ouvrés et les jours de fin de semaine. La fin de semaine est marquée par une augmentation majeure de la taille du

cluster 0. En majorité, dans l'échantillon étudié, les usagers n'utilisent pas le TC la fin de semaine et ils sont davantage le dimanche que le samedi. En regardant le prorata de composition des groupes, il s'observe que les groupes perdant le plus de passagers sont ceux associés à l'usage du métro. Ces observations coïncident avec les résultats obtenus précédemment.

En utilisant seulement cette figure, il est impossible de conclure sur les changements de comportement des usagers de l'échantillon. En effet, cette représentation explique comment évolue la composition des groupes dans la semaine, mais ne prend pas en compte les flux d'usagers. Pour visualiser de quel groupe proviennent et vers quel groupe s'en vont les cartes à chaque changement de jour, un diagramme de Sankey [67] est proposé. Ce diagramme peut représenter toute la période temporelle de l'étude. Cependant, pour des questions de lisibilité, il est conseillé d'afficher seulement une courte période. Pour l'analyse, le diagramme de Sankey de la semaine 42 avec le lundi de la semaine 43 est présenté à la figure 4.16.

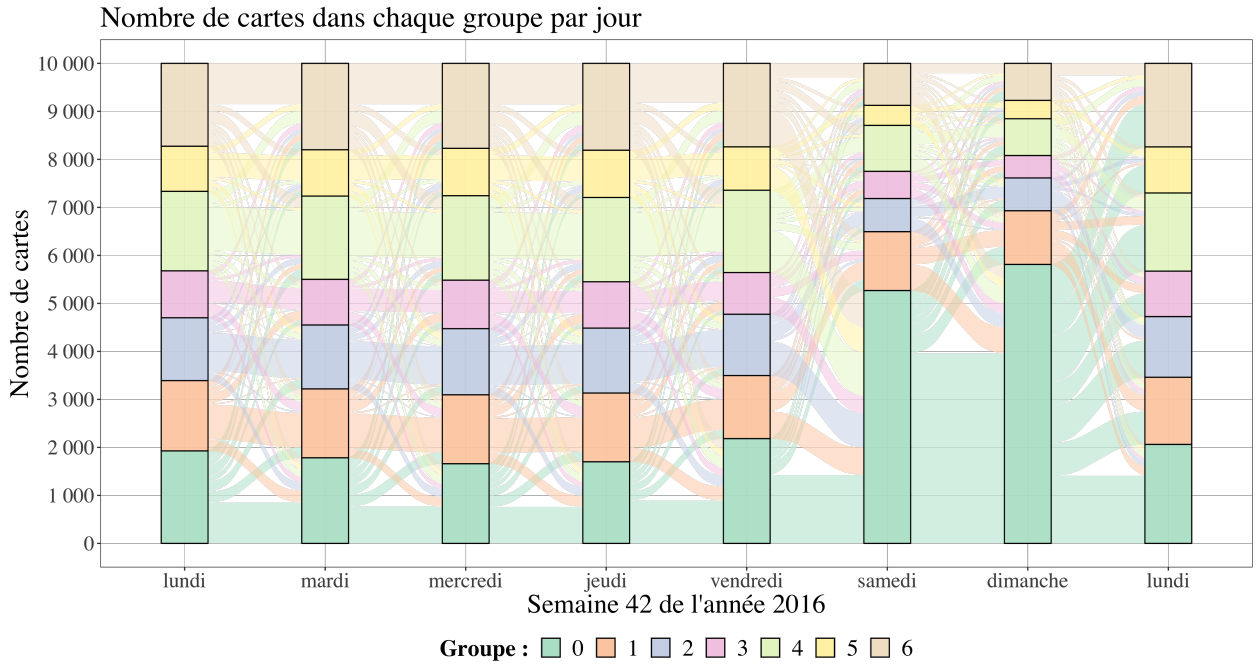


Figure 4.16 Diagramme de Sankey de la semaine 42 (du 17 au 24 octobre 2016)

La période couverte est du 17 au 24 octobre 2016. Il est choisi d'étudier cette semaine puisqu'elle est sur un seul mois, sans journée atypique et pendant une période d'activité normale. Le diagramme de Sankey montre le nombre de cartes dans chaque groupe par jour ainsi que les flux de cartes entre deux jours consécutifs. Les changements de groupes sont ainsi identifiables.

Tout d'abord, la différence de flux entre les jours ouvrés et les jours de fin de semaine est marquée. Pendant les jours ouvrés, les flux de cartes qui s'observent d'un jour à l'autre suivent les mêmes tendances, qui sont proches des proportions lues dans la matrice de probabilité de passage d'un groupe à un autre du tableau 4.7. Cependant, comme il n'y a pas deux semaines identiques, il y a des différences chaque jour.

Le samedi, le *cluster* 0 augmente de volume en absorbant des cartes en provenance de tous les autres groupes. De même, la proportion des cartes restant dans le *cluster* 1 est plus grande que celle des cartes restant dans le *cluster* 2. La proportion des cartes qui restent dans un groupe caractérisé par une correspondance est faible. De plus, le *cluster* 1 est celui qui attire le plus de cartes le samedi après le *cluster* 0. Cela signifie que le week-end, les usagers de l'échantillon étudié vont avoir tendance à ne pas utiliser le TC, ou alors en favorisant le bus et des trajets directs. Les mêmes tendances s'observent entre le samedi et le dimanche et, dans cet échantillon, beaucoup d'usagers ne se déplacent aucun des deux jours.

Entre le dimanche et le lundi, une partie des cartes du *cluster* 0 est redistribuée dans tous les autres groupes. De manière générale, pendant la semaine, les changements entre deux groupes différents sont peu nombreux. Finalement, dans l'échantillon utilisé, l'analyse révèle que le comportement des usagers est plus instable pendant la fin de semaine que durant les jours ouvrés, ce qui peut s'expliquer par la présence d'activités inhabituelles en fin de semaine, nécessitant d'autres itinéraires.

4.3.4 Positionnement géographique des groupes

La typologie des lignes présentée à la section 3.4 propose un regroupement des lignes du réseau en fonction des différentes zones qu'elles traversent (centre-ville, centre, Est et Ouest). Les dix combinaisons de lignes obtenues sont listées dans le tableau 4.8. Le repère attribué à chaque combinaison permet l'étude de l'utilisation des lignes dans les groupes.

Tableau 4.8 Combinaisons de régions des lignes de la STM

Centre-ville	Centre	Est	Ouest	Repère†	Nombre de lignes	Proportion	Illustration
0	0	0	1	A	29	12,94 %	E.1.a
0	0	1	0	B	7	3,13 %	E.1.b
0	1	0	0	C	30	13,39 %	E.1.c
0	1	0	1	D	54	24,11 %	E.1.d
0	1	1	0	E	45	20,09 %	E.2.a
0	1	1	1	F	3	1,34 %	E.2.b
1	0	0	0	G	6	2,68 %	E.2.c
1	1	0	0	H	34	15,18 %	E.2.d
1	1	0	1	I	12	5,36 %	E.2.e
1	1	1	0	J	4	1,79 %	E.2.f

† Repère : lettre associée à chaque combinaison de régions existante.

Tout d'abord, le tableau 4.8 montre que la majorité des lignes de la STM traversent le centre (182 lignes). Il y a également 98 lignes qui passent dans l'Ouest, 59 dans l'Est et 56 dans le centre-ville. L'Ouest est la plus grande région de l'Île de Montréal, ce qui explique la part importante des lignes qui la traversent. La majorité des lignes sont à la fois dans deux régions de l'Île. La plupart sont des lignes liant le centre à l'Est ou à l'Ouest. Ce sont des lignes de rabattement qui permettent d'aider les usagers éloignés du centre à rejoindre le métro. Certaines lignes permettent d'aller de l'Est à l'Ouest en traversant le centre et d'autres permettent de rejoindre le centre-ville. Toutefois, aucune ligne ne permet de traverser les quatre régions de l'Île de Montréal.

L'objectif de cette étude est alors de comprendre comment les lignes sont utilisées par les voyageurs de chaque groupe. Pour ce faire, l'histogramme de la figure 4.17 représente la proportion des lignes de chaque combinaison de régions empruntées par les usagers des groupes. Bien sûr, puisque le *cluster* 0 est composé des cartes-jours qui ne se déplacent pas, il n'est pas considéré dans cette analyse.

L'histogramme met en avant une différence d'utilisation des lignes du réseau selon les usagers de chaque ensemble. Les groupes associés à l'usage du bus (1 et 3) ont un taux d'utilisation des lignes dans les régions Est et Ouest plus élevé que ceux associés à l'usage du métro (2 et 5). La suite de cette partie analyse la répartition de chaque combinaison dans les groupes.

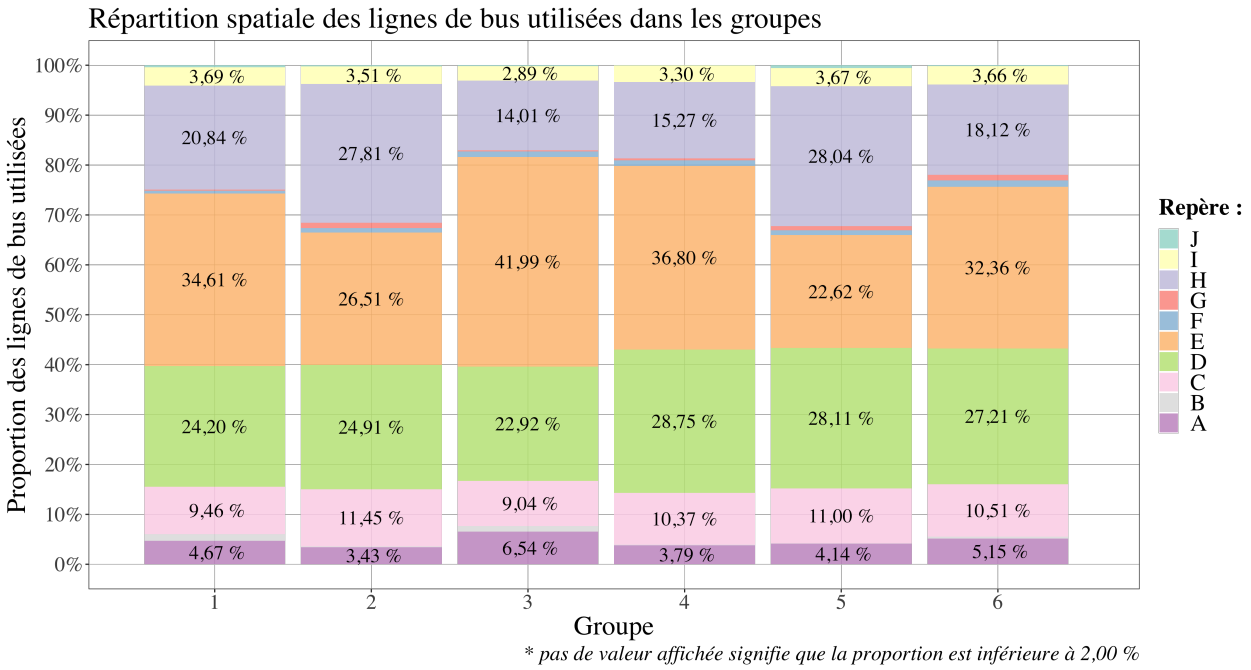


Figure 4.17 Répartition spatiale des lignes utilisées dans chaque groupe

En premier, la combinaison A caractérisant les lignes uniquement dans l'Ouest, est plus utilisée par les passagers de l'échantillon qui favorisent le bus. En particulier, ce sont les usagers qui font des correspondances BB qui ont le plus recours à ces lignes. Comme elles sont éloignées du centre et avec peu d'interactions avec le métro, les usagers doivent faire des correspondances s'ils souhaitent se rendre au cœur de la ville. La même observation peut se faire avec la combinaison B caractérisant les lignes seulement dans l'Est. Cependant, ces lignes n'ont aucune interaction avec le métro, ce qui explique pourquoi elles sont utilisées uniquement par les *clusters* 1 et 3.

Les lignes de la combinaison C sont celles n'étant que dans le centre. Elles sont plus utilisées par les *clusters* 2 et 5, puis le 4. Autrement dit, ces lignes de bus sont favorisées par les usagers de l'échantillon qui utilisent principalement le métro. Ce sont des voyageurs qui restent dans le centre ou qui changent de région en métro.

Ensuite, la combinaison D caractérise les lignes à la fois dans le centre et l'Ouest. Ce sont des lignes qui ont beaucoup d'interactions avec le métro. Dans l'échantillon étudié, elles sont plus utilisées par les usagers qui ont recours à des correspondances en métro ou qui utilisent plus d'une correspondance. Cependant, ces lignes sont autant utilisées par les voyageurs qui favorisent le bus que ceux qui favorisent le métro. Une supposition est que ces lignes de

rabattement sont utilisées par tous les passagers de l'échantillon qui souhaitent relier l'Ouest au centre. Arrivés dans le centre, certains choisissent de continuer en bus, d'autres en métro.

La différence d'utilisation est plus marquée pour la combinaison E, qui caractérise les lignes à la fois dans le centre et l'Est. Ces lignes ont moins d'interactions avec le métro que celles de la combinaison D et par conséquent, elles sont majoritairement utilisées par les voyageurs qui favorisent le bus ou des correspondances entre les deux modes. Comme précédemment, dans l'échantillon étudié, ce sont les usagers qui utilisent les lignes de rabattements pour relier l'Est et le centre. Mais de ce côté de l'Île, le bus est favorisé.

La combinaison F est constituée de trois lignes qui traversent le centre pour relier l'Est et l'Ouest. Elles sont autant utilisées dans tous les groupes, sauf le 1 où son usage est inférieur. Cela signifie que ces lignes ne sont pas prises seules. Les passagers les utilisent comme celles des combinaisons D et E.

Pour continuer, les lignes du centre-ville sont représentées par la combinaison G. Dans cet échantillon, elles sont principalement utilisées par les voyageurs des *clusters* 2, 5 et 6. Autrement dit, les lignes du centre-ville sont surtout empruntées par les passagers qui favorisent le métro. Il peut s'agir d'usagers travaillant dans le centre-ville et prenant le métro pour le rejoindre. La même observation peut être faite pour la combinaison H caractérisant les lignes du centre et du centre-ville. Ce sont des usagers qui restent dans le centre de l'Île et qui par conséquent favorisent le métro.

Enfin, les combinaisons I et J sont autant utilisées dans tous les groupes. Elles sont peut-être empruntées par les usagers se rendant au centre-ville depuis l'Est ou l'Ouest. Comme l'opportunité d'utiliser le métro est élevée, certains voyageurs de l'échantillon favorisent ensuite le métro alors que d'autres restent en bus.

En conclusion, les lignes les plus utilisées par les passagers sont celles qui sont à la fois dans le centre et l'Est, même si elles sont moins nombreuses que celle dans le centre et l'Ouest. Dans l'échantillon étudié, l'analyse révèle que les usagers favorisants le bus sont répartis partout sur le territoire alors que ceux favorisant le métro sont principalement au centre de l'Île. Cela s'explique par la présence du métro seulement dans le centre et le centre-ville. Peut-être que certains usagers favorisent le bus car ils n'ont pas d'accès direct au métro, ils ne peuvent pas faire autrement. Les passagers qui favorisent à la fois le bus et le métro sont ceux qui utilisent des lignes de rabattements pour rejoindre le métro afin d'aller dans le centre.

4.4 Utilisation des lignes du réseau par les usagers de l'échantillon

L'analyse de l'utilisation des lignes est faite afin de mesurer l'impact de chacune d'entre elles sur l'ensemble du réseau. Il s'agit de détecter les lignes fortement génératrices de correspondances et d'expliquer comment elles sont utilisées par les usagers de l'échantillon étudié. Cette partie s'intéresse premièrement à l'analyse de l'indice d'autonomie de chaque ligne (4.4.1) puis présente ensuite les graphes d'interactions (4.4.2). Finalement, les stations les plus utilisées sont localisées (4.4.3). L'expérimentation est faite à partir d'un échantillon de données sélectionné aléatoirement afin de valider les méthodes. Cet échantillon n'est pas nécessairement représentatif de l'utilisation réelle du réseau de la STM.

4.4.1 Indice d'autonomie des lignes

L'application des équations explicitées à la section 3.8.1 permet d'associer à chaque ligne un indice d'autonomie. Il s'agit d'un indicateur caractérisant l'usage d'une ligne pour correspondre. Il est basé sur le taux de correspondances généralisé et plus l'indice d'autonomie d'une ligne est élevé, moins elle est utilisée par les usagers de l'échantillon étudié pour faire des correspondances.

Cette partie de l'étude vise alors à détailler l'autonomie de chaque ligne. Elle débute par l'analyse de la distribution des indices, à la fois en fonction de la classification des lignes et de leur géolocalisation. Elle poursuit par le calcul des liens entre l'autonomie et la typologie et finit par le développement d'une méthode de prédiction de l'indice.

Distribution de l'indice d'autonomie

En premier, la figure 4.18 décrit comment est réparti l'indice d'autonomie au sein des lignes du réseau de la STM, dépendamment de l'échantillon utilisé. La classification des lignes est prise en compte dans la distribution afin de fournir davantage d'informations.

À partir de l'échantillon étudié, l'indice d'autonomie moyen est 0,68, ce qui signifie que la majorité des lignes ont un taux de correspondances généralisé inférieur à 40 %. Elles sont plus utilisées pour des trajets directs que pour des déplacements avec une correspondance. Toutes les lignes ont un indice compris entre 0,43 et 0,98. Le minimum est attribué à la ligne jaune, ce qui s'explique par sa faible couverture. Comme cette ligne n'a que deux stations, les voyageurs l'empruntent généralement avec une correspondance depuis ou vers un autre lieu. À l'inverse, la ligne la plus autonome est la 126, du réseau local. Son autonomie presque

intégrale signifie que ses passagers ne l'utilisent pas avec des correspondances. En réalité, c'est une ligne de transport courte et peu connectée permettant de faire le tour du secteur Bois-Franc, une zone industrielle de l'Ouest de l'Île. La majorité des lignes ont une autonomie comprise entre 0,55 et 0,75.

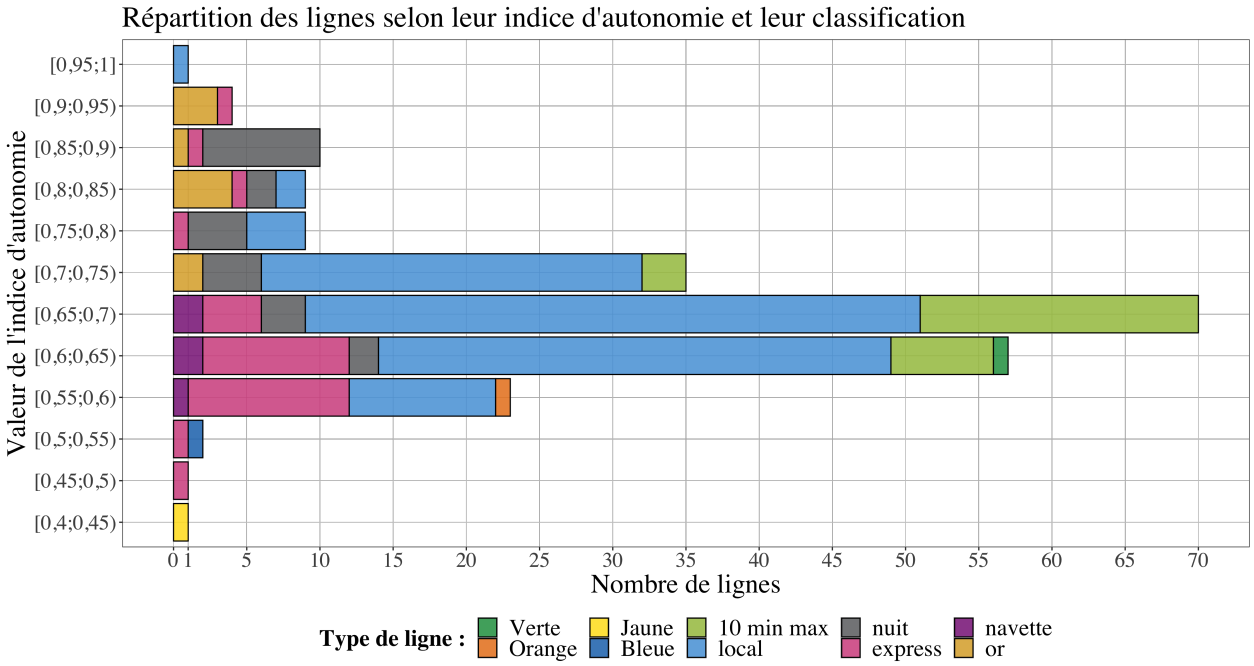


Figure 4.18 Distribution de l'indice d'autonomie des lignes selon leur classification

L'analyse appliquée sur l'échantillon de données révèle que la distribution de l'indice d'autonomie diffère selon la classification des lignes. En effet, toutes celles du réseau de navettes OR ont un niveau d'autonomie supérieur à 0,70. Cela est cohérent puisqu'elles sont conçues afin de faciliter les déplacements des aînées. Elles sont pensées pour desservir les centres d'intérêt de ces personnes sans qu'elles aient besoin de faire un changement. Par conséquent, leur taux de correspondances est bas et leur autonomie élevée. Les lignes du réseau de nuit ont également une autonomie supérieure à la moyenne. Comme elles sont utilisées à des heures où le reste du réseau est à l'arrêt, elles offrent peu de potentialité de correspondance. Elles sont autonomes car elles ont peu d'interactions avec d'autres lignes. À l'inverse, toutes les lignes du réseau haute fréquence sont comprises entre 0,60 et 0,75. Elles ont une autonomie plus faible qui marque une plus grande capacité à interagir avec les autres lignes. Dans l'échantillon d'usagers, ces lignes achalandées sont autant utilisées lors de déplacements directs qu'avec des correspondances. C'est aussi le cas pour les navettes. Enfin, certaines lignes du réseau express suivent les tendances des navettes OR alors que d'autres ont une faible autonomie.

Cela est cohérent puisque les lignes express sont conçues pour faire du rabattement vers le métro aux heures de pointe et pour desservir des destinations importantes. Elles ont ainsi un taux de correspondances élevé.

Les lignes de métro ont un indice d'autonomie moyen plus faible que celui des bus, respectivement 0,55 et 0,68. Cela s'explique par le plus grand nombre d'interactions qu'a une ligne de métro avec le reste du réseau. En effet, de par sa couverture et sa vitesse d'opération élevée, une ligne de métro attire les usagers, qui font des correspondances pour la rejoindre. Par conséquent, ce sont des lignes avec un fort taux de correspondances généralisé, ce qui conduit à un indice d'autonomie plus bas. En particulier, la ligne verte est la plus autonome, suivie de la orange, de la bleue et de la jaune. Dans l'échantillon étudié, les lignes de métro éloignées du centre-ville génèrent plus de correspondances.

Pour continuer, le diagramme de la figure 4.19 illustre la variabilité de l'autonomie des lignes en fonction de leur combinaison de régions, dépendamment de l'échantillon étudié. Quatre combinaisons se démarquent par une autonomie plus élevée. Les lignes traversant à la fois le centre-ville, le centre et l'Ouest ou l'Est sont les plus autonomes. Cela s'explique par leur bonne couverture spatiale. Comme elles sont étendues sur plusieurs zones, les usagers de l'échantillon n'ont pas besoin de faire des correspondances pour passer d'une région à l'autre. Ainsi, l'indice d'autonomie augmente. C'est la même raison pour les lignes qui sont à la fois dans le centre et le centre-ville. Cependant, cette explication ne s'applique pas pour les lignes de l'Est. Dans leur cas, l'indice d'autonomie élevé provient de leur moins bonne connectivité avec le reste du réseau. En effet, comme les lignes de l'Est ont moins d'interactions avec les autres lignes, elles génèrent moins de correspondances.

En conclusion, cette analyse montre que sur l'échantillon étudié, l'indice d'autonomie dépend à la fois de la classification et de la géolocalisation des lignes. Cependant, une autonomie élevée peut signifier deux situations distinctes. La première consiste en une ligne ayant une bonne couverture et pouvant ainsi déplacer les usagers de l'échantillon sans qu'ils aient besoin de faire une correspondance. La seconde à l'inverse consiste en une ligne ayant peu d'interactions avec le reste du réseau et offrant ainsi moins de possibilités de correspondre. Il faudrait pouvoir distinguer ces deux situations à l'aide de plusieurs indicateurs pour renforcer l'information fournie aux planificateurs.

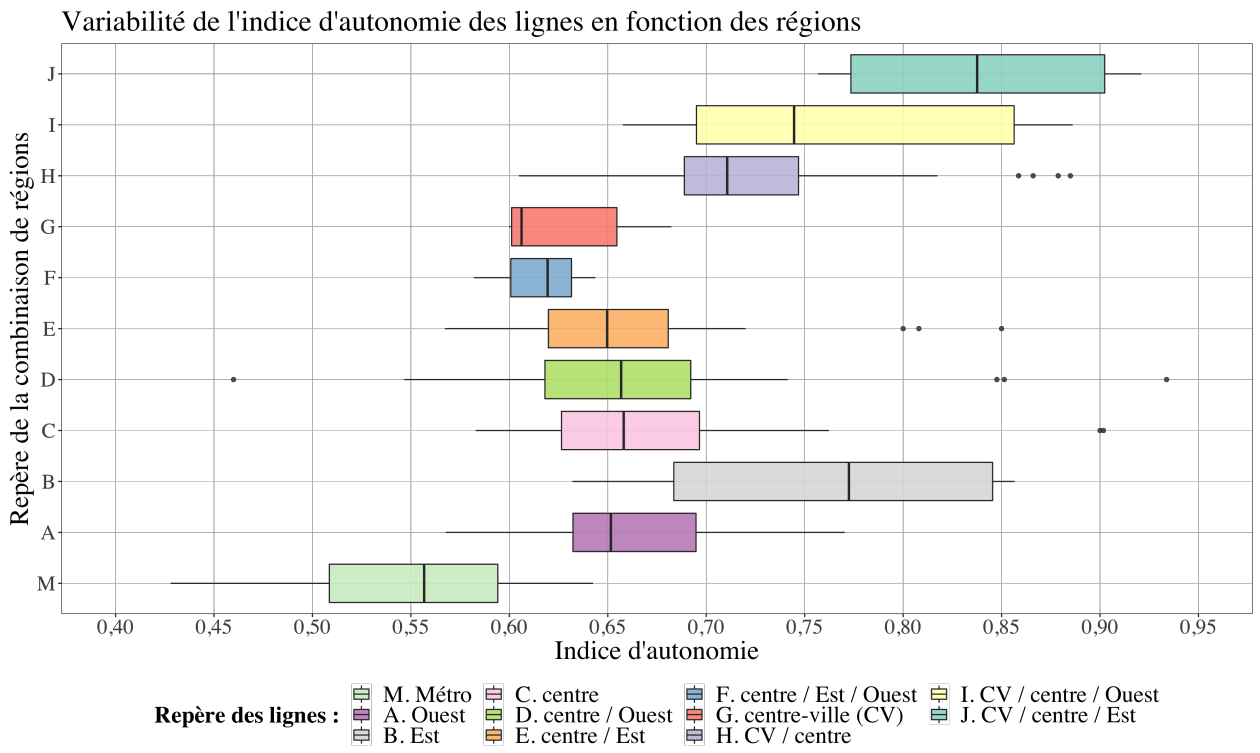


Figure 4.19 Variabilité de l'indice d'autonomie des lignes selon leur combinaison de régions

Liens entre la typologie d'une ligne et son autonomie

Comme le montre la partie précédente, l'indice d'autonomie est lié à certains attributs de la typologie des lignes. Il s'agit à présent d'identifier ceux qui ont un impact significatif. Dans cet objectif, la figure 4.20 présente une matrice de corrélation entre les variables de la typologie des lignes de bus et celles de l'indice d'autonomie. Cette matrice est réalisée avec un test de corrélation de Pearson, de Kendall et de Spearman. Si une case est vide, le niveau de significativité des corrélations entre les deux variables concernées est inférieur au seuil.

Premièrement, la matrice de corrélation confirme les observations précédentes en affirmant que la classification des lignes et les régions traversées ont un impact significatif sur l'autonomie, dépendamment de l'échantillon étudié. De même, elle spécifie que le nombre de lignes et d'arrêts de bus croisés ainsi que le nombre de stations de métro croisées ont une corrélation significative. Dans ce sens, augmenter les interactions entre les lignes tend à améliorer l'autonomie de chacune. Toutefois, le nombre de lignes de métro croisées n'a pas d'impact. Pour qu'une ligne gagne en autonomie, elle doit croiser davantage de stations, sans avoir à croiser plus de lignes de métro.

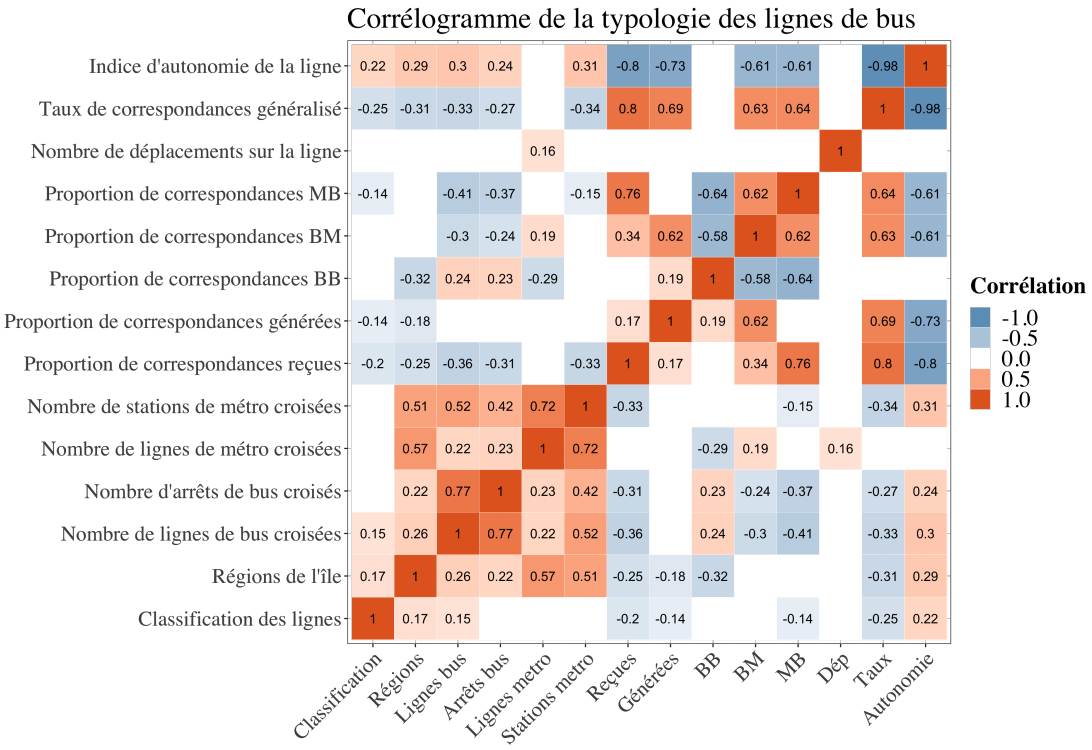


Figure 4.20 Matrice de corrélation des variables de la typologie des lignes de bus

Ensuite, il est à noter que le nombre de déplacements observés sur une ligne n'est corrélé à aucune variable, sauf le nombre de lignes de métro croisées. Cela s'explique par le fait que la typologie des lignes est constituée d'attributs opérationnels et géographiques non liés au nombre de déplacements. De même, l'indice d'autonomie est conçu pour être indépendant de l'achalandage de la ligne. La prise en compte du nombre de déplacements pourrait conduire à des observations différentes.

Pour finir, la matrice permet d'illustrer la corrélation entre chaque couple de variables. Cela montre que la typologie influe sur tous les types de correspondances, mais avec des impacts différents. De même, il y a une corrélation significative entre tous les types de correspondances. Sur une ligne, l'augmentation du nombre de correspondances BM est liée positivement à celle des correspondances MB et négativement à celle des correspondances BB. Ceci est en accord avec les observations de la section 4.2.3 sur la symétrie des déplacements pendulaires.

Explication de l'indice d'autonomie

L'indice d'autonomie repose sur les données d'usages des lignes. Pour l'obtenir, il faut prendre les données d'un échantillon de cartes à puce et appliquer l'algorithme VALIDEP afin de calculer le taux de correspondances généralisé de chaque ligne. Autrement dit, si la ligne n'existe pas encore, cette méthode ne permet pas de prédire son autonomie. À l'inverse, la typologie de la ligne s'obtient facilement à l'aide d'un système d'information géographique (GIS). Pour les planificateurs, expliquer l'indice d'autonomie d'une ligne à partir de sa typologie est alors un moyen d'analyser l'influence d'une stratégie de restructuration du réseau.

L'objectif de cette partie est de proposer une méthode d'estimation de l'indice d'autonomie en utilisant uniquement les informations fournies par la typologie des lignes. Quatre méthodes sont testées. Les trois premières sont des régressions linéaires multiples différenciées par la méthode d'encodage des données. La quatrième est l'application de la méthode de prédiction Random Forest. La comparaison des résultats montre que la régression linéaire multiple appliquée sur des variables catégorisées et cherchant à expliquer la valeur de l'indice d'autonomie normalisé $\log(I_l)$ est la plus précise, car elle minimise l'écart quadratique moyen (MSE). Ainsi, c'est cette méthode qui est présentée. Les tableaux 4.9 et 4.10 présentent respectivement les variables de prédictions utilisées et les résultats de cette régression linéaire multiple. Le tableau 4.11 donne quant à lui le coefficient associé à chaque variable de la régression linéaire multiple.

À partir des résultats du tableau 4.10, il est possible de conclure que sur l'échantillon étudié, 59 % de la variation de l'indice d'autonomie est explicable à partir des variables sélectionnées. Un coefficient de détermination ajusté plus élevé pourrait être souhaité, mais sa valeur moindre se justifie par l'irrégularité des voyageurs de l'échantillon dans leur utilisation des lignes du réseau. Toutefois, il est intéressant de noter que la valeur-P du modèle est inférieure à la statistique-F, ce qui indique que la régression linéaire multiple explique mieux la variation qu'un modèle sans variables indépendantes.

En conclusion, l'usage d'une régression linéaire multiple est approprié pour expliquer l'indice d'autonomie d'une ligne à partir de sa typologie. Ainsi, lorsque les planificateurs du réseau de transport étudient une nouvelle ligne, il peuvent estimer à l'avance son autonomie. Ils obtiennent ainsi des informations sur la façon dont les voyageurs vont l'utiliser et sur les correspondances qui seront induites, ce qui peut faciliter la prise de décision.

Tableau 4.9 Variables de la régression linéaire multiple

Variable	Format	Description	Valeur	Moyenne
max_10	Binaire	Ligne du réseau haute fréquence	1 Oui / 0 Non	13 %
local	Binaire	Ligne du réseau local	1 Oui / 0 Non	55 %
nuit	Binaire	Ligne du réseau de nuit	1 Oui / 0 Non	11 %
express	Binaire	Ligne du réseau espress	1 Oui / 0 Non	14 %
navette	Binaire	Ligne du réseau de navettes	1 Oui / 0 Non	2 %
centre_ville	Binaire	Ligne du centre-ville	1 Oui / 0 Non	25 %
centre	Binaire	Ligne du centre	1 Oui / 0 Non	82 %
est	Binaire	Ligne de l'Est	1 Oui / 0 Non	26 %
ouest	Binaire	Ligne de l'Ouest	1 Oui / 0 Non	44 %
nb_ar_bus	Classe	Nombre d'arrêts de bus croisés	[0 ; 10[, etc.	54
nb_ar_metro	Classe	Nombre de stations de métro croisées	[0 ; 3[, etc.	2
autonomie	Numérique	Autonomie de la ligne	[0 ; 1]	0,68

Tableau 4.10 Résultats globaux de la régression linéaire multiple

Indicateur	Description	Valeur
Coefficient de détermination R^2	Proportion de la variation de la réponse expliquée par le modèle linéaire : qualité de la régression.	0,5907
Coefficient de détermination ajusté R^2_{ad}	R^2 ajusté pour tenir compte du nombre de prédicteurs dans le modèle.	0,5251
Valeur-P (P-Value)	Probabilité que le résultat de la prédiction soit le fruit du hasard. Il doit être inférieur à 0,01.	$2{,}2 \times 10^{-16}$
Statistique-F (F-Statistic)	Indicateur de significativité entre les variables. Si P-Value $<$ F-Statistic, le modèle de régression correspond mieux aux données que le modèle sans variable indépendante.	8,998
Erreur quadratique moyenne MSE	Moyenne des carrés des erreurs : il est inférieur à celui des autres modèles essayés.	$2{,}791 \times 10^{-3}$

Tableau 4.11 Contribution des variables de la régression linéaire multiple

	Coefficient	Erreur std.	Valur-T	Valeur-P
(Intercept)	-0.1779	0.0343	-5.18	0.0000
max_10	-0.2436	0.0322	-7.57	0.0000
local	-0.2409	0.0274	-8.80	0.0000
nuit	-0.1316	0.0402	-3.27	0.0013
express	-0.3080	0.0307	-10.04	0.0000
navette	-0.3776	0.0488	-7.73	0.0000
centre_ville	0.0860	0.0163	5.27	0.0000
centre	-0.0351	0.0156	-2.25	0.0257
nb_ar_bus [60,70)	0.0562	0.0290	1.94	0.0537
nb_ar_bus [120,130)	0.2513	0.0829	3.03	0.0028
nb_ar_bus [200,210)	-0.1749	0.0921	-1.90	0.0592
nb_ar_metro [3,6)	0.0361	0.0170	2.13	0.0346
nb_ar_metro [6,9)	0.0858	0.0282	3.05	0.0027

4.4.2 Interactions entre les lignes

Les graphes d'interactions entre les lignes peuvent être tracés selon diverses agrégations temporelles, en sélectionnant les types de correspondances et en filtrant les données pour afficher le nombre idéal de sommets. Dans la suite de cette section, les liens des graphes affichent le nombre moyen de correspondances observées quotidiennement entre les deux lignes impliquées à partir de l'échantillon d'utilisateurs étudié. Un graphe général est d'abord présenté puis chaque type de correspondance est analysé.

Interactions entre toutes les lignes

Le premier graphe trace les 31 arêtes majoritaires observées quotidiennement sur le réseau, sans séparation des types de correspondances et à partir des données de l'échantillon d'utilisateurs. Cela représente 27 % des correspondances effectuées quotidiennement sur le réseau. Ces 31 arêtes permettent de lier ensemble 20 sommets, c'est-à-dire 20 lignes différentes. Ce graphe représentant tous les types de correspondances est visible avec la figure 4.21.

Plusieurs observations peuvent se faire à partir de ce graphe. Premièrement, parmi les 20 lignes concernées, il y a trois lignes de métro (la verte, la orange et le bleue) et 17 lignes de bus. Il s'observe également 4 arêtes de correspondances MM, 17 arêtes de correspondances MB et 10 arêtes de correspondances BM. Il n'y a aucune arête de correspondances BB. Cela

s'explique par la plus petite proportion de correspondances BB ainsi que par le nombre plus élevé de couples de lignes pouvant être utilisées pour passer d'un bus à l'autre. De même, puisque les graphes représentent les nombres d'interactions, les lignes les plus achalandées apparaissent plus facilement, même si leur taux de correspondances est plus bas.

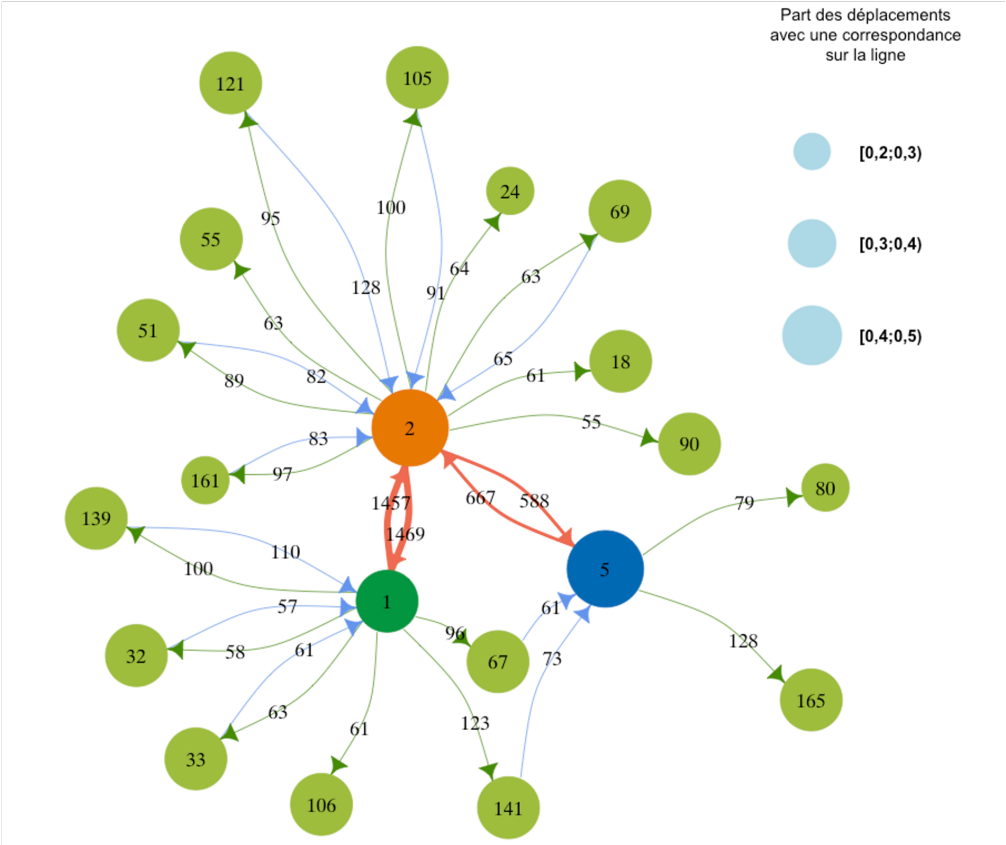


Figure 4.21 Graphe des interactions de tous les types de correspondances confondus

Le graphe montre que les lignes de bus les plus utilisées par les usagers de l'échantillon pour correspondre sont toutes des lignes du réseau haute fréquence (10 minutes max.). Ces lignes sont plus utilisées par les usagers ayant à faire une correspondance, car étant des lignes très régulières, le temps d'attente est plus faible et la correspondance est alors plus commode. De même, toutes les lignes qui ont été identifiées dans l'étude des déplacements pendulaires sont dans ce graphe. Ce sont des lignes favorisées par les passagers de l'échantillon pour leurs déplacements quotidiens.

La ligne de métro ayant le plus d'interactions avec les lignes de bus est la orange. C'est également celle qui a le taux de correspondances généralisé le plus élevé. La ligne verte a un taux de correspondances généralisé inférieur et interagit avec moins de lignes. Enfin, la ligne bleue est en interaction avec quatre lignes seulement, malgré son taux de correspondances

élevé. Cette différence s'explique par l'achalandage plus bas dans la ligne bleue que dans les lignes orange et verte.

Enfin, la majorité des lignes de bus sont impliquées à la fois dans des correspondances BM et MB. Cependant, certaines d'entre elles sont favorisées dans un seul sens, principalement MB. Comme les nombres moyens de correspondances quotidiennes BM et MB sont proches, cela signifie que certaines lignes encouragent un type de correspondance en particulier. Spécialement, les lignes 67 et 141 sont intéressantes puisqu'elles favorisent des correspondances MB depuis la ligne verte et BM vers la ligne bleue. Ce sont des bus empruntés par les usagers de la ligne verte voulant prendre la ligne bleue, sans utiliser le tronçon de la ligne orange entre Berri-UQAM et Jean-Talon, qui est fortement achalandé.

La figure 4.22 montre les lignes du graphe 4.21 sur la carte de Montréal. Toutes les lignes concernées sont des lignes du centre de l'Île ayant un croisement avec le métro.

Figure 4.22 Carte des lignes les plus utilisées pour faire des correspondances

Interactions avec des correspondances bus – bus

Le second graphe étudié est celui des interactions entre deux bus. Il est composé de 35 arêtes pour 20 sommets et représente 4 % des correspondances BB obtenues avec l'échantillon d'usagers. Ce graphe est présenté à la figure 4.23.

Tout d'abord, l'étude des interactions BB montre que sur l'échantillon étudié, les lignes principalement utilisées sont des lignes du réseau haute fréquence. Cependant, plusieurs lignes du réseau express et local sont présentes. Il est normal qu'il y ait plus de correspondances entre les lignes haute fréquence puisqu'elles sont plus régulières et permettent de rendre les transferts plus rapides et moins lourds. De même, ce sont des lignes fortement achalandées. Les lignes express sont moins utilisées pour correspondre, car leur service est proposé uniquement pendant les heures de pointe.

Certaines lignes ont des interactions avec plusieurs autres lignes, notamment la 67, la 139, la 141 et la 165. Ce sont les lignes qu'il faut planifier en fonction des correspondances pour pouvoir améliorer le service. De plus, la majorité des lignes concernées ont un taux de correspondances élevé, ce qui signifie qu'une proportion non négligeable des passagers de l'échantillon ne les utilisent pas seulement pour des trajets directs. Enfin, dans la plupart des cas, il y a le même nombre moyen de correspondances entre deux lignes quel que soit l'ordre du changement. Cela signifie que dans cet échantillon, les usagers changent d'une ligne à l'autre, quel que soit le sens du déplacement. Ceci résulte de la symétrie des déplacements allers-retours.

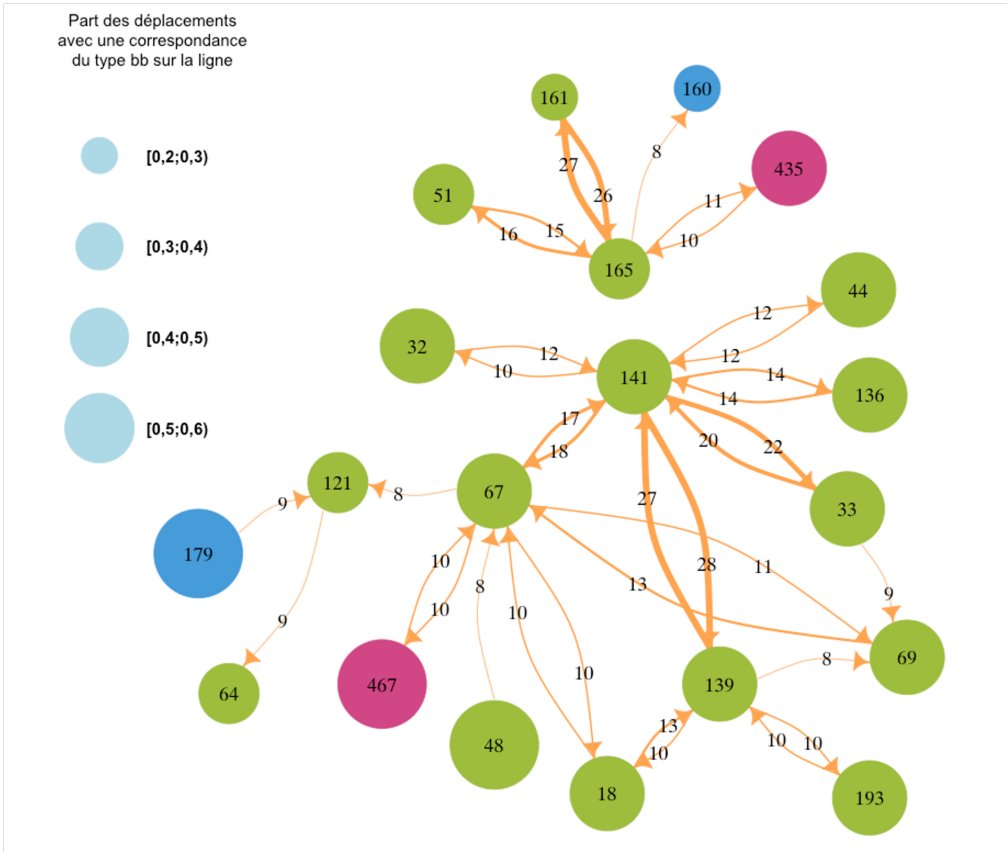


Figure 4.23 Graphe des correspondances entre deux bus

Interactions avec des correspondances bus – métro et métro – bus

Le graphe de la figure 4.24 présente les interactions entre le bus et le métro avec des correspondances BM. Il y a 21 arêtes pour 23 sommets, ce qui représente 35 % de toutes les correspondances BM provenant de l'échantillon étudié. Comme précédemment, il y a majoritairement des lignes du réseau haute fréquence. La plupart d'entre elles génèrent des correspondances vers la ligne orange et à l'inverse, seulement trois lignes génèrent beaucoup de correspondances avec la bleue. La ligne orange est celle qui croise le plus de lignes de bus, et c'est la plus grande du réseau. Il est ainsi normal que ce soit la ligne qui soit la plus concernée par des correspondances. Seule la ligne 165 génère des correspondances vers deux lignes de métro.

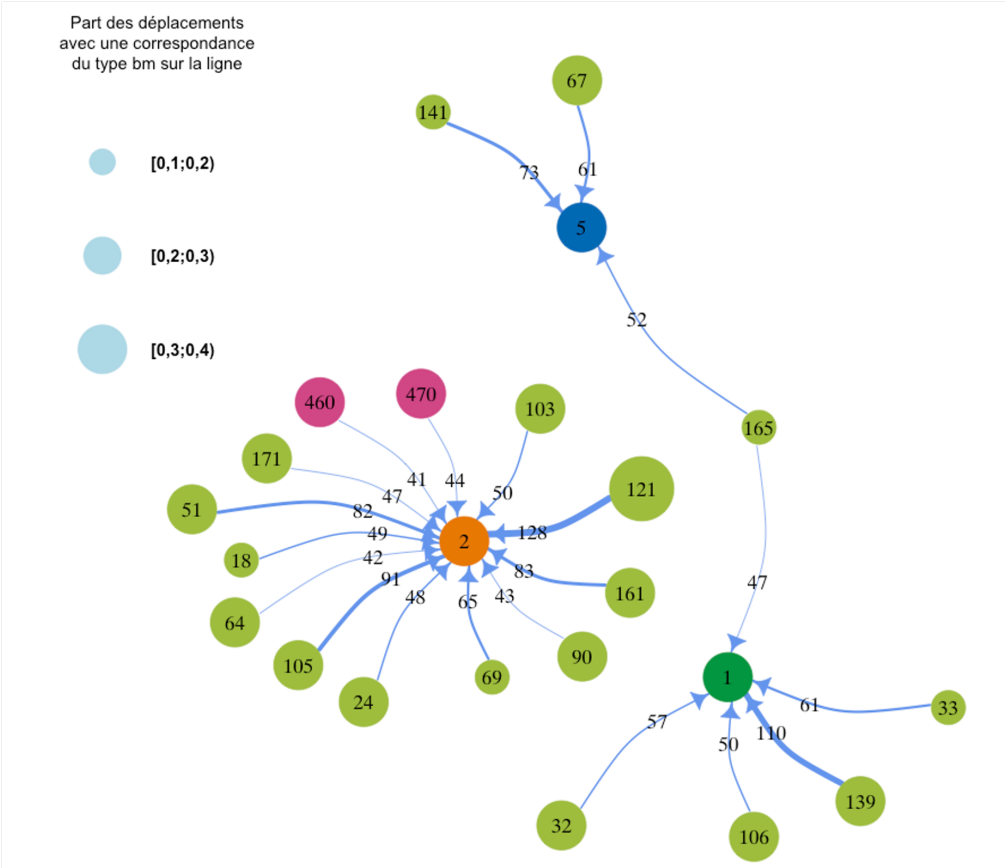


Figure 4.24 Graphe des correspondances d'un bus vers un métro

Les mêmes observations peuvent se faire à partir de la figure 4.25 qui représente 45 % des correspondances MB obtenues à partir de l'échantillon d'usagers étudié. Les deux graphes mettent en avant les mêmes lignes, ce qui caractérise leur utilisation pour des déplacements allers-retours.

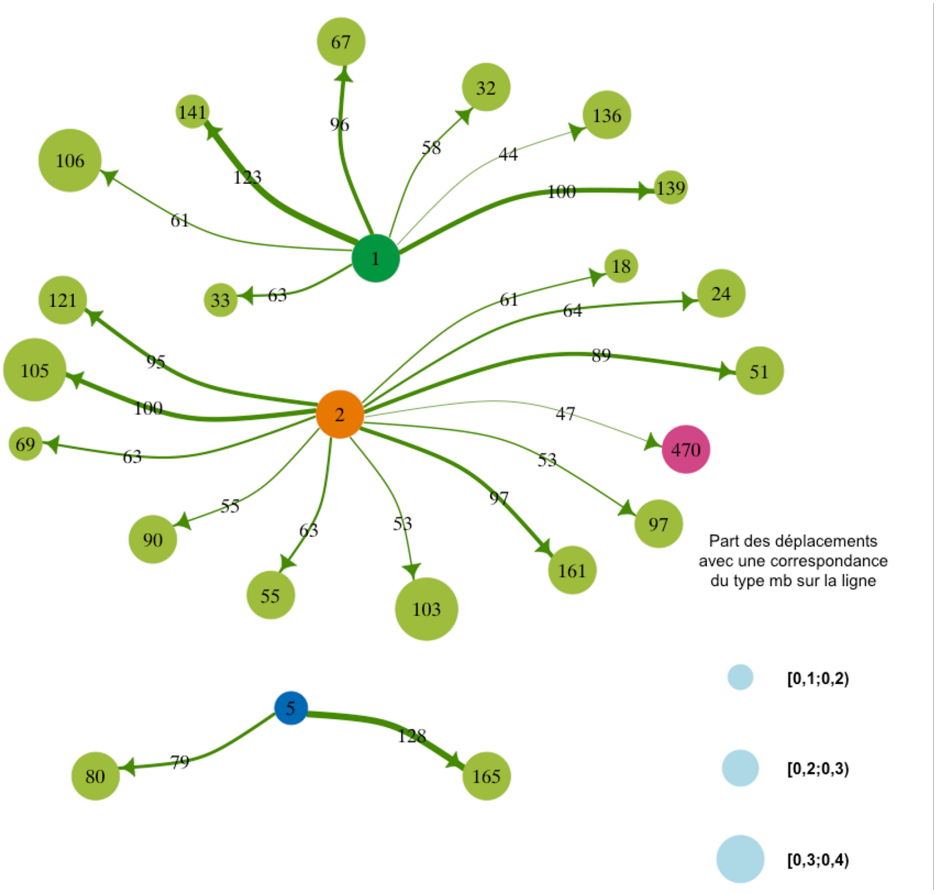


Figure 4.25 Graphe des correspondances d'un métro vers un bus

Interactions avec des correspondances métro – métro

Le dernier graphe proposé est celui des interactions entre les quatre lignes de métro. Il est présenté à la figure 4.26. La ligne orange est la seule ligne du réseau à permettre une correspondance vers toutes les autres. Elle est centrale et c'est elle qui génère le plus de correspondances dans l'échantillon. À l'inverse, la ligne jaune est la moins achalandée et peu de correspondances sont détectées entre cette ligne et les autres. Ceci s'explique par l'absence des données de cartes à puce en provenance de la station Longueuil – Université-de-Sherbrooke. En effet, comme cette station n'est pas sur l'Île de Montréal, elle n'est pas gérée par la STM et les usagers entrant dans cette station doivent avoir un autre titre de transport. Ainsi, dans l'échantillon étudié, seule la station Jean-Drapeau est à l'origine des entrées dans la ligne jaune. La majorité des correspondances MM ont lieu entre la ligne orange et la ligne verte. Cela est cohérent puisque ce sont les deux lignes les plus achalandées dans cette étude.

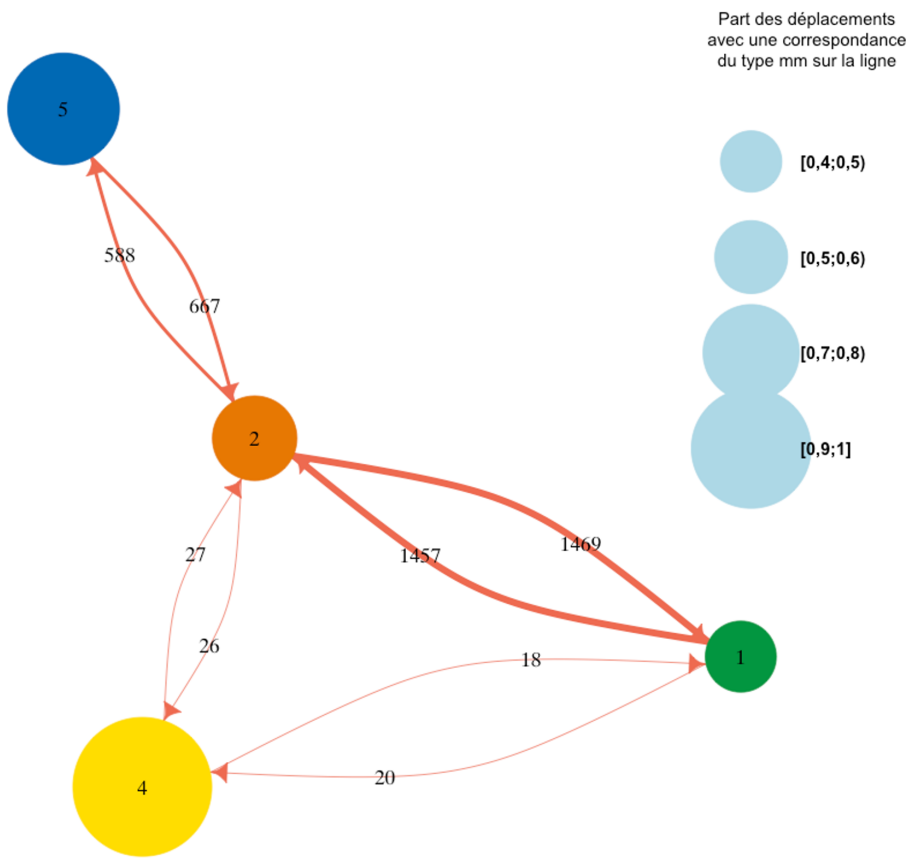


Figure 4.26 Graphe des correspondances entre deux métros

Exemple d'interprétation des graphes

La ligne 141 prend place dans tous les graphes incluants du bus. Cela signifie que, dans l'échantillon étudié, elle génère beaucoup de correspondances. En effet, elle engendre à la fois des correspondances BB (autant vers que depuis d'autres lignes), des correspondances BM principalement vers la ligne bleue, et des correspondances MB principalement depuis la ligne verte. De plus, l'indice d'autonomie de cette ligne est de 0,66, soit un taux de correspondances généralisé de 34 %. C'est ainsi une ligne très utilisée.

Comme cette ligne est achalandée, il peut être bénéfique d'améliorer son niveau de service. Il s'agit d'augmenter la fréquence de la ligne ou de maximiser sa fiabilité pour limiter les délais et les imprévus. La mise en place de mesures préférentielles sur cette ligne peut être envisagée. De plus, comme beaucoup de correspondances sont générées, il peut être utile d'instaurer un horaire cadencé entre cette ligne et les principales autres impliquées. Cela réduit le temps d'attente lors des transferts ce qui diminue la lourdeur des correspondances.

En particulier, dans l'échantillon étudié, l'analyse révèle qu'un horaire cadencé avec les lignes 139 et 67 est envisageable, car elles génèrent elles aussi beaucoup de correspondances. Il s'agit de faire concorder l'arrivée d'un bus de la ligne 141 avec celle d'un bus de la ligne 139 sur le croisement Jean-Talon / Pie-IX et avec celle d'un bus de la ligne 67 sur le croisement Jean-Talon / Saint-Michel.

De la même façon, bien que cela soit plus rare, il est possible d'accorder le départ d'un bus sur la ligne 141 avec l'arrivée en station d'un métro vert à Honoré-Beaugrand et de faire coïncider l'arrivée du bus 141 avec le départ d'un métro bleu à la station Saint-Michel.

4.4.3 Détection des principaux arrêts et stations de correspondances

À l'aide de l'algorithme VALIDEP et des graphes obtenus précédemment, il est dorénavant possible de localiser les principaux arrêts qu'utilisent les voyageurs de l'échantillon pour faire des correspondances BB. De même, les stations de métro peuvent être classées suivant leur taux de correspondances de chaque type (BM, MB et MM). Cette dernière partie vise à décrire où les usagers de l'échantillon favorisent des correspondances. Cela permet d'identifier les arrêts et les stations où améliorer l'expérience du transfert peut avoir un effet bénéfique pour beaucoup d'utilisateurs.

Arrêts principaux des correspondances BB

La figure 4.27 situe sur une carte de Montréal les arrêts utilisés pour correspondre entre les principaux couples de lignes de bus déduits du graphe 4.23. La valeur affichée présente le nombre moyen d'échanges quotidiens observés à partir de l'échantillon de données. Tout d'abord, les arrêts les plus utilisés sont ceux situés dans le centre et dans l'Est de l'Île. Il s'agit d'une zone dense où le métro n'est pas présent. Par conséquent, les déplacements en bus sont plus nombreux, ce qui explique le plus grand nombre de correspondances. Cette observation s'accorde avec le positionnement des groupes proposé à la section 4.3.4.

Ensuite, la répartition des arrêts dessine des axes de bus nord – sud et est – ouest. Le boulevard Saint-Michel, le boulevard Pie-IX, la rue Jean-Talon, le boulevard Henri-Bourassa, le chemin de la Côte-des-Neiges et le boulevard de la Côte-Vertu sont ainsi identifiés. Ce sont de grands axes structurants de l'Île sur lesquels circulent des lignes de bus à haute fréquence permettant le rabattement vers le métro. Cette visualisation laisse supposer que les usagers de l'échantillon ont recours à des correspondances entre deux bus afin d'embarquer dans des lignes permettant de rejoindre le centre-ville ou le réseau de métro.

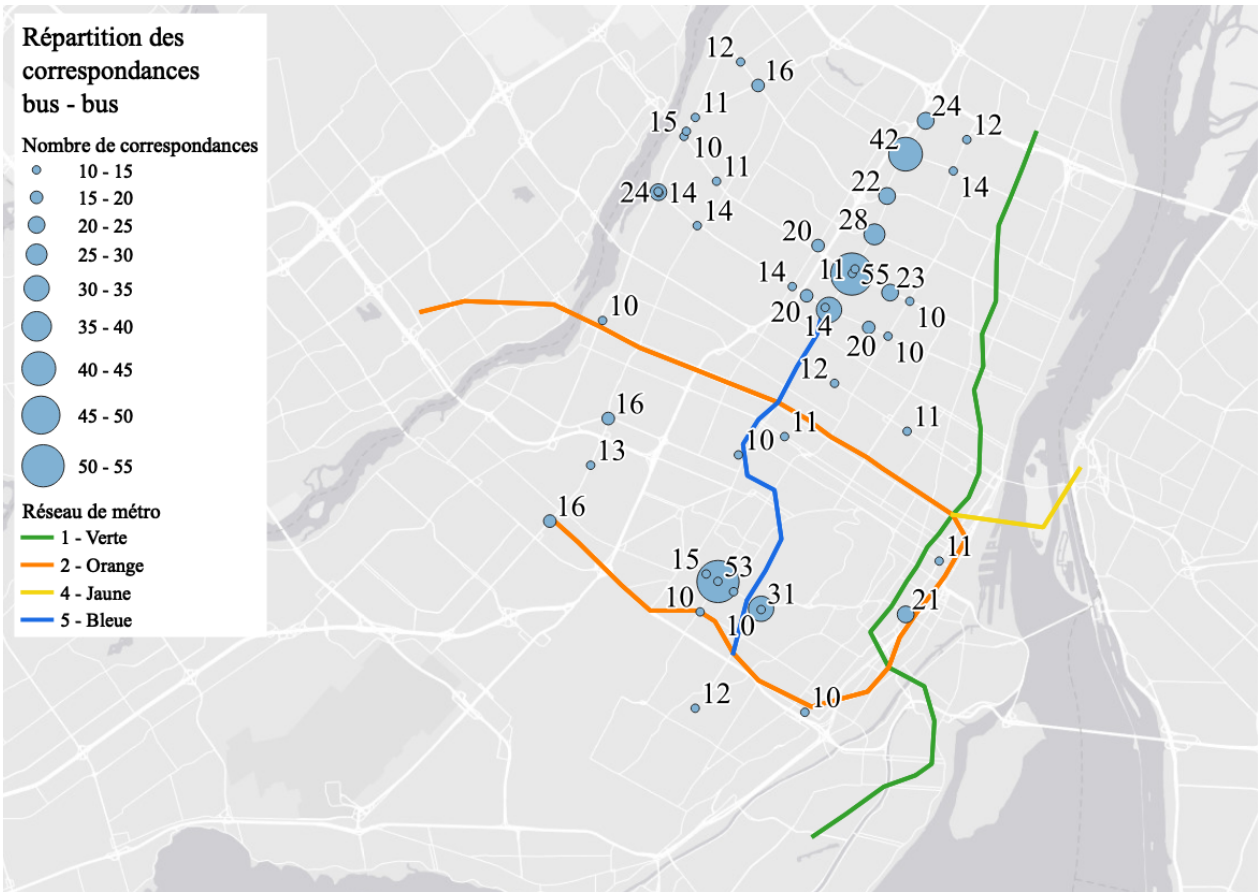


Figure 4.27 Répartition géographique des correspondances bus – bus

L'axe principal qui apparaît avec cette carte est la rue Jean-Talon, à l'est de la station Saint-Michel (terminus de la ligne bleue). Cela signifie que les passagers de l'échantillon correspondent vers la ligne 141, qu'ils empruntent pour rejoindre la ligne bleue du métro. Prolonger la ligne bleue à l'est de la station Saint-Michel permettra d'offrir un meilleur accès au métro pour les habitants du nord-est de Montréal, leur permettant de se rendre au centre-ville plus facilement. Ce prolongement fait partie des grands projets de la STM en 2020, ainsi que la mise en place d'un service rapide par bus (SRB) sur le boulevard Pie-IX.

Pour terminer, cette visualisation permet d'identifier les arrêts où beaucoup d'usagers de l'échantillon transitent entre deux bus. Les croisements Van-Horne / Côte-des-Neiges, Jean-Talon / Pie-IX et Jean-Talon / Langelier sont les plus gros nœuds de correspondance. Pour améliorer le service, il faut concentrer les actions sur les arrêts situés dans ces croisements. En particulier, sécuriser les passages piétons, proposer des espaces d'attente couverts et fournir de l'information en temps réel peut diminuer la lourdeur des correspondances et améliorer la qualité perçue par les usagers. De même, ajouter des services sur ces secteurs sera bénéfique.

Usage des stations de métro pour les correspondances BM et MB

De même que précédemment, les figures 4.28 et 4.29 illustrent la proportion respective de correspondances BM et MB dans chaque station de métro observée à partir de l'échantillon d'utilisateurs. En premier, il s'observe que la répartition des correspondances des deux types est proche. La symétrie des déplacements pendulaires étudiée à la section 4.2.3 permet de justifier cette similarité. La différence s'explique par le fait que le lieu des correspondances MB est estimé par l'algorithme VALIDEP à l'aide d'un calcul de proximité des lignes, alors que le lieu des correspondances BM est précis.

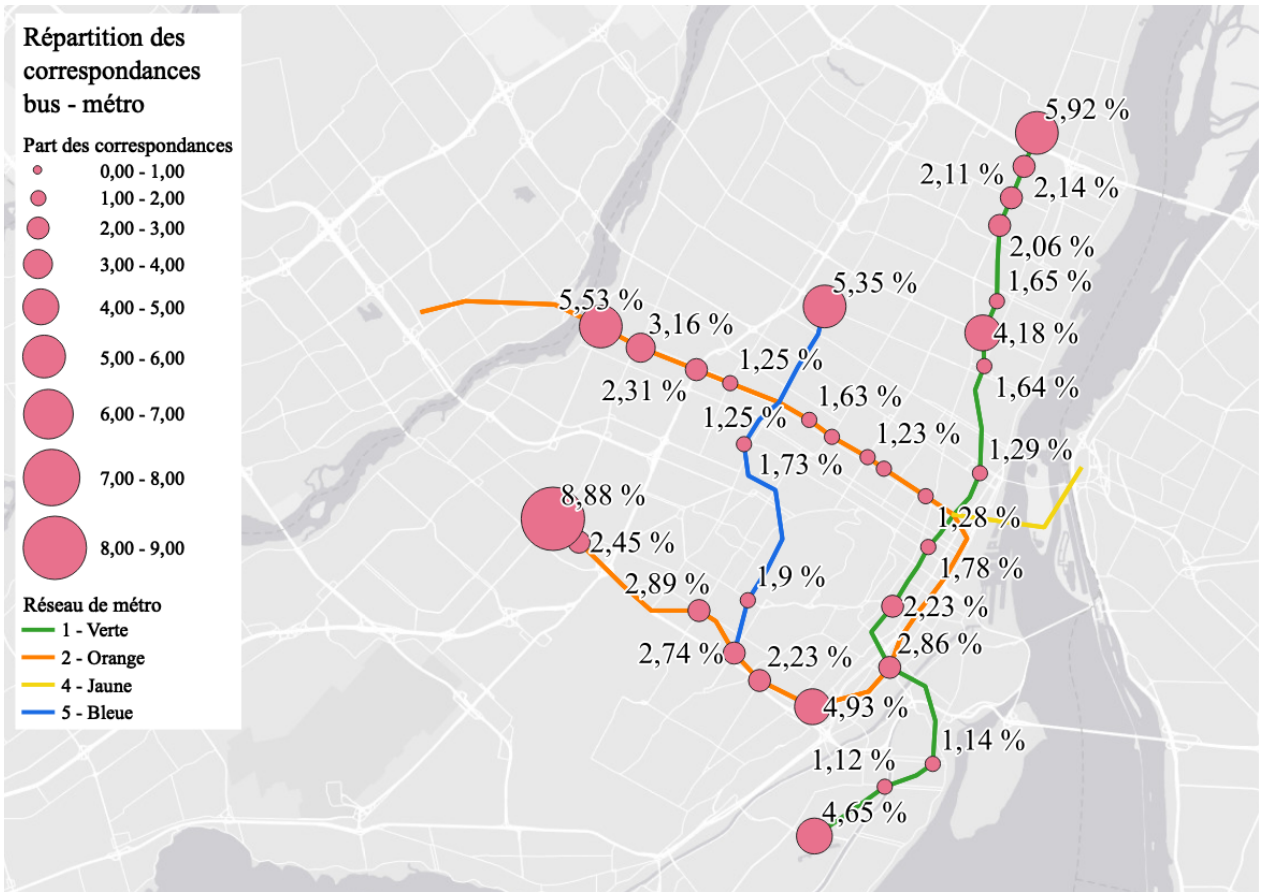


Figure 4.28 Répartition géographique des correspondances bus – métro

À partir de l'échantillon étudié, les deux figures montrent que les stations où il y a le plus de correspondances avec un bus sont celles situées aux extrémités des lignes de métro. Certaines stations situées sur un boulevard structurant sont également très impliquées. C'est le cas par exemple des stations Pie-IX ou Henri-Bourassa. Autrement dit, dans l'échantillon étudié, l'analyse révèle que les usagers de l'Est et l'Ouest utilisent le bus pour rejoindre le métro à un

terminus. Quant à eux, les voyageurs originaires du centre favorisent les lignes de rabattement des grands boulevards pour aller prendre le métro au cœur du réseau.

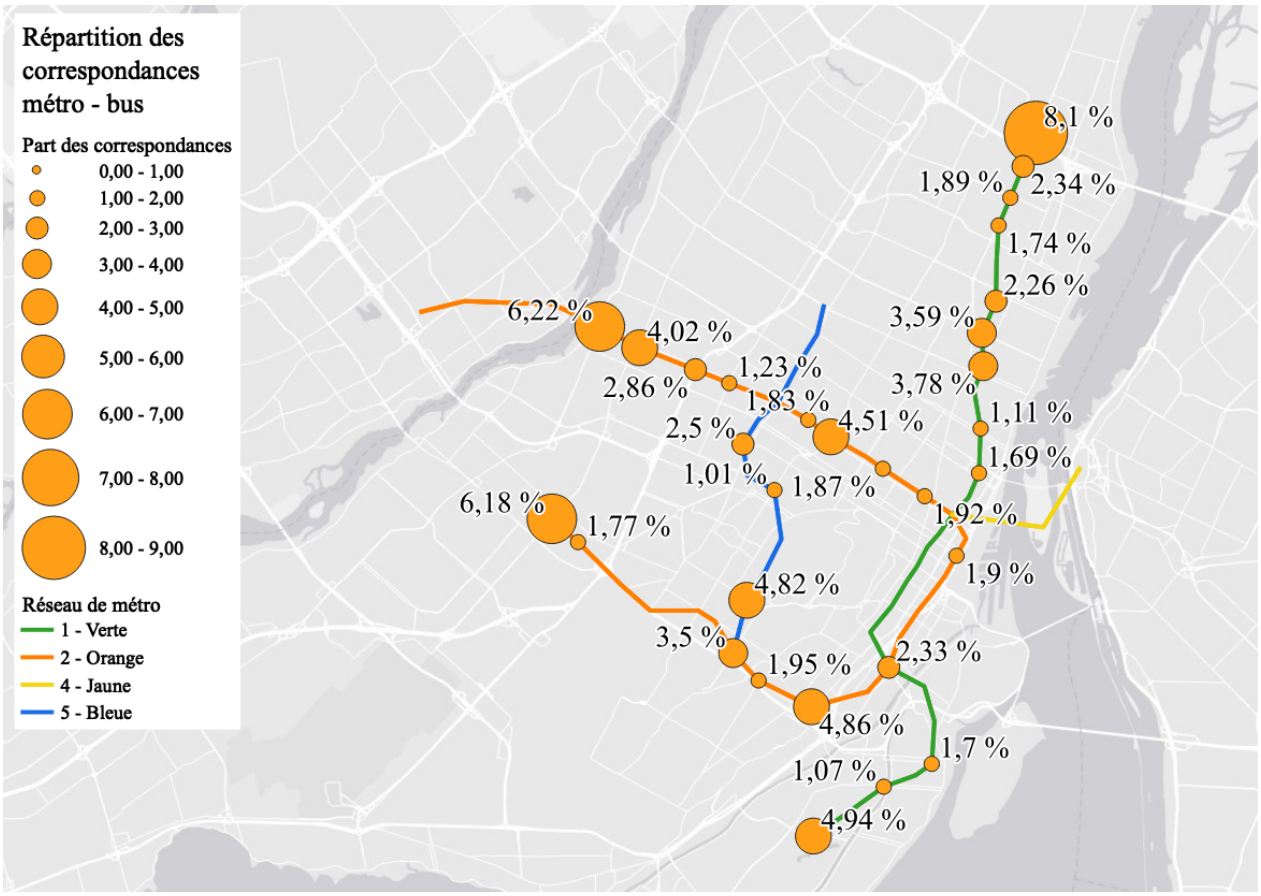


Figure 4.29 Répartition géographique des correspondances métro – bus

Dans l'échantillon, les stations de la ligne bleue sont moins utilisées pour correspondre. Deux explications peuvent être données. Premièrement, les lignes de rabattement qui se dirigent vers la ligne orange sont plus nombreuses que celles allant vers la ligne bleue. Deuxièmement, pour quitter le centre depuis la ligne bleue, il est plus facile de rester dans le métro pour aller prendre la ligne orange que d'utiliser des bus. Pour les mêmes raisons, il s'observe très peu de correspondances BM et MB dans les stations du centre-ville.

Usage des stations de métro pour les correspondances MM

Pour terminer, la figure 4.30 illustre, la distribution des correspondances MM sur les quatre stations du réseau permettant de changer de ligne. Dans l'échantillon analysé, Lionel-Groulx est la station la plus utilisée, suivie de Berri-UQAM, Jean-Talon et Snowdon.

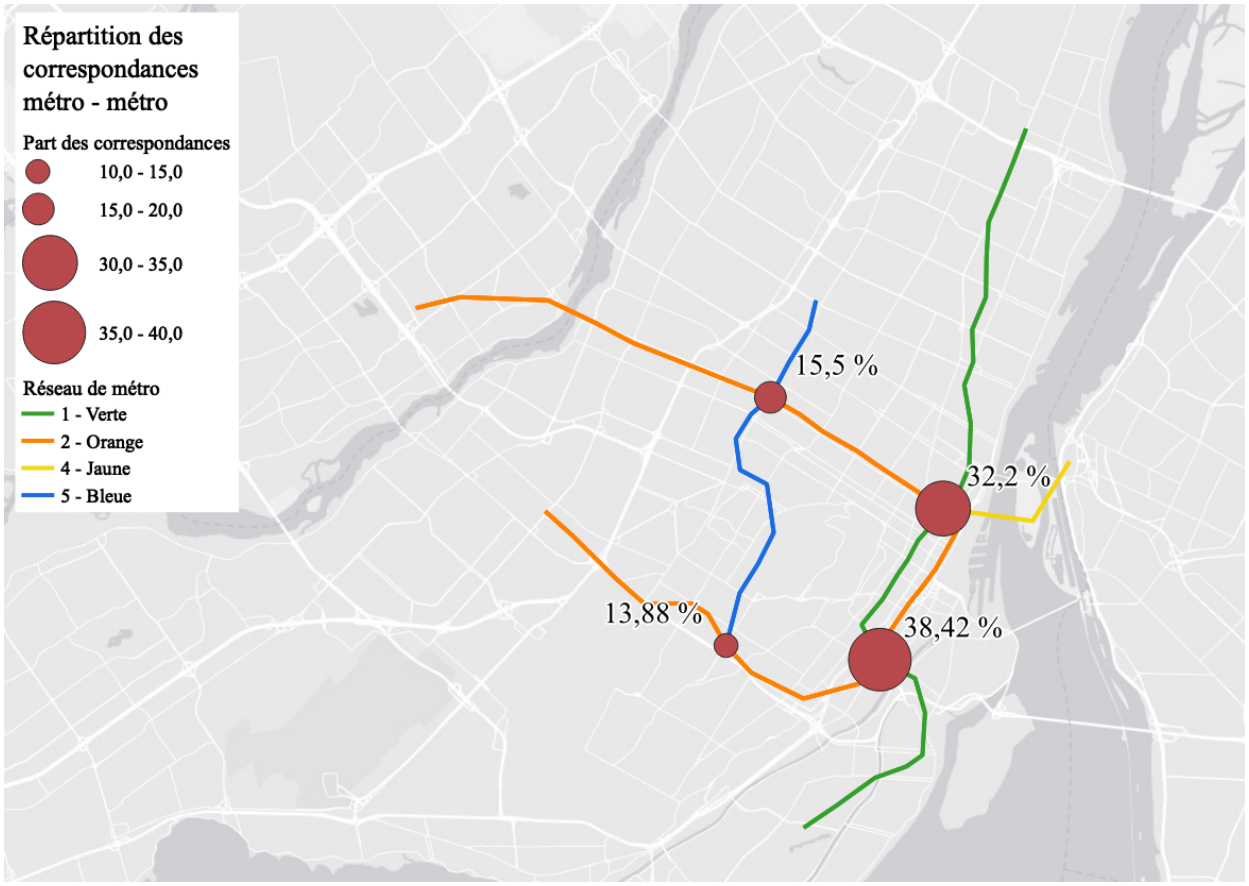


Figure 4.30 Répartition géographique des correspondances métro – métro

La plus faible proportion de correspondances dans les stations Jean-Talon et Snowdon s'explique par l'achalandage moins élevé d'utilisateurs de l'échantillon sur la ligne bleue. Comme il y a moins d'usagers, le nombre de correspondances effectuées est plus faible. De même, la proportion plus grande à Jean-Talon est due à un achalandage supérieur dans la partie Est de la ligne orange que dans la partie Ouest.

Les stations Lionel-Groulx et Beri-UQAM accueillent beaucoup de correspondances MM, notamment car elles lient les deux lignes les plus achalandées de l'échantillon et encadrent le centre-ville. La différence de proportion entre les deux stations n'est pas significative puisqu'elle dépend de la matrice de correspondances station / station utilisée dans l'algorithme

VALIDEP. Toutefois, le design de la station Lionel-Groulx, pensé pour faciliter les correspondances, peut expliquer la proportion légèrement supérieure. En revanche, dans cette étude, les usagers du métro en provenance de Longueuil ne sont pas considérés. S'ils l'étaient, la proportion de correspondances sur ces deux stations pourrait être différente. De même, dans toute cette étude, les résultats sont obtenus à partir d'un échantillon aléatoire de données de cartes à puce. Ils permettent de justifier la méthode, mais ne sont pas nécessairement représentatifs de l'utilisation réelle du réseau de la STM.

Finalement, les planificateurs doivent accorder de l'importance à toutes les stations permettant de changer de ligne de métro. En effet, l'amélioration des correspondances MM permet d'offrir une meilleure connectivité, essentielle aux usagers du réseau.

CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS

Le chapitre 5 clôture l'ensemble de ce mémoire dont l'objectif est d'utiliser des données de transactions par cartes à puce pour élaborer une méthode universelle d'analyse de l'utilisation des correspondances. Pour ce faire, l'étude est orientée autour de quatre objectifs spécifiques :

- développer un algorithme de détection, de comptage et de localisation des correspondances afin de transformer les données de transactions en données de déplacements;
- mesurer la variabilité d'utilisation des correspondances au sein des déplacements quotidiens afin de caractériser la pendularité des usagers d'un échantillon;
- élaborer une méthode permettant l'analyse comportementale d'un échantillon d'usagers du transport en commun selon leur utilisation des correspondances ;
- proposer une étude du réseau afin de repérer les lignes qui génèrent des correspondances et dénombrer leurs interactions.

La conclusion commence par la synthèse des travaux réalisés afin de lister les contributions de cette maîtrise. Ensuite, les limitations associées à la méthode et son application sont énoncées pour finalement proposer de nouvelles perspectives de recherche.

5.1 Synthèse des travaux

Dans un premier temps, une revue de littérature permet de définir l'ensemble des notions essentielles de cette recherche. Elle met en avant le potentiel d'utilisation des données de cartes à puce à des fins de planification du transport en commun. Ces données ont l'avantage d'être spatio-temporelles et de bonne qualité même si leurs traitements nécessitent le développement de nouveaux algorithmes. La revue de littérature explique que la correspondance contraint les usagers dans leurs déplacements en transport en commun. L'information disponible, les temps de marche et d'attente ainsi que la sécurité sont les éléments les plus pénalisants pour les voyageurs. Toutefois, la revue montre que les études spatio-temporelles de l'usage des correspondances sont encore peu développées. C'est pour cela que ce projet est réalisé avec des données de cartes à puce. Enfin, la revue de littérature expose les éléments méthodologiques nécessaires pour la conception de la méthodologie, justifiant notamment le choix de la méthode des k-moyennes pour caractériser le comportement des usagers.

Dans un second temps, l'ensemble de la méthodologie est explicitée. Elle commence par la présentation du contexte de Montréal et expose la composition des données utilisées pour ce

projet, 10 000 cartes à puce d'usagers de la Société de transport de Montréal (STM) ayant des abonnements mensuels ou annuels durant l'année 2016, sélectionnées aléatoirement. L'algorithme VALIDEP est ensuite expliqué étape par étape. Il permet de détecter, compter et localiser les correspondances faites pendant les déplacements. Pour continuer, deux éléments méthodologiques sont établis. Le premier est une typologie des lignes qui permet de les caractériser, à la fois suivant leur offre de service, leur localisation et leurs potentielles interactions avec les autres lignes du réseau. Le deuxième est une catégorisation des déplacements à l'aide de leur série de modes. C'est un attribut qui décrit le nombre, le type et l'ordre des correspondances qui constituent un déplacement. Trois méthodes sont ensuite expliquées, permettant d'apporter une réponse à chacun des objectifs spécifiques du projet. La première présente comment les données sont agrégées dans le but de décrire la répartition de l'utilisation des correspondances au cours d'une journée et illustre, avec deux arbres de décisions, l'analyse de la symétrie d'utilisation du réseau chez les passagers pendulaires de l'échantillon étudié. La seconde détaille l'approche utilisée pour caractériser le comportement des usagers de l'échantillon. La méthode des k-moyennes est choisie car elle est parfaitement adaptée aux données du problème. Enfin, la partie sur l'analyse de l'utilisation du réseau expose les équations permettant le calcul de l'indicateur d'autonomie de ligne, la méthode de représentation des interactions sous forme de graphes et la technique de localisation des arrêts et des stations où il y a le plus de correspondances.

Finalement, cette méthodologie est expérimentée sur un échantillon aléatoire de données de cartes à puce de la STM. Sur cet échantillon, l'algorithme VALIDEP a la capacité de traiter 97,48 % des données de transactions, fournissant une table composée de 5 887 074 déplacements et détectant 4 983 278 correspondances. L'analyse des répartitions annuelle et quotidienne des déplacements obtenus en comparaison à l'ensemble des données de la STM justifie l'efficacité et la fiabilité de cet algorithme. Cette table est alors utilisée pour appliquer les trois méthodes développées. Toutefois, l'échantillon de données est utilisé pour présenter et valider les méthodes mises en place, mais n'est pas nécessairement représentatif de l'utilisation du réseau de la STM.

En premier, l'étude de la répartition des correspondances dans les déplacements quotidiens montre que les voyageurs de l'échantillon n'ont pas la même utilisation du réseau selon l'heure de la journée. En particulier, bien que les deux modes soient autant utilisés, l'analyse rapporte que pendant les heures de pointe les voyageurs favorisent le métro et les déplacements avec une correspondance, alors qu'ils favorisent le bus et les trajets directs en dehors des pointes. De plus, la faible proportion de déplacements avec plus d'une correspondance traduit le désintérêt des voyageurs de l'échantillon pour ce type de trajet. Comme il s'observe

une inversion des tendances d'utilisation des correspondances entre le matin et le soir, l'analyse de la symétrie des déplacements est réalisée. Elle montre sur l'échantillon étudié que le transport en commun génère des activités de courtes et de longues durées et que les usagers réalisent majoritairement deux ou trois déplacements par jours. En restreignant la table aux cartes-jours faisant deux déplacements les jours ouvrés, l'analyse montre que la majorité des passagers se rendent et reviennent d'une activité à l'aide d'un aller-retour avec le même itinéraire dans les deux sens. Le métro et les lignes hautes fréquences du centre-ville sont les lignes les plus utilisées par ses voyageurs. Enfin, dans leurs déplacements allers-retours, les utilisateurs de l'échantillon accordent plus de flexibilité à l'usage du bus que celui du métro.

Ensuite, la segmentation des cartes-jours décrit sept groupes de comportements distincts. L'analyse commence par la caractérisation des groupes et explique que, sur l'échantillon étudié, les usagers ayant accès à des trajets directs sont plus stables que ceux faisant des déplacements avec une correspondance. Les voyageurs privilégient davantage un mode de transport plutôt qu'un type de déplacement. La matrice de probabilité de changement de groupe exprime une régularité plus importante chez les usagers du métro, mais avec des changements moins prévisibles. À l'inverse, dans cet échantillon de données, les usagers du bus changent plus souvent de groupe, mais sans changer de mode. Ils alternent entre les séries "B" et "B,B". Ensuite, comme la segmentation est réalisée en utilisant des données quotidiennes, il devient possible de regrouper les comportements en fonction des jours de la semaine. La taille moyenne des groupes selon le jour met en avant une régularité de comportement, mais illustre une différence notable entre les jours ouvrés et ceux de fins de semaine. De plus, le diagramme de Sankey illustrant les flux quotidiens des cartes entre chaque groupe confirme les observations apportées par la matrice de probabilité de changement et explique que les week-ends, les usagers de l'échantillon utilisent davantage les bus et les trajets directs. Pour terminer, l'analyse du positionnement géographique des groupes explique que dans l'échantillon étudié, les passagers favorisant le bus sont répartis partout sur le territoire alors que les usagers en métro ont tendance à rester dans les régions centrales de l'Île.

Enfin, l'utilisation des lignes du réseau est analysée à l'aide du calcul d'un indice d'autonomie des lignes. Il permet de les classer selon leur taux de correspondances afin d'identifier celles qui nécessitent une attention particulière. L'étude appliquée aux données de l'échantillon montre que cet indicateur dépend à la fois du service et de la géolocalisation des lignes. Celles dont l'autonomie est grande sont caractérisées par un service spécifique, une large couverture spatiale ou par peu d'interactions avec le reste du réseau. En moyenne sur l'échantillon, les lignes de bus sont plus autonomes que les lignes de métro. De plus, une régression linéaire multiple prouve que l'indice d'autonomie d'une ligne peut être expliqué à l'aide des variables

de sa typologie. Cela permet à la fois d'identifier les lignes à prioriser pour augmenter l'offre de service et d'estimer l'influence des futurs itinéraires proposés. Ensuite, des graphes sont utilisés pour observer les interactions entre les couples de lignes. Ils peuvent être tracés selon plusieurs échelles temporelles et divers attributs. En particulier, la séparation des types de correspondances en plusieurs graphes montre que sur l'échantillon étudié, les lignes hautes fréquences du centre-ville sont les plus concernées par des correspondances. La vision des interactions entre les bus et les métros rend possible le choix des lignes nécessitant une amélioration. Des solutions telles que la mise en place d'un horaire cadencé entre des lignes ou l'augmentation de la fréquence sont alors envisageables. Enfin, les graphes permettent de détecter les arrêts de bus et les stations où il y a le plus de correspondances. Dans l'échantillon étudié, l'analyse révèle que les correspondances BB sont principalement réparties sur les arrêts des boulevards structurants de l'Île, montrant la force des lignes de rabattements vers le métro. De même, les stations qui génèrent le plus de correspondances avec un métro sont celles des terminus et celles situées sur les grands boulevards.

En conclusion, à l'aide du développement de l'algorithme VALIDEP et l'expérimentation de trois analyses, ce projet apporte une nouvelle méthodologie d'étude de l'utilisation des correspondances sur un réseau de transport en commun avec des données de cartes à puce. Cette étude permet de caractériser à la fois la pendularité des voyageurs à l'aide de l'analyse quotidienne de l'utilisation des correspondances, le comportement des usagers sur toute la période disponible et l'utilisation des lignes du réseau en fonction des correspondances qu'elles génèrent. Finalement, la force de cette méthodologie est sa possibilité d'être appliquée à tout réseau de transport doté d'un système de perception des titres par cartes à puce. Elle peut être adaptée à toute taille de réseau, peut accueillir de nouveaux modes de transport (tram ou train par exemple) et peut fonctionner quelle que soit la période désirée.

Avec ce projet, les planificateurs reçoivent des connaissances sur les agissements des usagers face aux correspondances. Cela permet de formuler de nouvelles propositions pour améliorer le service offert et augmenter la qualité perçue par les voyageurs. La réponse aux trois derniers objectifs permet de proposer une liste non exhaustive de solutions :

- favoriser l'implantation de lignes haute fréquence pendant les heures de pointe pour aider les usagers dans leurs déplacements allers-retours;
- proposer une nouvelle grille de tarifaire basée sur l'utilisation des correspondances; abonnements autorisant uniquement l'accès au bus, au métro, etc.;
- ajouter des lignes dans les secteurs où il y a le plus de correspondances, notamment le long des boulevards structurants de la ville.

5.2 Limitations

Les travaux réalisés dans le cadre de ce mémoire comportent plusieurs limitations. Celles-ci proviennent principalement des données utilisées et de la méthode mise en place. Cette partie liste les principales limitations de ce projet en suivant l'ordre des études réalisées.

Algorithme VALIDEP

L'algorithme VALIDEP est utilisé pour traiter les données brutes de cartes à puce afin de regrouper les transactions en déplacements et de détecter les correspondances. Toutefois, les données brutes manquent d'informations et peuvent contenir des erreurs, rendant le développement de l'algorithme complexe et conduisant à plusieurs potentiels biais dans les résultats. Tout d'abord, les données de cartes à puce utilisées ne fournissent pas d'information sur la localisation des embarquements en bus. Cela est problématique à la fois, car il n'est pas possible de localiser les correspondances bus – bus et, car cela oblige de faire des approximations pour localiser les correspondances métro – bus. De même, l'absence de données sur la sortie d'un mode complexifie l'étude. L'application de la méthode sur un réseau "tap-in / tap-out" permettrait de régler ces problèmes.

Ensuite, la supposition de la station de sortie du métro est une étape indispensable pour localiser les correspondances. Cependant, cette supposition repose sur l'hypothèse de He et Trépanier (2015) [19] consistant à dire que le lieu de sortie d'une ligne est le plus proche arrêt possible du lieu d'entrée dans la ligne suivante. Or, ce n'est pas toujours le cas, à la fois car les lieux d'embarquements en bus sont manquants et, car un usager peut se déplacer autrement qu'en TC pendant la journée. De plus, cette méthode ne peut s'appliquer aux transactions solitaires en métro. Une partie des données initiales est alors indisponible dans la table des déplacements. Dans ce projet, il s'agit donc d'une estimation du nombre de correspondances, pas d'une réalité.

Enfin, la détection des correspondances métro – métro passe par l'utilisation d'une matrice indiquant pour chaque couple de stations entrée / sortie le nombre de lignes à prendre et le lieu des correspondances à faire. Cependant, pour certains couples, plusieurs itinéraires sont possibles. La matrice choisit l'itinéraire avec l'hypothèse suivante : le passager minimise le nombre de correspondances et de stations à parcourir. Cette hypothèse peut conduire à un biais dans l'estimation du nombre et du lieu des correspondances entre deux métros. Toutefois, le nombre de déplacements est correct même s'il y a une erreur dans le nombre de correspondances détectées. La connaissance du couple OD de chaque déplacement permettrait d'améliorer la détection des stations de correspondance.

L'algorithme VALIDEP est conçu dans l'objectif d'apporter une solution à un maximum des problèmes potentiels qui apparaissent en traitement des données de cartes à puce. Cela fait de lui un outil performant, précis et évolutif qui peut s'adapter à plusieurs autres réseaux de transport. Toutefois, des modifications sont obligatoires. Le format des données brutes doit être adapté, les informations propres aux lignes doivent être ajoutées et une nouvelle matrice de correspondances station / station doit être générée. Sur un réseau avec beaucoup de lignes de métro (Paris par exemple), la création de la matrice de correspondances nécessitera le développement de nouveaux outils de calculs, tels qu'un algorithme du plus court chemin.

Distribution des correspondances et symétrie des déplacements

L'étude de la distribution des séries de modes vise à expliquer comment et quand les usagers d'un échantillon ont recours à une correspondance. Cependant, cette étude est limitée par l'absence d'information sur l'heure des correspondances. En effet, à cause du manque de localisation des embarquements et des descentes, il n'est pas possible de prédire avec précision l'heure d'une correspondance. Ainsi, seule l'heure de début du déplacement peut être connue. De même, il n'est pas possible de connaître la durée du déplacement.

Ce manque d'information conduit à des limitations dans l'étude des symétries. Tout d'abord, les données de cartes à puce ne permettent pas de connaître le motif des déplacements et la durée des activités générées. De plus, les déplacements se faisant sans l'usage du transport en commun ne sont pas visibles. Ainsi, la répartition des durées d'activités pourrait varier avec l'intégration d'autres moyens de transport. Enfin, pour simplifier l'étude, l'analyse des symétries se fait uniquement sur les cartes-jours avec deux déplacements pendant les heures de travail. Est-il possible de réaliser cette même étude sur toutes les cartes, quel que soit leur nombre de déplacements ?

Segmentation des usagers

La segmentation des passagers utilise la méthode des k-moyennes. La limitation principale de cette méthode repose sur le choix du nombre de groupes k. Dans cette étude, $k = 6$ et un septième groupe est ajouté a posteriori. Cependant, choisir un autre nombre de groupes pourrait apporter des résultats et des observations différentes. De même, le choix du vecteur de la segmentation a un impact. Y a-t-il d'autres éléments à prendre en compte avant la segmentation ? Finalement, la méthode rend difficile l'association des comportements observés avec des données exogènes qui ne sont pas considérées dans le modèle. Les comportements

des usagers et leurs changements sont visibles, mais expliquer ces variations est complexe. Par exemple, la segmentation ne peut pas prendre en compte les titres de transport utilisés par les usagers ou leur localisation.

Utilisation du réseau

L'analyse de l'utilisation des lignes permet de détecter celles qui sont importantes pour améliorer la connectivité du réseau. Cependant, cette étude est limitée par l'absence de localisation des correspondances. En effet, dans certains cas les interactions entre deux lignes sont possibles à plusieurs arrêts. Pour obtenir plus d'informations, il faudrait que le modèle puisse détecter avec précision où a lieu chaque correspondance BB et MB. Enfin, dans cette étude ce sont les interactions moyennes qui sont étudiées. Mais les résultats pourraient être différents si les données étaient agrégées sur chaque jour ou chaque heure séparément.

La typologie des lignes développée permet aussi d'apporter de l'information sur l'utilisation du réseau. Cependant, le manque d'information sur les lignes de métro rend impossible l'application de la méthode de prédiction de l'indice d'autonomie par la régression linéaire multiple sur ces lignes. Pour améliorer la prédiction, il faudrait à la fois intégrer les lignes de métro dans le modèle et enrichir la typologie en ajoutant d'autres informations, telles que la longueur de la ligne ou sa priorité de passage. Enfin, à l'heure actuelle, la mesure de l'indice d'autonomie ne permet pas de distinguer les lignes ayant une bonne couverture spatiale des lignes qui ont peu d'interactions avec le reste du réseau.

Pour terminer, l'usage des graphes permet de visualiser les couples de lignes les plus utilisés pour correspondre. Cependant, cet outil est basé sur le nombre moyen de correspondances entre deux lignes. Les lignes très achalandées sont alors mises en avant puisque même avec un faible taux de correspondances, elles génèrent plus de transferts que des lignes avec une capacité plus faible. Pour éviter cela, il faudrait utiliser un indice normalisé permettant de travailler sur un volume de correspondances indépendamment de la capacité des lignes.

Autre limitation

Dans l'ensemble de ce projet, les expérimentations sont réalisées à l'aide d'un jeu de données de cartes à puce. Ces cartes sont sélectionnées aléatoirement suivant certaines conditions. Les données permettent ainsi de tester et de valider la méthodologie développée, ce qui apporte de l'information sur le fonctionnement d'un groupe d'usagers du TC. Cependant, ce jeu de données n'est pas nécessairement représentatif de l'usage réel du réseau de la STM.

5.3 Perspectives de recherches

L'étude proposée cherche à constituer une méthodologie universelle solide d'analyse de l'utilisation des correspondances. Toutefois, les limitations précédentes ainsi que d'autres pistes potentielles d'explorations conduisent à de nouvelles perspectives de recherche.

Tout d'abord, dans cette étude, les données de cartes à puce utilisées ne donnent aucun renseignement sur le type de tarif des cartes. Ajouter une analyse des titres de transport durant chaque étape de la méthodologie serait un moyen d'apporter de nouvelles informations. En effet, il s'agirait d'expliquer le lien entre le tarif utilisé par un usager et son recours aux correspondances. Notamment, l'ajout d'une analyse des titres dans la segmentation permettrait d'observer s'il y a une corrélation entre la composition des groupes et le titre de transport de chaque carte-jour. Cela fournirait aussi une détection des activités plus complète et une meilleure explication de l'utilisation symétrique des déplacements. Cet enrichissement permettrait également de formuler des conseils pour l'amélioration de la grille tarifaire des sociétés de transport.

Ensuite, la segmentation telle que réalisée dans ce mémoire fournit des informations sur les comportements quotidiens d'un échantillon d'usagers et leurs variations jour après jour. Une perspective d'amélioration consisterait à poursuivre la segmentation en regroupant les cartes-jours en cartes-semaines. Cela permettrait de caractériser des comportements hebdomadaires typiques chez les voyageurs. De même, enrichir la segmentation avec l'ajout de plusieurs données exogènes (météo, usage des lignes du réseau, etc.) pourrait peut-être expliquer pourquoi une telle composition des groupes est observée. Il est également essentiel de poursuivre les recherches en étudiant plus en profondeur les caractéristiques du *cluster* "Autre". L'objectif est de comprendre ce qui conduit les usagers à adopter un comportement en particulier.

Pour continuer, il paraît intéressant d'améliorer le fonctionnement de l'algorithme VALIDEP afin qu'il puisse intégrer davantage d'informations dans les données générées. En particulier, il s'agit de trouver une solution pour traiter les transactions solitaires en métro, pour calculer l'heure des correspondances et pour améliorer leurs localisations. De plus, adapter l'application du programme VALIDEP à d'autres villes rendra possible la comparaison des résultats. Dans cet objectif, il faut automatiser l'obtention de la matrice des correspondances station / station à l'aide d'un algorithme de calcul de chemin.

Enrichir le calcul de la symétrie des déplacements est également une source d'amélioration du modèle. Par exemple, rendre cette méthode applicable à toutes les cartes-jours, ajouter de l'information spatiale et proposer une solution pour identifier les usagers qui réalisent

des déplacements pendulaires est une perspective de recherche future. De même, intégrer davantage la typologie des lignes au modèle, à la fois pour le repérage des symétries et pour l'analyse de l'utilisation du réseau, permettrait d'ajouter une dimension spatiale de meilleure qualité. Utiliser davantage d'informations spatiales serait un moyen de comprendre si les usagers ont recours aux correspondances par choix ou par contrainte.

En conclusion, l'ensemble de ces travaux doit permettre d'apporter aux planificateurs une meilleure connaissance du réseau et de ses usagers. À terme, il faudrait adapter ces méthodologies pour qu'elles fournissent des coefficients d'utilités nécessaires à l'intégration des correspondances dans les modèles classiques d'affectation.

RÉFÉRENCES

[1] Radio Canada, “Où s’entasse-t-on le plus dans les villes ? la réponse en carte,” 2016. [En ligne]. Disponible : https://ici.radio-canada.ca/nouvelle/809865/metropoles-urbanisation-grandes-villes-monde-carte

[2] V. R. Vuchic, *Urban transit : operations, planning, and economics.* John Wiley & Sons, 2005.

[3] K. M. Shelfer et J. D. Procaccino, “Smart card evolution,” *Association for Computing Machinery. Communications of the ACM*, vol. 45, n°. 7, p. 83–83, 2002.

[4] M.-P. Pelletier, M. Trépanier et C. Morency, “Smart card data use in public transit : A literature review,” *Transportation Research Part C : Emerging Technologies*, vol. 19, n°. 4, p. 557–568, 2011.

[5] P. Mounier-Khun, “Invention de la carte à puce,” Encyclopædia Universalis, 2019. [En ligne]. Disponible : https://www.universalis.fr/encyclopedie/carte-a-puce/

[6] P. T. Blythe, “Improving public transport ticketing through smart cards,” dans *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, vol. 157, n°. 1. Citeseer, 2004, p. 47–54.

[7] M. Bagchi et P. White, “What role for smart-card data from bus systems ?” *Municipal Engineer*, vol. 157, n°. 1, p. 39–46, 2004.

[8] M. Trépanier, “L’exploitation des données de cartes à puce à des fins de planification des transports collectifs urbains,” *Recherche Transports Sécurité*, vol. 28, n°. 2, p. 139–152, 2012.

[9] M. Bagchi et P. R. White, “The potential of public transport smart card data,” *Transport Policy*, vol. 12, n°. 5, p. 464–474, 2005.

[10] M. Munizaga, C. Palma et P. Mora, “Public transport od matrix estimation from smart card payment system data,” dans *Proceedings from 12th World Conference on Transport Research, Lisbon, Paper*, n°. 2988, 2010.

[11] B. Agard, C. Morency et M. Trépanier, “Mining public transport user behaviour from smart card data,” *IFAC Proceedings Volumes*, vol. 39, n°. 3, p. 399–404, 2006.

[12] A.-S. Briand, E. Côme, M. Trépanier et L. Oukhellou, “Analyzing year-to-year changes in public transport passenger behaviour using smart card data,” *Transportation Research Part C : Emerging Technologies*, vol. 79, p. 274–289, 2017.

[13] E. Deschaintres, C. Morency et M. Trépanier, “Analyzing transit user behavior with 51 weeks of smart card data,” *Transportation Research Record*, vol. 2673, n°. 6, p. 33–45, 2019.

[14] M. Trépanier, C. Morency, B. Agard, E. Descoimps et J. Marcotte, “Using smart card data to assess the impacts of weather on public transport user behavior,” dans *Conference on Advanced Systems for Public Transit-CASPT12, Santiago, Chile*, 2012.

[15] F. Devillaine, M. Munizaga et M. Trépanier, “Detection of activities of public transport users by analyzing smart card data,” *Transportation Research Record*, vol. 2276, n°. 1, p. 48–55, 2012.

[16] A. Dubos-Golain, M. Trépanier et C. Morency, “Understanding transit use patterns in montreal,” CIRRELT, Rapport technique, 2017.

[17] M. S. Ghaemi, B. Agard, M. Trépanier et V. Partovi Nia, “A visual segmentation method for temporal smart card data,” *Transportmetrica A : Transport Science*, vol. 13, n°. 5, p. 381–404, 2017.

[18] S. Hao et D.-H. Lee, “Using smart card data to identify critical transfer hubs and characterize transfer time between metro and bus system,” *Transportation Research Board*, 2018.

[19] L. He et M. Trépanier, “Estimating the destination of unlinked trips in transit smart card fare data,” *Transportation Research Record*, vol. 2535, n°. 1, p. 97–104, 2015.

[20] L. He, M. Trépanier et B. Agard, “Evaluating the impacts of a bus-rapid transit on users’ temporal patterns using cross correlation distance and sampled hierarchical clustering applied to smart card data,” *Transportation Research Board*, 2017.

[21] L. He, B. Agard et M. Trépanier, “A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method,” *Transportmetrica A : Transport Science*, p. 1–20, 2018.

[22] M. Hofmann et M. O’Mahony, “Transfer journey identification and analyses from electronic fare collection data,” dans *Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.* IEEE, 2005, p. 34–39.

[23] W. Li, Q. Luo, Q. Cai et X. Zhang, “Using smart card data trimmed by train schedule to analyze metro passenger route choice with synchronous clustering,” *Journal of Advanced Transportation*, vol. 2018, 2018.

[24] C. Morency, M. Trépanier et B. Agard, “Measuring transit use variability with smart-card data,” *Transport Policy*, vol. 14, n°. 3, p. 193–203, 2007.

[25] H. Nishiuchi, T. Todoroki et Y. Kishi, “A fundamental study on evaluation of public transport transfer nodes by data envelop analysis approach using smart card data,” *Transportation Research Procedia*, vol. 6, p. 391–401, 2015.

[26] L. Sun, J. G. Jin, D.-H. Lee et K. W. Axhausen, “Characterizing multimodal transfer time using smart card data : the effect of time, passenger age, crowdedness, and collective pressure,” *Transportation Research Board*, 2015.

[27] Y. Sun et P. M. Schonfeld, “Schedule-based rail transit path-choice estimation using automatic fare collection data,” *Journal of Transportation Engineering*, vol. 142, n°. 1, 2015.

[28] F. Toqué, M. Khouadjia, E. Come, M. Trepanier et L. Oukhellou, “Short & long term forecasting of multimodal transport passenger flows with machine learning methods,” dans *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, p. 560–566.

[29] M. Trépanier, N. Tranchant et R. Chapleau, “Individual trip destination estimation in a transit smart card automated fare collection system,” *Journal of Intelligent Transportation Systems*, vol. 11, n°. 1, p. 1–14, 2007.

[30] M. Trépanier, C. Morency et B. Agard, “Calculation of transit performance measures using smartcard data,” *Journal of Public Transportation*, vol. 12, n°. 1, p. 5, 2009.

[31] M. Trépanier et C. Morency, “Assessing transit loyalty with smart card data,” dans *12th World Conference on Transport Research, Lisbon, Portugal*, 2010.

[32] M. Trépanier, K. M. Habib et C. Morency, “Are transit users loyal ? revelations from a hazard model based on smart card data,” *Canadian Journal of Civil Engineering*, vol. 39, n°. 6, p. 610–618, 2012.

[33] A. Viallard, M. Trépanier et C. Morency, “Assessing the evolution of transit user behavior from smart card data,” *Transportation Research Record*, vol. 2673, n°. 4, p. 184–194, 2019.

[34] B. T. Yen, W.-C. Tseng, C. Mulley, Y.-C. Chiou et M. Burke, “Assessing interchange effects in public transport : A case study of south east queensland, australia,” *Transportation Research procedia*, vol. 25, p. 4019–4037, 2017.

[35] A. Páez, M. Trépanier et C. Morency, “Geodemographic analysis and the identification of potential business partnerships enabled by transit smart cards,” *Transportation Research Part A : Policy and Practice*, vol. 45, n°. 7, p. 640–652, 2011.

[36] E. Deakin et S. Kim, *Transportation technologies : Implications for planning*. UC Berkeley : University of California Transportation Center, 2001, n°. 536.

[37] A. Ceder et S. J. Chowdhury, “Interchange attributes influencing public-transport user’s intention to use routes with transfers,” *Transportation Research Board*, 2012.

[38] American Public Transit Association and others, *Glossary of transit terminology.* Washington, DC : American Public Transit Association, 1994.

[39] H. Laichour, S. Maouche et R. Mandiau, “Traffic control assistance in connection nodes,” dans *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236)*, vol. 2. IEEE, 2001, p. 1355–1360.

[40] X. Wu, Y. Lu, Y. Lin et Y. Yang, “Measuring the destination accessibility of cycling transfer trips in metro station areas : A big data approach,” *International journal of environmental research and public health*, vol. 16, n°. 15, p. 2641, 2019.

[41] H. Iseki et B. D. Taylor, “Not all transfers are created equal : Towards a framework relating transfer connectivity to travel behaviour,” *Transport Reviews*, vol. 29, n°. 6, p. 777–800, 2009.

[42] Z. Guo et N. H. Wilson, “Assessing the cost of transfer inconvenience in public transport systems : A case study of the london underground,” *Transportation Research Part A : Policy and Practice*, vol. 45, n°. 2, p. 91–104, 2011.

[43] B. De Keizer, K. Geurs et G. Haarsman, “Interchanges in timetable design of railways : A closer look at customer resistance to interchange between trains,” p. 1–13, 2012.

[44] D. Lois, A. Monzón et S. Hernández, “Analysis of satisfaction factors at urban transport interchanges : Measuring travellers’ attitudes to information, security and waiting,” *Transport policy*, vol. 67, p. 49–56, 2018.

[45] A. Garcia-Martinez, R. Cascajo, S. R. Jara-Diaz, S. Chowdhury et A. Monzon, “Transfer penalties in multimodal public transport networks,” *Transportation Research Part A : Policy and Practice*, vol. 114, p. 52–66, 2018. [En ligne]. Disponible : http://www.sciencedirect.com/science/article/pii/S0965856417303117

[46] C. A. Cheriyan et S. Sinha, “Assessment of transit transfer experience : Case of bangalore,” *Transportation Research Procedia*, 2019.

[47] F. Jin, E. Yao, Y. Zhang et S. Liu, “Route choice behavior analysis considering the perceived transfer threshold for urban rail transit passengers,” Transportation Research Board, 2017.

[48] A. Ceder, S. Chowdhury, N. Taghipouran et J. Olsen, “Modelling public-transport users’ behaviour at connection point,” *Transport Policy*, vol. 27, p. 112–122, 2013.

[49] P. Bonnel, *Prévoir la demande de tranport*, 1er éd. Presses Ponts et Chaussées, 2004.

[50] H. Iseki et B. D. Taylor, “Style versus service ? an analysis of user perceptions of transit stops and stations,” *Journal of Public Transportation*, vol. 13, n°. 3, p. 2, 2010.

[51] M. Wardman, “Public transport values of time,” *Transport policy*, vol. 11, n°. 4, p. 363–377, 2004.

[52] Y. Hadas et P. Ranjitkar, “Modeling public-transit connectivity with spatial quality-of-transfer measurements,” *Journal of Transport Geography*, vol. 22, p. 137–147, 2012.

[53] S. Chowdhury, A. A. Ceder et B. Schwalger, “The effects of travel time and cost savings on commuters’ decision to travel on public transport routes involving transfers,” *Journal of Transport Geography*, vol. 43, p. 151–159, 2015.

[54] S. Mishra, T. F. Welch et M. K. Jha, “Performance indicators for public transit connectivity in multi-modal transportation networks,” *Transportation Research Part A : Policy and Practice*, vol. 46, n°. 7, p. 1066–1085, 2012.

[55] Y. Hadas, “Assessing public transport systems connectivity based on google transit data,” *Journal of Transport Geography*, vol. 33, p. 105–116, 2013.

[56] A. Ceder et S. Perera, “Detecting and improving public-transit connectivity with case studies of two world sport events,” *Transport Policy*, vol. 33, p. 96–109, 2014.

[57] A. Frappier, C. Morency et M. Trépanier, “Measuring the quality and diversity of transit alternatives,” *Transport Policy*, vol. 61, p. 51–59, 2018.

[58] ARTM, “Programme des immobilisations 2019 – 2028,” 2019. [En ligne]. Disponible : https://www.artm.quebec/wp-content/uploads/2019/03/Programme-des-immobilisations-2019-2028-1.pdf

[59] Société de transport de Montréal, “Budget 2016,” 2016. [En ligne]. Disponible : http://stm.info/sites/default/files/pdf/fr/budget2016.pdf

[60] Société de transport de Montréal, “Budget 2020,” 2020. [En ligne]. Disponible : http://stm.info/sites/default/files/pdf/fr/budget2020.pdf

[61] EXO, “Organismes de transport du grand montréal, qui fait quoi ?” 2018. [En ligne]. Disponible : https://exo.quebec/fr/actualites/nouvelles-evenements/nouvelles/exo-organismes-transport-qui-fait-quoi2?

[62] Société de transport de Montréal, “Règlement r-105 concernant les conditions au regard de la possession et de l’utilisation de tout titre de transport pour les services de transport offerts par ou pour la société de transport de montréal,” 2017. [En ligne]. Disponible : http://www.stm.info/fr/a-propos/gouvernance-d-entreprise/les-reglements-et-politiques/reglement-r-105-artm

[63] A. Viallard, “Méthodologie pour l’étude de l’évolution des comportements des voyageurs de transport collectif urbain,” Mémoire de maîtrise, École Polytechnique de Montréal, 2018.

[64] J. H. Ward Jr, “Hierarchical grouping to optimize an objective function,” *Journal of the American statistical association*, vol. 58, n°. 301, p. 236–244, 1963.

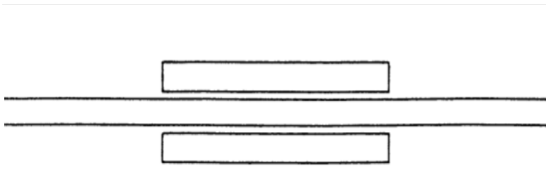
[65] Société de transport de Montréal, “La mobilité des montréalais,” 2016.

[66] Services Québec, “Durée du travail et aménagement du temps de travail,” 2020. [En ligne]. Disponible : http://www4.gouv.qc.ca/FR/Portail/Citoyens/Evenements/travailleur-avec-salaire/Pages/duree-travail-amenagement-temps.aspx

[67] J. C. Brunson, “Alluvial plots in ggplot2,” 2019. [En ligne]. Disponible : https://cran.r-project.org/web/packages/ggalluvial/vignettes/ggalluvial.html

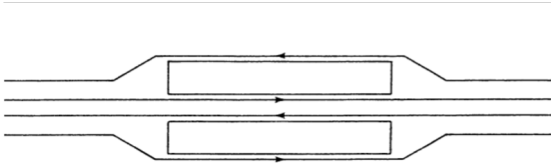
[68] V. R. Vuchic, *Urban transit systems and technology.* John Wiley & Sons, 2007.

ANNEXE A EXEMPLES D'OPTIMISATION DES CORRESPONDANCES

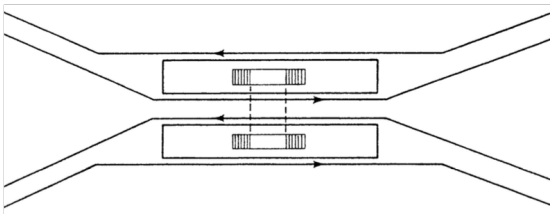


(a) Station conventionnelle

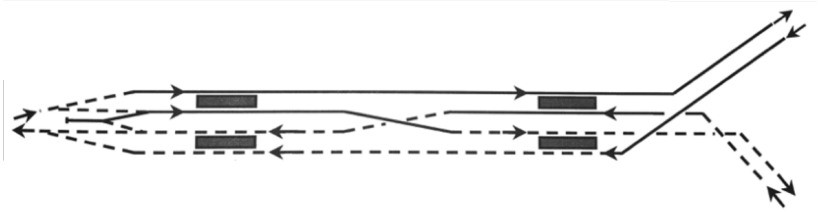
(b) Station à plateforme centrale



(c) Station à deux quais centraux



(d) Stations à deux quais communicants



(e) Double station à deux étages

Figure A.1 Configurations de stations optimisant les correspondances [2, 68]

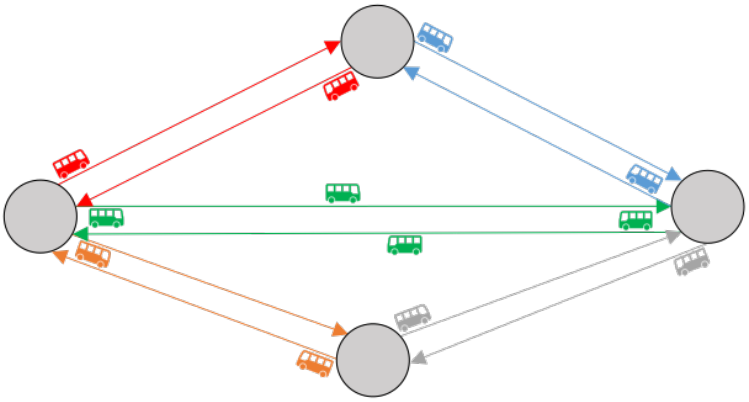
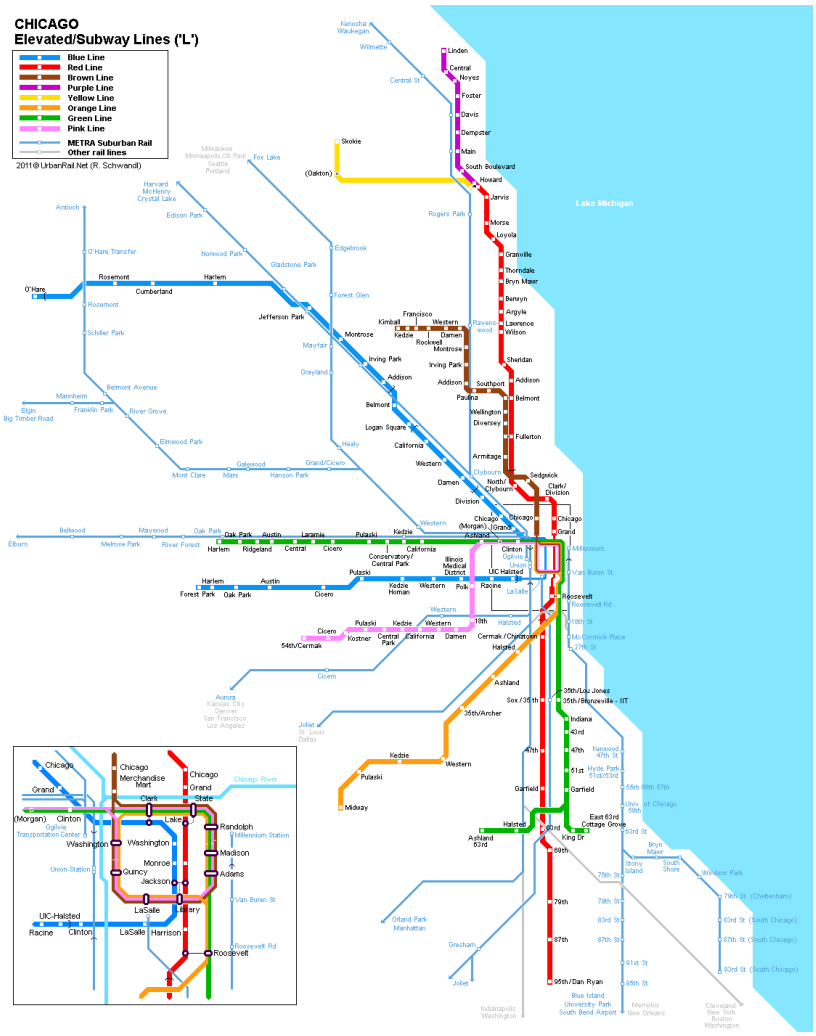
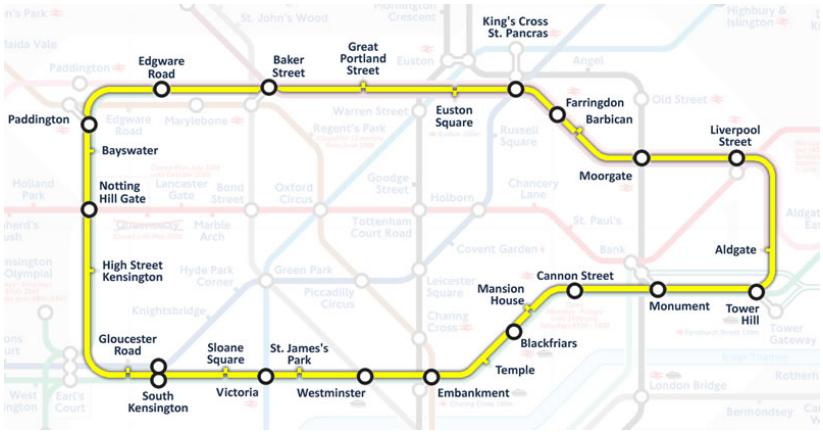


Figure A.2 Horaires cadencés avec quatre points focaux



(a) Plan du métro de Chicago (tiré de http://www.orangesmile.com/chicago.htm)



(b) Ligne circonférentielle du métro de Londres (tiré de http://www.britevents.com/features/a-day-out-on-the-circle-line/)

Figure A.3 Réseaux de métro optimisant les correspondances

ANNEXE B PLAN DU RÉSEAU DE BUS ET DE MÉTRO DE LA STM

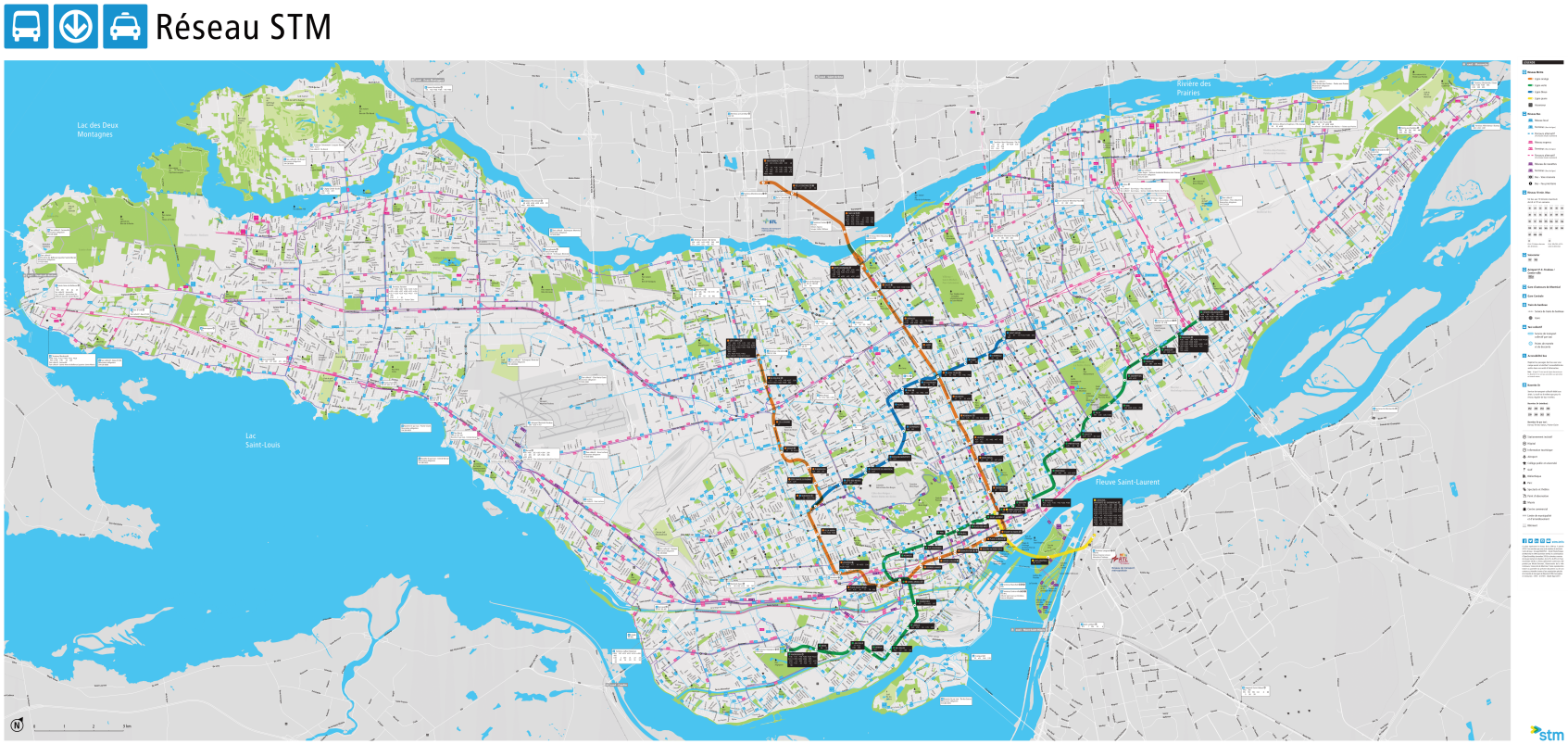


Figure B.1 Plan du réseau de bus et de métro de la STM

Tiré du site officiel de la Société de transport de Montréal : http://www.stm.info/fr/infos/reseaux/plans-des-reseaux.

Figure B.2 Plan du métro de Montréal

Tiré du site officiel de la Société de transport de Montréal :
http://www.stm.info/fr/infos/reseaux/plans-des-reseaux.

ANNEXE C TITRES DE TRANSPORT PROPOSÉS PAR LA STM

(a) Carte OPUS (b) CPO (c) CMJ

Figure C.1 Cartes de transport de la STM

Tiré du site officiel de la Société de transport de Montréal :
http://stm.info/fr/infos/titres-et-tarifs.

Tarifs

En vigueur 1er juillet 2019

BUS ET MÉTRO

TITRES[1]	TARIF ORDINAIRE	TARIF RÉDUIT[2] 6-17 ANS 65 ANS ET +	TARIF RÉDUIT[2] ÉTUDIANTS 18 ANS ET +	À BORD DU BUS	OPUS[3]			VALIDE 747	VALIDE VIA LAVAL • LONGUEUIL	EN VENTE DÉTAILLANTS AUTORISÉS
1 passage	3,50 $	2,50 $[4]	.....	●[4]	●	●[5]	●[5]		●	●
2 passages (Un seul détenteur)	6,50 $	4,25 $[4]	.....		●	●[6]	●[6]		●	●
10 passages	29,00 $	17,50 $	.....		●					●
Soirée illimitée (18 h à 5 h) Première utilisation avant minuit	5,50 $	.....	.....		●	●			●	●
1 jour (24 h)	10,00 $	.....	.....		●	●		●	●	●
Week-end illimité (Du vendredi 16 h au lundi 5 h)	14,00 $	.....	.....		●	●		●	●	●
3 jours (Consécutifs, jusqu'à 23 h 59 le 3e jour)	19,50 $	.....	.....		●	●		●	●	●
Hebdo (Du lundi au dimanche 23 h 59)	26,75 $	16,00 $	.....		●			●		●
Mensuel (Du 1er à la fin du mois)	86,50 $	52,00 $	52,00 $		●			●		●
4 mois (Consécutifs)	.....	204,00 $	204,00 $		●			●		●
Groupe[7] (1 adulte + 10 enfants de 6-13 ans)	.....	17,50 $	.....				●		●	
Aéroport Mtl-Trudeau (747) (Valide 24 h)	10,00 $	.....	.....	●				●	●	
Passage 747 ouest[8]	10,00 $	.....	.....	●				●		

TRANSPORT ADAPTÉ ET SERVICES PAR TAXI

TITRES	TARIF ORDINAIRE	TARIF RÉDUIT[2] 6-17 ANS 65 ANS ET +	TARIF RÉDUIT[2] ÉTUDIANTS 18 ANS ET +	À BORD DU BUS	OPUS		EN VENTE DÉTAILLANTS AUTORISÉS	TRANSPORT ADAPTÉ – DÉPLACEMENTS MÉTROPOLITAINS
Montant exact dans le véhicule[9]	2,90 $	1,75 $	.....	●				
1 passage	3,50 $	2,50 $	.....			●		
Hebdo (Du lundi au dimanche 23 h 59)	26,75 $	16,00 $	.....		●		●	
Mensuel (Du 1er à la fin du mois)	86,50 $	52,00 $	52,00 $		●		●	
4 mois (Consécutifs)	.....	204,00 $	204,00 $		●		●	
Vers les territoires du RTL et de la STL								Double tarification STM
Vers les autres destinations à l'intérieur du territoire délimité par l'ARTM								Triple tarification STM

SORTIES EN FAMILLE

GRATUIT pour les enfants de 5 ans et moins en tout temps et pour les enfants de 6 à 11 ans à certaines périodes dans le cadre du programme Sorties en famille[10].

19504_01

(1) Les titres de transport ne sont pas vendus dans les véhicules. Monnaie exacte requise pour le paiement comptant à bord des bus.
(2) Le client sans carte OPUS avec photo, lorsque celle-ci est exigée, s'expose à une amende plus les frais administratifs.
(3) Des frais d'émission de 6 $ sont exigés pour une carte OPUS avec photo pour les 6-11 ans et 15 $ pour les 12-17 ans, les étudiants de 18 ans et plus, ainsi que les 65 ans et plus. La carte OPUS sans photo coûte 6 $ et un titre doit obligatoirement y être chargé.
(4) Titres 1 passage et 2 passages à tarif réduit : • **6-11 ans :** possibilité de payer comptant sans carte OPUS avec photo. Carte photo requise pour l'achat chez un détaillant autorisé. • **12-17 ans :** tarif réduit sur carte OPUS avec photo seulement. • **65 ans et plus :** possibilité de payer comptant sur présentation de la carte OPUS avec photo.
(5) Le titre 1 passage au tarif ordinaire sur carte *L'occasionnelle* est disponible seulement chez les détaillants autorisés. Auprès du changeur et à la distributrice automatique du métro, ce titre est disponible sur carte magnétique.
(6) Le titre 2 passages au tarif ordinaire est disponible sur *L'occasionnelle* et au tarif réduit sur carte magnétique.
(7) Ce titre donne droit à un passage simultané pour le groupe.
(8) Paiement aux bornes Stationnement de Montréal désignées. Valide en direction de l'aéroport Montréal-Trudeau seulement.
(9) Le comptant à bord est accepté pour les déplacements au transport adapté et les dessertes de navettes Or par taxi. Il n'est pas possible de payer comptant à bord des dessertes par taxi collectif.
(10) Permet à un maximum de 5 enfants de 6 à 11 ans de voyager gratuitement lorsqu'accompagnés d'un adulte détenant un titre de transport valide. Programme en vigueur la fin de semaine (du vendredi 16 h au dimanche à la fin du service), les jours fériés ainsi que du 21 juin au 27 août 2019, du 20 décembre (16 h) 2019 au 5 janvier 2020 et du 28 février (16h) 2020 au 8 mars 2020. Programme non valide à bord des bus de la ligne 747 Aéroport Montréal-Trudeau.

De plus, l'application des tarifs et l'utilisation des titres de transport y correspondant doivent s'effectuer conformément aux dispositions du Règlement R-105. Détails et conditions à stm.info.

Figure C.2 Grille tarifaire de la STM

Tiré du site officiel de la Société de transport de Montréal : http://stm.info/fr/infos/titres-et-tarifs.

ANNEXE D TABLES DES TRANSACTIONS ET DES DÉPLACEMENTS

Tableau D.1 Table des transactions après l'étape 1 de l'algorithme VALIDEP

Attribut	Format	Description	Exemple
id_validation	Entier	Identifiant unique de la transaction	2 846 348 716
date_heure	Balise temporelle	Date et heure de la transaction	2016-01-01 19:50:42
date	Balise temporelle	Date de la transaction	2016-01-01
heure	Balise temporelle	Heure de la transaction	19:50:42
id_carte	Entier	Identifiant unique de chaque carte à puce	232418
code_produit	Caractère	Code du produit (titre de transport)	CAMOP
code_ligne	Entier	Numéro de la ligne utilisée	211
code_direction	Entier	Direction de la ligne utilisée	6
code_station	Entier	Station de métro de la transaction	0
mode	Caractère	Mode de la transaction	B
act_cal	Caractère	Champ d'activité calculée	DC

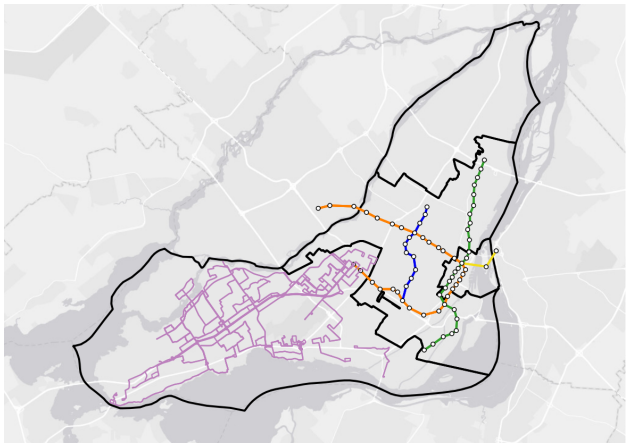
Tableau D.2 Table des transactions enrichies avec les correspondances

Attribut	Format	Description	Exemple
id_dep	Entier	Identifiant unique de chaque déplacement	684
id_validation	Entier	Identifiant unique de la transaction	2 846 348 716
date	Balise temporelle	Date de la transaction	2016-01-01
heure	Balise temporelle	Heure de la transaction	19:50:42
act_cal	Caractère	Champ d'activité calculée	DC
id_carte	Entier	Identifiant unique de chaque carte à puce	232418
code_produit	Caractère	Code du produit (titre de transport)	CAMOP
modes	Caractère	Mode utilisé	B
code_ligne	Entier	Ligne utilisée	211
station_entree	Entier	Station d'entrée dans le métro	0
station_sortie	Entier	Station de sortie du métro	0
BB	Entier	Nombre de correspondances bus - bus	0
BM	Entier	Nombre de correspondances bus - métro	1
BM_loc	Entier	Station de la correspondance bus - métro	36
MB	Entier	Nombre de correspondances métro - bus	0
MB_loc	Entier	Station de la correspondance métro - bus	0
MM	Entier	Nombre de correspondances métro - métro	0
MM_loc_1	Entier	Station de la première correspondance MM	0
MM_loc_2	Entier	Station de la deuxième correspondance MM	0

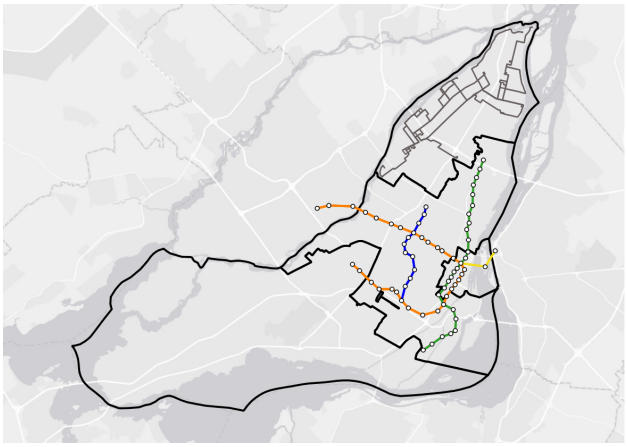
Tableau D.3 Table des déplacements enrichis avec les correspondances

Attribut	Format	Description	Exemple
id_dep	Entier	Identifiant unique de chaque déplacement	684
nb_val	Entier	Nombre de transactions du déplacement	3
date	Balise temporelle	Jour du déplacement	2016-01-01
heure	Balise temporelle	Heure du début du déplacement	19:38:09
id_carte	Entier	Identifiant unique de chaque carte à puce	232418
modes	Caractère	Succession des modes utilisés	B,B,M,M
nb_corresp	Entier	Nombre total de correspondances	3
BB	Entier	Nombre de correspondances bus - bus	1
BM	Entier	Nombre de correspondances bus - métro	1
BM_loc	Entier	Station de la correspondance bus - métro	36
MB	Entier	Nombre de correspondances métro - bus	0
MB_loc	Entier	Station de la correspondance métro - bus	0
MM	Entier	Nombre de correspondances métro - métro	1
MM_loc_1	Entier	Station de la première correspondance MM	49
MM_loc_2	Entier	Station de la deuxième correspondance MM	0

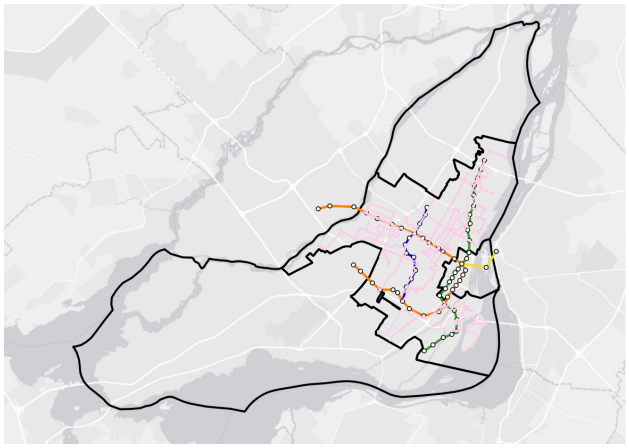
ANNEXE E LIGNES DE CHAQUE COMBINAISON DE RÉGIONS



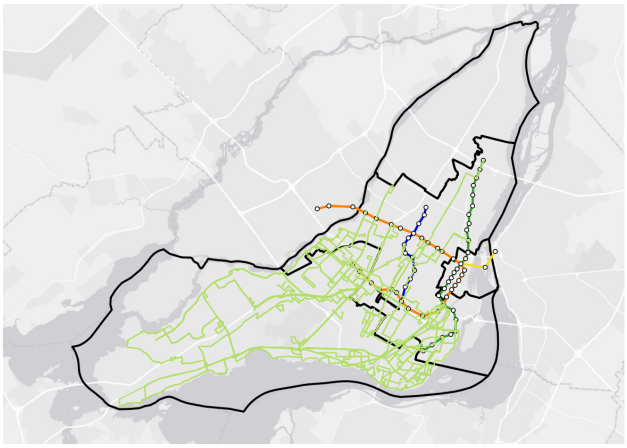
(a) Ouest, repère A



(b) Est, repère B

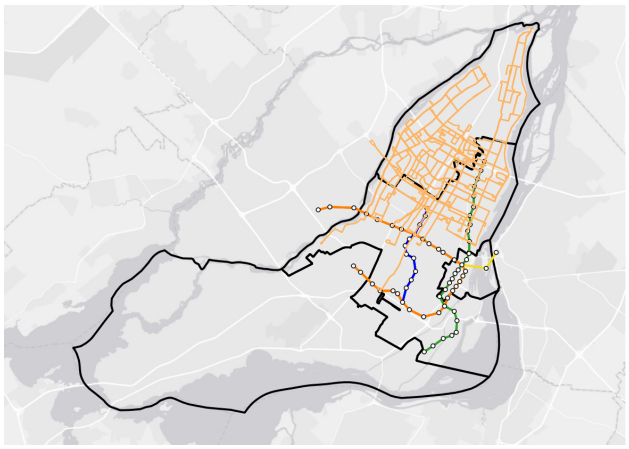


(c) Centre, repère C

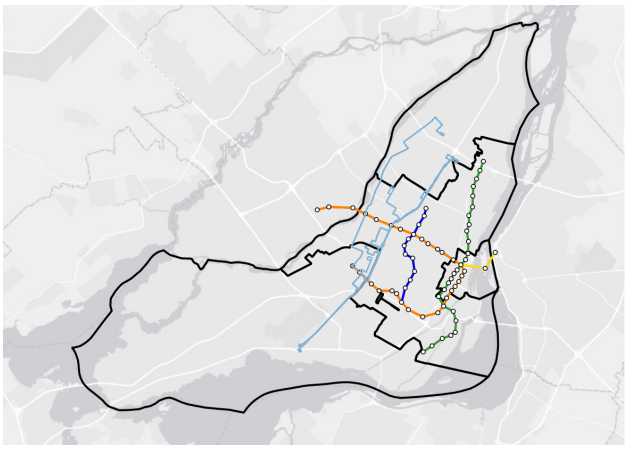


(d) Centre & Ouest, repère D

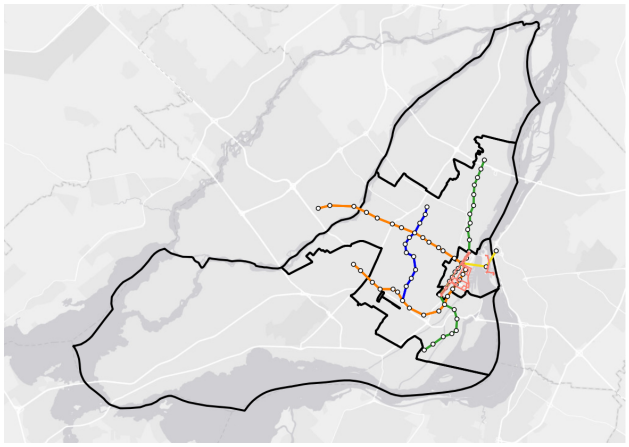
Figure E.1 Lignes selon leur localisation dans les régions de Montréal



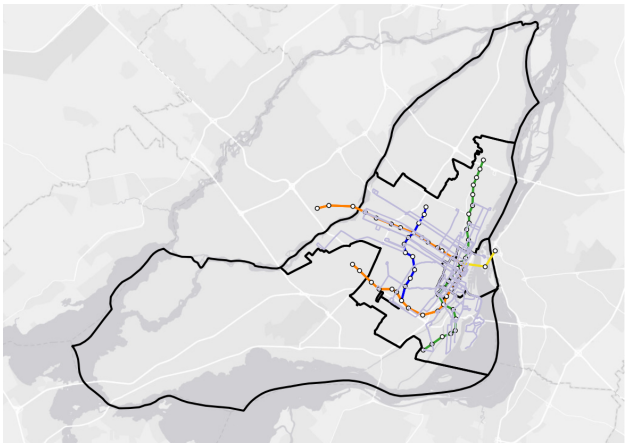
(a) Centre & Est, repère E



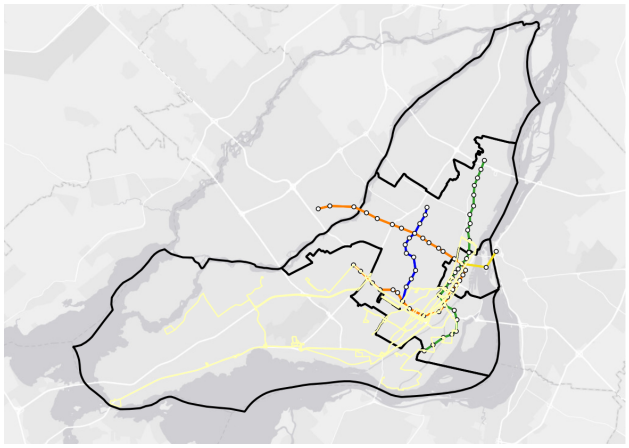
(b) Centre & Ouest & Est, repère F



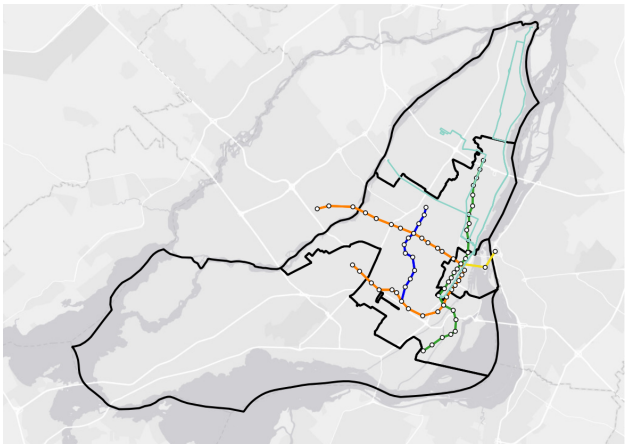
(c) Centre-ville, repère G



(d) Centre-ville & centre, repère H



(e) Centre-ville & centre & Ouest, repère I



(f) Centre-ville & centre & Est, repère J

Figure E.2 Lignes selon leur localisation dans les régions de Montréal (suite)

ANNEXE F LISTE DES JOURS ATYPIQUES DE L'ANNÉE 2016

- Jours fériés [1]
  - † 2016-01-01 : Jour de l'an.
  - † 2016-03-25 : Vendredi Saint.
  - † 2016-03-28 : Lundi de Pacques.
  - † 2016-05-23 : Journée nationale des patriotes.
  - † 2016-06-24 : Fête nationale du Québec.
  - † 2016-07-01 : Fête du Canada.
  - † 2016-09-05 : Fête du travail.
  - † 2016-10-10 : Action de grâce.
  - † 2016-12-25 : Noël.
- Vacances scolaires [1]
  - † Du 2016-02-29 au 2016-03-04 : Relâche.
  - † Du 2016-06-22 au 2016-06-27 : Congé.
  - † Du 2016-07-25 au 2016-08-05 : Fermeture CSDM.
  - † Du 2016-12-26 au 2016-12-31 : Noël.
- Semaines incomplètes en 2016 [1]
  - † Du 2016-01-01 au 2016-01-03 : Semaine 53 de 2015.
    Le premier lundi de 2016 est le 04 janvier. Par conséquent, les trois premiers jours de 2016 sont dans la 53e semaine de 2015.
  - † Du 2016-12-26 au 2016-12-31 : Semaine 52 de 2016.
    La dernière semaine de 2016 inclut le premier dimanche de 2017. Par conséquent, la 52e semaine de 2016 est incomplète.
- Nuit blanche de Montréal [2]
  - † 2016-02-28 : La nuit blanche de Montréal est un événement culturel durant laquelle le service de transport en commun est disponible sans interruption toute la journée. En 2016, la nuit blanche de Montréal s'est déroulée le 28 février (2016-02-28).

1. Calendrier officiel de 2016 – https://www.cspq.gouv.qc.ca/conges-et-vacances.
2. Site officiel de la nuit blanche de Montréal – https://www.nuitblanchemtl.com.