POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Modélisation des interactions entre les modes de transport aux abords des stations de métro

SIMON LEPAGE

Département des génies civil, géologique et des mines (CGM)

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

Génie civil

Novembre 2019

© Simon Lepage, 2019.

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Modélisation des interactions entre les modes de transport aux abords des stations de métro

présenté par **Simon LEPAGE**

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

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

Martin TRÉPANIER, président

Catherine MORENCY, membre et directrice de recherche

Alfred Ka Kee CHU, membre externe

REMERCIEMENTS

Tout d'abord, je tiens à remercier Catherine Morency, ma directrice de recherche, pour son support, ses précieux conseils et les diverses opportunités qu'elle m'a offertes. Son dévouement et son dynamisme sont inspirants. Sa passion pour le domaine du transport est contagieuse!

Je souhaite également remercier mes collègues de bureau qui m'ont partagé leurs connaissances et qui m'ont aidé à résoudre mes problèmes techniques. Merci à Jean-Simon Bourdeau, à Hubert Verreault et à Élodie Deschaintres pour leur aide avec le prétraitement des données et leurs conseils méthodologiques.

Je tiens à remercier le soutien financier de la Chaire de recherche du Canada sur la mobilité des personnes, du conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) et du Fonds québécois de la recherche sur la nature et les technologies (FRQNT). Je tiens également à remercier les partenaires de la Chaire Mobilité, ainsi que les différentes organisations ayant fourni les données, soit la Société de transport de Montréal (STM), Bixi et Taxi Diamond.

RÉSUMÉ

Les modèles de prévision de la demande en transport s'appuient typiquement sur la représentation d'un jour moyen de semaine, en utilisant des données d'enquête regroupant plusieurs jours d'observation. Les données d'enquête ne permettent pas d'expliquer la variabilité d'usage à court terme attribuable à certains évènements occasionnels. De plus, les données d'enquête sont souvent limitées par leur couverture temporelle discontinue et par le nombre restreint de modes considérés.

Cette étude propose plutôt d'utiliser des flux continus de données afin de modéliser la demande de transport. Les données utilisées permettent non seulement d'analyser les variations du niveau d'usage dans le temps, mais aussi dans l'espace. Contrairement à la majorité des recherches qui ne s'intéressent qu'à un seul mode de transport, la présente étude propose l'analyse simultanée de plusieurs modes afin d'en comparer l'usage.

L'objectif général de la recherche est de mieux comprendre les corrélations d'usage entre différents modes de transport. Plus spécifiquement, le projet vise à modéliser le niveau d'usage du métro, du vélopartage, du taxi et du bus aux abords des stations de métro, en utilisant trois années de flux passifs de données transactionnelles. Le territoire à l'étude est celui de la région métropolitaine de Montréal. Les déplacements sont agrégés par voisinage de station de métro afin de simplifier la composante spatiale. Trois différentes analyses sont complétées, soit l'analyse des patrons types d'usage, la création d'une typologie de stations de métro et l'analyse de l'impact d'évènements atypiques sur l'usage des différents modes.

La première analyse portant sur les patrons types d'usage a comme objectif de mieux comprendre les cycles d'utilisation annuels, hebdomadaires et journaliers. Des vecteurs contenant les patrons d'usage sont créés pour chaque combinaison de modes et de stations, puis le nombre de déplacements est normalisé afin de permettre la comparaison de vecteurs ayant des nombres de déplacements d'ordres de grandeur différents. La méthode de partitionnement de données des k-moyennes est utilisée afin de regrouper les patrons similaires. Les résultats démontrent que les patrons annuels d'usage du métro et du bus sont très similaires. De son côté, l'usage du taxi a tendance à être élevé en début d'année et à diminuer tout au long de l'année. Les stations de métro Côte-des-Neiges et Université-de-Montréal ont des patrons annuels qui se distinguent par un usage particulièrement faible durant la période estivale, ce qui est expliqué par les vacances des nombreuses institutions scolaires à proximité. Pour continuer, l'analyse des patrons hebdomadaires

relève que l'usage du vélopartage est généralement aussi élevé la fin de semaine que la semaine durant l'été. Les stations situées dans la partie Ouest du centre-ville et près des universités se distinguent par un achalandage très faible la fin de semaine. Par ailleurs, l'analyse des patrons journaliers démontre que le vélopartage a une pointe du soir plus importante que celle du matin et plus tardive que celle des autres modes. Du côté du taxi, son usage est important le soir, particulièrement pour les stations à l'Est du centre-ville qui sont populaires pour leurs activités sociales. Cela confirme que le métro et le bus sont des modes complémentaires, étant donné que l'usage du bus augmente durant la période de fermeture nocturne du métro.

La deuxième analyse a comme objectif de créer une typologie de stations de métro basée sur les caractéristiques du voisinage des stations. Les stations sont classées dans cinq groupes selon un algorithme hiérarchique de partitionnement de données. Cette typologie de stations est comparée à une seconde classification des stations basée sur les patrons journaliers d'usage, afin de déterminer s'il est possible de prédire le type d'usage d'une station à partir de ses caractéristiques. Les résultats indiquent que les deux classifications ont plusieurs similarités.

La troisième analyse s'intéresse aux fluctuations d'achalandage dues à divers évènements tels que des évènements météorologiques, des interruptions de service du métro et des activités (telles que des concerts ou des évènements sportifs). Les variables ayant un impact significatif sur l'usage sont identifiées indépendamment pour chaque mode à l'aide de modèles de type GAM (*Generalised additive model),* soit un modèle utilisant des fonctions pour transformer certaines variables explicatives. Ensuite, l'impact de ces variables sur l'usage est quantifié. Les résultats suggèrent que l'usage de chaque mode de transport est influencé par la présence de pluie, la température, les interruptions de service et les activités à proximité des stations. L'usage du vélopartage est également influencé par la vitesse du vent. La présence de pluie diminue l'usage du vélopartage, du métro et du bus, alors qu'elle augmente l'usage du taxi. Pour continuer, des modèles sont calibrés pour chaque type de station défini dans la typologie de stations. Deux types de modèles descriptifs sont comparés, soit le modèle GAM et le modèle ARIMA (un modèle autorégressif propre aux séries temporelles). Finalement, des modèles sont calibrés pour chaque station de métro. Les résultats des modèles descriptifs suggèrent que la pluie a un impact moins important sur les stations près du centre-ville. Aussi, les stations proposant des activités touristiques à proximité semblent davantage affectées par la météo. Par ailleurs, l'intensité d'usage est prédite sur différents horizons à l'aide de modèles prédictifs de forme GAM et ARIMA. Le modèle ARIMA semble

mieux performer pour des prédictions à court terme, alors que le modèle GAM performe mieux pour des prédictions à long terme. Le modèle ARIMA semble être trop flexible pour faire des prévisions à long terme, causant du surapprentissage.

Cette étude propose l'analyse simultanée de quatre différents modes de transport à l'aide de flux continus de données, ce qui a très peu été exploré dans la littérature. De plus, l'impact de plusieurs types d'évènements sur l'usage à court terme est évalué dans un seul modèle, alors que les autres études consultées ne prennent en compte qu'un seul type d'évènement à la fois. Cette étude permet de mieux comprendre les habitudes de déplacement de la population montréalaise. Elle permet de considérer un ajustement plus dynamique de l'offre de transport pour correspondre à la demande. Après avoir explicité les contributions, ce mémoire se termine en présentant certaines limites et en proposant quelques perspectives de recherche.

ABSTRACT

Demand modeling is typically based on an average working day, since it often uses data from travel surveys that aggregate multiple days of observation. Also, travel surveys are often performed for a limited time frame rather than a continuous period and neglect low usage transportation modes. Survey data do not allow to capture the effect of occasional events on ridership.

Instead, this study proposes using continuous data flows to model travel demand. The data allows not only to analyze demand variations in time, but also in space. Unlike most research that focus on one mode of transportation only, this study proposes a simultaneous analysis of several modes to compare their demand.

The main objective of this research is to better understand ridership correlations between modes of transportation. More specifically, the project aims to model demand for subway, bikesharing, taxi and bus near subway stations, using three years of transactional data. The area under study is that of the metropolitan area of Montreal. Trips are aggregated by subway station neighborhood to simplify the spatial component. Three different analysis are performed, namely an analysis of ridership patterns, the creation of a typology for subway stations and the impact of events on transportation demand.

The first analysis aims to better understand annual, weekly and daily demand cycles. Vectors containing ridership patterns are created for each combination of mode and station, then demand is normalized. Similar vectors are grouped using k-means grouping method. Results show that subway and bus annual patterns are very similar. For taxi, demand tends to be high at the beginning of the year and to decrease gradually. Subway stations *Côte-des-Neiges* and *Université-de-Montréal* show annual patterns that are characterized by very low demand during the summer period, which can be attributed to summer holidays for many schools nearby. Furthermore, the analysis of weekly patterns indicates that bikesharing demand is generally as high on the weekend as it is on weekdays during summer. Stations located in the western part of downtown and near the *University of Montreal* are characterized by very low demand on weekends. Moreover, the analysis of daily patterns shows that the evening peak for bikesharing is higher than the morning peak, and that it is shifted later in the day. For taxi, demand is high in the evening, especially for stations east of downtown that are popular for their social activities. There seems to be complementarity between subway and bus services, since bus demand increases during night while subway is closed.

The second analysis aims to create a typology of subway stations based on some characteristics of their neighborhood. Stations are classified into five groups using hierarchical grouping. This typology is compared to a second classification of the stations based on daily ridership patterns, in order to determine if the ridership pattern of a station can be estimated based on its characteristics. Results indicate that both classifications have several similarities.

The third analysis focuses on demand fluctuations due to various events such as weather events, subway service disruptions and activities (such as concerts or sporting events). Variables that have a significant impact on demand are selected independently for each mode using GAM models, which is a model using functions to transform explanatory variables. Then, the impact of those variables on demand is estimated. Results suggest that demand for each mode is influenced by the presence of rain, by temperature, by subway service disruptions and by activities near stations. Bikesharing is the only mode also impacted by wind speed. The presence of rain reduces demand for bikesharing, subway and bus, while increasing demand for taxi. Furthermore, models are calibrated for each station type defined by the station typology. Two types of descriptive models are compared being GAM and ARIMA (an autoregressive model for time series). Moreover, models are calibrated for each subway station. Results of the descriptive models suggest that rain has less impact on stations near downtown. Also, stations with recreational activities nearby seem more affected by weather than others. In addition, demand intensity is predicted on different horizons using predictive GAM and ARIMA models. The ARIMA model seems to perform better for short-term predictions, while the GAM model performs better for long-term predictions. The ARIMA model seems too flexible for long-term predictions, causing what appears to be overfitting.

Finally, this study analyses transportation demand of four modes simultaneously using continuous data flows, which has been little explored in the literature. In addition, a single model is used to evaluate the impact of several types of events on short-term demand, while other studies consider only one type of event at a time. This study provides a better understanding of travel behaviors in Montreal. It allows to consider a more dynamic adaptation of transportation supply to match travel demand. After explaining the contributions, this thesis presents some limitations and perspectives.

TABLE DES MATIÈRES

REMERCIEMENTS III

RÉSUMÉ IV

ABSTRACT VII

TABLE DES MATIÈRES IX

LISTE DES TABLEAUX XIII

LISTE DES FIGURES XV

LISTE DES ANNEXES XIX

CHAPITRE 1 INTRODUCTION 1

1.1 Mise en contexte 1

1.2 Problématique 2

1.3 Objectifs 2

1.4 Structure du mémoire 3

CHAPITRE 2 REVUE DE LITTÉRATURE 5

2.1 Facteurs influençant le choix modal individuel 5

2.2 Facteurs influençant la demande 7

2.2.1 Service de transport 8

2.2.2 Facteurs temporels 9

2.2.3 Météorologie 10

2.2.4 Activités 12

2.2.5 Économie 12

2.3 Modélisation de la demande 13

2.3.1 Modèles unimodaux 14

2.3.2 Modèles multimodaux 15

2.4 Zone d'attraction des stations de métro 17

CHAPITRE 3 MÉTHODOLOGIE 19

3.1 Méthodologie générale 19

3.2 Territoire à l'étude 21

3.3 Données 23

3.3.1 Données transactionnelles 23

3.3.2 Données explicatives 26

3.3.3 Caractéristiques du voisinage des stations 34

3.4 Prétraitement des données 38

3.4.1 Filtrage 38

3.4.2 Agrégation spatiale 38

3.4.3 Agrégation temporelle 40

3.4.4 Corrections 40

3.4.5 Imputation 41

3.4.6 Bonification 46

3.4.7 Fusion 46

CHAPITRE 4 ANALYSE DES PATRONS TYPIQUES D'USAGE 48

4.1 Patrons annuels 48

4.1.1 Méthodologie 48

4.1.2 Résultats 53

4.2 Patrons hebdomadaires 57

4.2.1 Méthodologie 57

4.2.2 Résultats 60

4.3 Patrons journaliers 69

4.3.1 Méthodologie 69

4.3.2 Résultats 73

CHAPITRE 5 TYPOLOGIE DE STATIONS 82

5.1 Typologie de stations selon les caractéristiques 82

5.2 Classification selon les patrons journaliers 88

5.2.1 Classification selon le volume et la répartition journalière 88

5.2.2 Classification selon la répartition journalière 94

5.3 Association des patrons journaliers à la typologie 98

CHAPITRE 6 IMPACT D'ÉVÈNEMENTS ATYPIQUES SUR L'USAGE 101

6.1 Modèles sur l'ensemble des stations 101

6.1.1 Filtrage des données 102

6.1.2 Normalisation de la demande 102

6.1.3 Fusion de données 103

6.1.4 Transformation de la variable d'intérêt 104

6.1.5 Sélection des variables 105

6.1.6 Formation des périodes horaires 105

6.1.7 Résultats des modèles 114

6.2 Modèles par type de station 126

6.2.1 Modèles GAM 126

6.2.2 Modèles ARIMA 127

6.2.3 Comparaison des modèles 128

6.3 Modèles par station 131

6.3.1 Modèles descriptifs 131

6.3.2 Modèles prédictifs 133

CHAPITRE 7 CONCLUSION 135

7.1 Synthèse de la recherche 135

7.2 Contributions 137

7.3 Limites 139

7.3.1 Limites dues aux données 139

7.3.2 Limites méthodologiques 140

7.4 Perspectives 141

BIBLIOGRAPHIE 142

ANNEXES 145

LISTE DES TABLEAUX

Tableau 2-1 : Facteurs influençant la demande dans le temps 7

Tableau 3-1 : Données transactionnelles collectées selon le mode 24

Tableau 3-2 : Jours avec données transactionnelles disponibles 26

Tableau 3-3 : Statistiques descriptives sur les variables météo horaires 27

Tableau 3-4 : Statistiques descriptives sur les variables météo journalières 28

Tableau 3-5 : Nombre d'arrêts de service du métro par ligne 30

Tableau 3-6 : Services provisoires de métro 32

Tableau 3-7 : Caractéristiques du voisinage des stations 35

Tableau 3-8 : Statistiques descriptives sur les caractéristiques des stations 36

Tableau 3-9 : Heures de début d'activité imputées par catégorie 43

Tableau 3-10 : Durées d'activité imputées par catégorie 45

Tableau 3-11 : Extrait de la base de données transactionnelle 47

Tableau 4-1 : Extrait de la table de données contenant les vecteurs 51

Tableau 4-2 : Répartition des vecteurs parmi les groupes 53

Tableau 4-3 : Extrait de la table de données contenant les vecteurs 59

Tableau 4-4 : Répartition des vecteurs parmi les groupes 60

Tableau 4-5 : Extrait de la table de données contenant les vecteurs 72

Tableau 4-6 : Répartition des vecteurs parmi les groupes 73

Tableau 5-1 : Centres des groupes 87

Tableau 5-2 : Extrait des vecteurs utilisés pour la classification 89

Tableau 5-3 : Matrice de confusion 98

Tableau 6-1 : Périodes horaires formées selon le mode 112

Tableau 6-2 : Périodes horaires formées pour le taxi 113

Tableau 6-3 : Définition des variables 114

Tableau 6-4 : Variation relative de l'intensité selon chaque variable 117

Tableau 6-5 : Critère d'information d'Akaike (AIC) des différents modèles 125

Tableau 6-6 : Variation relative de l'intensité selon la pluie et le vent, pour le vélopartage.....126

Tableau 6-7 : Variation relative de l'intensité selon la variable $pluie3h + A$, pour le vélopartage 129

Tableau 6-8 : AIC des modèles selon le type de station 130

Tableau 6-9 : Erreur RMSE par type de modèle, pour le vélopartage 130

Tableau 6-10 : Erreur (RMSE) par validation croisée pour le vélopartage à la station Mont-Royal 134

Tableau 7-1 : Statistique du test de Kolmogorov–Smirnov pour chaque paire de jours.....145

Tableau 7-2 : Coefficients du modèle du vélopartage.....149

Tableau 7-3 : Coefficients du modèle du taxi (suite) 150

Tableau 7-4 : Coefficients du modèle du métro 151

Tableau 7-5 : Coefficients du modèle du bus.....152

Tableau 7-6 : Coefficients du modèle GAM, pour le vélopartage 153

Tableau 7-7 : Coefficients du modèle ARIMA, pour le vélopartage 154

LISTE DES FIGURES

Figure 3-1 : Diagramme méthodologique général 20

Figure 3-2 : Carte de la région métropolitaine de Montréal (Google, s. d.).....21

Figure 3-3 : Réseaux de transport à l'étude.....22

Figure 3-4 : Distribution de la température observée.....27

Figure 3-5 : Distribution de la vitesse du vent 28

Figure 3-6 : Distribution de la quantité de pluie journalière 29

Figure 3-7 : Distribution de la quantité de neige journalière 29

Figure 3-8 : Nombre d'interruptions par station de métro 31

Figure 3-9 : Distribution de la durée des interruptions de service du métro 31

Figure 3-10 : Nombre d'activités par catégorie 33

Figure 3-11 : Nombre d'activités par station 33

Figure 3-12 : Zones d'influence des stations de métro 39

Figure 3-13 : Voisinage de la station de métro Édouard-Montpetit.....40

Figure 3-14 : Distribution des heures de début d'activité par catégorie 42

Figure 3-15 : Distribution des durées d'activité par catégorie 44

Figure 4-1 : Intensité d'usage moyenne par mode 50

Figure 4-2 : Exemple de vecteurs pour le taxi, avant et après le lissage.....51

Figure 4-3 : Vecteurs exclus aberrants.....52

Figure 4-4 : Nombre de groupes optimal par la méthode du coude 53

Figure 4-5 : Répartition des groupes par mode 54

Figure 4-6 : Centres des groupes des patrons annuels 54

Figure 4-7 : Attribution des groupes aux stations, pour le bus 55

Figure 4-8 : Attribution des groupes aux stations, pour le vélopartage 56

Figure 4-9 : Intensité moyenne d'usage par mode 58

Figure 4-10 : Nombre de groupes optimal par la méthode du coude 60

Figure 4-11 : Centres des groupes des patrons hebdomadaires 61

Figure 4-12 : Répartition des modes par groupe 61

Figure 4-13 : Distribution temporelle des groupes les plus fréquents par mode.....63

Figure 4-14 : Distribution temporelle des groupes par mode pour la station Université-de-Montréal 64

Figure 4-15 : Groupes le plus fréquent à chaque station, pour le métro 65

Figure 4-16 : Répartition des groupes à chaque station, pour le métro.....66

Figure 4-17 : Répartition des groupes à chaque station, pour le vélopartage 67

Figure 4-18 : Répartition des groupes à chaque station, pour le taxi 68

Figure 4-19 : Intensité moyenne d'usage par mode, pour les jours de semaine.....71

Figure 4-20 : Intensité moyenne d'usage par mode, pour les jours de fin de semaine 71

Figure 4-21 : Nombre de groupes optimal par la méthode du coude 73

Figure 4-22 : Centres des groupes des patrons hebdomadaires 74

Figure 4-23 : Répartition des groupes par mode 74

Figure 4-24 : Distribution temporelle des groupes les plus fréquents par mode.....76

Figure 4-25 : Distribution des vecteurs par jour de la semaine et par mode 77

Figure 4-26 : Répartition des groupes à chaque station pour les jours de semaine, pour le métro 78

Figure 4-27 : Répartition des groupes à chaque station, pour le vélopartage 79

Figure 4-28 : Répartition des groupes à chaque station, pour le taxi 80

Figure 4-29 : Répartition des groupes à chaque station, pour le bus 81

Figure 5-1 : Dendrogramme résultant de la classification selon les caractéristiques.....84

Figure 5-2: Répartition des stations par groupe 85

Figure 5-3 : Dendrogramme résultant de la classification selon les patrons journaliers.....91

Figure 5-4 : Répartition spatiale des stations par groupe 92

Figure 5-5 : Centres des groupes.....93

Figure 5-6 : Dendrogramme résultant de la classification selon les patrons journaliers.....95

Figure 5-7 : Répartition spatiale des stations par groupe 96

Figure 5-8 : Centres des groupes.....97

Figure 5-9 : Spatialisation des différences de classification 99

Figure 6-1 : Distribution des intensités de déplacement selon la transformation 104

Figure 6-2 : Intensité horaire moyenne selon le jour, pour le vélopartage.....106

Figure 6-3 : Distribution des intensités de déplacements par heure, pour le vélopartage.....106

Figure 6-4 : Distance maximale obtenue par le test KS pour une paire d'heures 107

Figure 6-5 : Matrice de distances maximales pour chaque paire d'heures, pour le vélopartage .108

Figure 6-6 : Nombre de périodes formées selon la distance seuil choisie, pour le vélopartage ..109

Figure 6-7 : Erreur du modèle selon la distance seuil choisie, pour le vélopartage.....109

Figure 6-8 : Erreur du modèle selon le nombre de périodes choisi, pour le vélopartage.....110

Figure 6-9 : Périodes horaires formées selon le seuil maximal de distance, pour le vélopartage 110

Figure 6-10 : Distribution des déplacements par période, pour le vélopartage.....111

Figure 6-11 : Intensité moyenne horaire selon le jour, pour le taxi 112

Figure 6-12 : Matrice de corrélation entre les variables sélectionnées pour le vélopartage 116

Figure 6-13 : Variation relative de l'intensité selon la température.....119

Figure 6-14 : Variation relative de l'intensité selon la durée d'interruption de service.....120

Figure 6-15 : Variation de l'intensité des départs selon l'activité et le mode.....122

Figure 6-16 : Résidus selon les valeurs prédites, pour le modèle du vélopartage.....123

Figure 6-17 : Graphe quantile-quantile des résidus 124

Figure 6-18 : Valeurs prédites par rapport aux valeurs réelles, pour le modèle du vélopartage..124

Figure 6-19 : Intensité des déplacements selon la température par type de station, pour le vélopartage 127

Figure 6-20 : Exemple de prévision de l'intensité pour les stations de type 3, pour le vélopartage 129

Figure 6-21 : Variation relative de l'intensité pour le vélopartage selon la variable $pluie3h + A$ 132

Figure 6-22 : Exemple de prévision de l'intensité pour le vélopartage à la stations Laurier 133

Figure 7-1 : Distribution horaire des déplacements par jour, pour le bus 145

Figure 7-2 : Matrice des distances et périodes horaires formées, pour le bus 146

Figure 7-3 : Erreur du modèle selon la distance seuil, pour le bus 146

Figure 7-4 : Distribution horaire des déplacements par jour, pour le métro 147

Figure 7-5 : Matrice des distances et périodes horaires formées, pour le métro 147

Figure 7-6 : Erreur du modèle selon la distance seuil, pour le métro 148

LISTE DES ANNEXES

Annexe A – Formation des périodes horaires 145

Annexe B – Coefficients des modèles 149

Annexe C – Coefficients des modèles par type de station 153

CHAPITRE 1 INTRODUCTION

Tout d'abord, cette section présente une brève mise en contexte des différents systèmes de transport de la région de Montréal. Ensuite, la problématique et les objectifs de l'étude seront discutés, avant la présentation de la structure du mémoire.

1.1 Mise en contexte

Dans un contexte où la technologie évolue rapidement, les modes de transport se diversifient et leurs interactions deviennent de plus en plus complexes. Les modes partagés comme le vélopartage et l'autopartage prennent de l'ampleur dans les grandes villes et offrent de nouvelles opportunités de déplacements aux usagers. En outre, les usagers combinent régulièrement plusieurs modes de transport pour compléter un déplacement et varient les modes empruntés. Ces comportements entrainent des variations importantes de la demande à court terme, qui sont difficilement pris en compte dans les modèles actuels de prévision de la demande. Il devient pertinent d'étudier le comportement des usagers de plusieurs modes de transport simultanément, afin de bien comprendre les comportements de mobilité de la population.

Les modèles de prévision de la demande s'appuient typiquement sur des données d'enquêtes. Les déplacements d'un échantillon de la population sont recensés pour une journée donnée. Les motifs de déplacement et les modes empruntés sont recensés. Ces données permettent de brosser un portrait général des habitudes de déplacement de la population pour un jour typique de semaine. Elles permettent également d'expliquer comment les caractéristiques sociodémographiques influencent les déplacements effectués.

Une alternative aux données d'enquête est l'utilisation de flux continus de données. La disponibilité de données en continu permet maintenant d'observer et de modéliser la variabilité d'usage de différents modes, mais surtout de croiser ces variations avec des facteurs qui modulent en continu les comportements. Ainsi, les données des systèmes opérationnels de transport peuvent être utilisées afin de quantifier les déplacements des usagers dans différents systèmes de transport.

1.2 Problématique

Les modèles typiques de prévision de la demande s'appuyant sur des données d'enquêtes comportent certaines limitations au niveau de leur couverture temporelle, des modes considérés, des types de jours observés et du taux d'échantillonnage. Tout d'abord, ces enquêtes sont souvent menées sur des périodes temporelles restreintes plutôt qu'en continue, ce qui ne permet pas de capturer le cycle annuel de fluctuation de la demande. Ensuite, ces enquêtes négligent généralement les modes de transport de faible usage, tel que le taxi, le vélopartage et l'autopartage. Pour continuer, plusieurs enquêtes se limitent aux jours ouvrables, délaissant les jours de fin de semaine et les jours fériés. Aussi, les taux journaliers d'échantillonnage sont typiquement faibles, ce qui oblige à regrouper plusieurs jours ouvrables en un jour moyen de semaine. Or, cette technique élimine la variabilité entre les jours observés et ne permet donc pas d'observer des variations d'usage à court terme.

Par ailleurs, plusieurs modèles de prévision de la demande ignorent les facteurs externes pouvant influencer les décisions des usagers à court terme. Différents évènements ponctuels tels que des évènements météorologiques ont certainement une incidence sur le niveau d'utilisation des divers modes. Or, cet impact ne peut être quantifié à partir des données d'enquête qui se basent sur un jour moyen de semaine. L'usage de flux de données passifs permet de quantifier la variabilité d'usage en continu, pour ainsi évaluer l'impact de différents facteurs sur la demande.

La littérature relève plusieurs modèles d'analyse de la demande se basant sur des flux continus de données. Or, ces études portent généralement sur un seul mode de transport, ne permettant pas d'observer les corrélations d'usage entre les modes. De plus, peu d'études traitent des modes émergents, bien que leur niveau d'usage soit en augmentation dans plusieurs villes.

1.3 Objectifs

L'objectif général de la recherche est de mieux comprendre les corrélations d'usage entre différents modes de transport. Plus spécifiquement, le projet vise à modéliser l'usage du métro, du vélopartage, du taxi et du bus aux abords des stations de métro, en utilisant des flux passifs de données transactionnelles. Il vise à tenir compte de la variabilité d'usage des modes dans le temps et dans l'espace. Le cas à l'étude porte sur la région métropolitaine de Montréal et utilise trois

années de données transactionnelles. Or, les méthodes proposées pourraient être appliquées pour tous les modes de transport urbains et pourraient être adaptées à d'autres villes.

La recherche comporte quatre objectifs spécifiques. Le premier consiste à expérimenter différentes méthodes de traitement, d'analyse longitudinale et de modélisation de flux passifs de données.

Le second objectif spécifique vise à mieux comprendre les cycles d'usage de différents modes de transport. L'objectif est d'utiliser des techniques de partitionnement de données afin d'analyser les cycles d'usage annuels, hebdomadaires et journaliers.

Le troisième objectif spécifique est de mieux comprendre l'usage de chaque station de métro, en créant une typologie de station. Une telle information permet de mieux comprendre les corrélations spatiales. Cela permet aussi d'évaluer des scénarios, tel que l'ajout d'une station de métro.

Le quatrième objectif spécifique est d'évaluer l'impact de différents évènements sur le niveau d'usage à court terme. Les évènements étudiés sont les évènements météorologiques, les activités majeures (par exemple des concerts ou des évènements sportifs) et les interruptions de service du métro. Le but est d'identifier les variables ayant un impact significatif sur l'usage de chaque mode, puis de quantifier cet impact.

Cette recherche contribue à l'amélioration des modèles de prévision et de planification stratégique des réseaux de transport en permettant une meilleure compréhension des fluctuations d'usage des divers modes de transport. Une meilleure compréhension des variations d'usage permet d'évaluer le niveau d'adéquation entre l'offre et la demande et permet d'envisager une adaptation plus dynamique de l'offre. Les résultats permettront de soutenir une prise de décision plus éclairée tant sur la planification stratégique d'un réseau intégré et multimodal de transport, ainsi que sur la priorisation de certains modes lors de l'allocation d'espace urbain et de ressources.

1.4 Structure du mémoire

Ce mémoire est composé de sept chapitres. Tout d'abord, le premier chapitre introduit la recherche par une mise en contexte du territoire à l'étude, la description de la problématique et la présentation des objectifs.

Le deuxième chapitre présente une revue de littérature des thèmes abordés dans ce mémoire. Les thèmes discutés sont les facteurs influençant le choix modal, les facteurs influençant la demande, la modélisation de la demande et la zone d'attraction d'une station de métro.

Le troisième chapitre présente la méthodologie générale du projet. Les données utilisées sont présentées, puis les différentes étapes du pré-traitement des données sont détaillées.

Le quatrième chapitre consiste en une analyse des patrons types d'usage. L'objectif est de mieux comprendre les cycles d'usage annuels, hebdomadaires et journaliers pour les différents modes.

Le cinquième chapitre présente une typologie des stations de métro, soit une classification de celles-ci basée sur leurs caractéristiques. Cette typologie est ensuite mise en relation avec les patrons journaliers d'usage.

Le sixième chapitre porte sur l'impact d'évènements atypiques sur le niveau d'usage. Tout d'abord, les variables ayant un impact significatif sur le niveau d'usage sont sélectionnées à l'aide de modèles regroupant l'ensemble des stations. Des modèles sont calibrés pour chaque type de station défini par la typologie proposée au chapitre cinq. Ensuite, différents modèles sont calibrés pour chaque station afin d'en faire la comparaison.

Le septième chapitre résume les différents résultats de la recherche, présente les contributions apportées, énumère les limites et formule certaines perspectives pour de futures recherches.

CHAPITRE 2 REVUE DE LITTÉRATURE

Cette section présente tout d'abord une revue non exhaustive des facteurs individuels influençant le choix modal. Ensuite, les facteurs qui influencent la demande sont présentés et différentes méthodes d'analyse de la demande sont expliquées. Finalement, différentes définitions de la zone d'attraction d'une station de métro sont relevées.

2.1 Facteurs influençant le choix modal individuel

Les déplacements en milieu urbain peuvent généralement être effectués en utilisant différents modes de transport, ou même différentes combinaisons de modes. Avant d'effectuer un déplacement, chaque usager fait un choix plus ou moins conscient du mode à emprunter.

Plusieurs recherches s'intéressent à mieux comprendre le choix des usagers. Martel Poliquin (2012) identifie plusieurs facteurs déterminants du choix modal et les met en corrélation avec les modes déclarés dans l'enquête Origine-Destination de Montréal de 2008. Il classifie ces facteurs en cinq catégories, soit les facteurs individuels, ceux liés au ménage, à l'environnement, au déplacement et les autres facteurs. Les autres facteurs incluent entre autres des facteurs psychologiques ainsi que les conditions dans lesquelles s'effectuent le déplacement (telles que les conditions météorologiques). Le fort achalandage d'un service de transport et la disponibilité de service peuvent aussi influencer les usagers. Certains usagers pourraient même décider de ne pas se déplacer à cause de certains facteurs tels que la météorologie (Martel Poliquin, 2012). Une autre étude de Pnevmatikou (2015) a modélisé le choix modal des usagers à la suite d'une panne de métro prolongée, en se basant sur les caractéristiques de la personne et du déplacement effectué. Certaines variables influençant le choix modal sont plus facilement mesurables, alors que plusieurs sont qualitatives.

Pour commencer, plusieurs facteurs sont propres au déplacement effectué. Le choix modal pour un déplacement peut être influencé entre autres par le motif, la distance, la durée, le dénivelé et les propriétés urbanistiques. En effet, la topographie influence davantage l'utilisation des modes actifs. Les trajets présentant de fortes déclivités présentent généralement des parts modales plus faibles du vélo et de la marche (Martel Poliquin, 2012).

Ensuite, de nombreuses propriétés urbanistiques influencent l'attractivité des différents modes. Par exemple, des infrastructures de transport actif telles que des pistes cyclables et des trottoirs

encouragent leur utilisation. Par ailleurs, un quartier dense diminue les distances de déplacement vers les différents points d'intérêt. Des densités de population et d'emplois élevées sont généralement corrélées avec une forte utilisation des modes actifs et collectifs (Martel Poliquin, 2012). Pour continuer, la connectivité est une mesure de la quantité de connexions disponibles sur le réseau. Elle peut être représentée par la densité d'intersections par exemple. Les réseaux routiers ayant une forte connectivité ont généralement une forte utilisation de la marche (Dill, 2013). De son côté, la tortuosité représente le caractère sinueux d'un trajet, par un ratio de la distance réseau sur la distance euclidienne. En milieu urbain, une tortuosité importante décourage généralement les automobilistes (Martel Poliquin, 2012). Ensuite, la disponibilité du stationnement et sa tarification sont également des facteurs qui remettent en question l'utilisation de la voiture. Une faible disponibilité du stationnement et des coûts élevés favorisent plutôt l'utilisation du transport collectif (Mattson, 2008). Finalement, la distance d'accès à un arrêt de transport en commun influence son utilisation. Une faible distance d'accès à un arrêt à partir du domicile et du travail augmente la part modale du transport collectif, particulièrement pour les modes sur rail (Martel Poliquin, 2012).

Pour continuer, certains facteurs influençant le choix modal sont propres à la personne et au ménage. Certains facteurs sont intrinsèques tels que le genre, l'âge et l'ethnicité de la personne, alors que d'autres sont extrinsèques tels que l'occupation principale, le revenu, la possession automobile, la possession d'un permis de conduire ou d'un titre de transport en commun (Martel Poliquin, 2012). Le genre influence les habitudes de déplacement selon plusieurs études, dû aux rôles des femmes et des hommes qui diffèrent dans notre société, même si ces différences s'atténuent progressivement (Resenbloom, 2006). L'âge est généralement un indicateur de l'occupation principale de la personne, influençant le type de déplacement effectué (Martel Poliquin, 2012). D'autres facteurs facilitent ou restreignent l'utilisation de certains modes, comme par exemple la possession automobile et du permis de conduire. Le revenu peut restreindre l'utilisation de certains modes qui sont plus dispendieux. Finalement, la connaissance des différentes alternatives de transport est un facteur difficile à mesurer, mais qui a certainement un impact sur l'utilisation de certains modes.

Bien que de nombreux facteurs individuels influencent le choix modal, la demande totale d'un mode de transport est relativement stable étant donné le grand nombre d'usagers. La section

suivante présente des facteurs qui touchent l'ensemble des usagers et qui influencent donc la demande totale d'un mode de transport.

2.2 Facteurs influençant la demande

Cette section s'intéresse aux facteurs qui influencent l'achalandage d'un mode de transport dans le temps, soit des éléments qui impactent l'ensemble des usagers d'un système. Le Tableau 2-1 présente une liste non exhaustive de différents facteurs relevés dans la littérature. Ceux-ci sont ensuite discutés dans les sous-sections qui suivent.

Tableau 2-1 : Facteurs influençant la demande dans le temps

Catégorie	Facteur
Service de transport	Heures d'ouverture Quantité de service Qualité du service Interruptions de service Frais d'utilisation
Facteurs temporels	Heure Jour de la semaine Période de l'année Congés fériés Périodes de vacances
Météorologie	Précipitations Température Vent Humidité
Activités	Activités majeures
Socioéconomique	Prix de l'essence Tarification Population
Utilisation du sol	Utilisation du sol

2.2.1 Service de transport

Le choix modal des usagers est certainement influencé par l'offre de transport et par les frais imposés à ses utilisateurs. Une offre variée, un service fiable et de faibles frais d'usage sont susceptibles d'augmenter l'attractivité d'un mode.

Tout d'abord, l'offre des systèmes de transport présente certaines variations. En effet, certains services ne sont pas accessibles toute l'année, comme c'est le cas pour la majorité du réseau cyclable montréalais qui demeure fermé durant l'hiver. Le service de vélopartage Bixi est aussi fermé durant la période froide, soit de la mi-novembre à la mi-avril (BIXI, 2018). Par ailleurs, la quantité de service du transport en commun peut varier selon le jour de la semaine et selon la saison, comme c'est le cas pour plusieurs autobus et trains à Montréal dont les fréquences sont réduites la fin de semaine et les jours fériés (EXO, 2018). Du côté du vélopartage, le service Bixi modifie la position de certaines stations en fonction de la demande, par exemple lors d'évènements spéciaux ou de travaux municipaux (BIXI, 2018). Pour continuer, la qualité du service impacte aussi l'usage des modes. Des interruptions de service du transport en commun diminuent son attractivité. Du côté du vélopartage, l'indisponibilité de vélos ou d'ancrages rend son utilisation ardue. Similairement, l'indisponibilité de taxi ou de véhicules d'autopartage diminue aussi leur attractivité. La qualité du service est également définie par le confort à bord des véhicules et aux stations de correspondance.

Les interruptions de service encouragent les usagers à utiliser des modes de transport alternatifs pour compléter leur déplacement. La littérature sur le sujet est plutôt limitée. Lin (2017) a mené une enquête à préférence déclarée sur les usagers du métro de Toronto et a comparé les résultats à une enquête à préférences révélées. Pour cette étude, une interruption de service a été définie comme un retard au service planifié de 10 minutes ou plus. Les répondants déclarent dans 89% des cas qu'ils cherchent un mode alternatif lors d'une interruption de service du métro, alors que les autres préfèrent attendre le rétablissement du service. Cependant, l'enquête à préférences révélées indique que le nombre d'usagers qui attendent le rétablissement du service serait aussi élevé que 66%. Cette enquête indique que 17% des usagers choisissent de prendre la navette ou un autre mode de transport en commun lors d'une interruption de service, alors que 11% des usagers choisissent de marcher et 6% prennent leur voiture privée. L'enquête à préférence déclarée suggère que des conditions météorologiques défavorables entrainent une diminution de l'usage de la

marche et du vélo et une augmentation de l'usage du bus. Sans trop de surprises, plus la durée de l'interruption de service est élevée, plus les usagers choisissent d'emprunter des modes alternatifs au métro.

Finalement, dans le cadre de cette revue de littérature, aucune étude consultée n'estime l'impact d'une interruption de service sur la demande d'un autre mode de transport.

Ensuite, les frais liés à l'utilisation d'un système de transport ont certainement un impact sur son utilisation. Pour les modes collectifs et partagés, la tarification peut être sous forme d'abonnement ou payable à l'utilisation. Afin de quantifier l'impact d'un changement de tarification sur l'achalandage, un indicateur nommé « élasticité » peut être calculé. L'élasticité est le ratio de la variation proportionnelle d'achalandage sur la variation proportionnelle du tarif. Cette valeur est généralement négative, indiquant une relation inverse entre les deux variables (Paulley et al., 2006). En d'autres mots, une augmentation du tarif a généralement la conséquence de diminuer le nombre d'usages. En ce qui concerne la voiture personnelle, elle présente des frais liés à l'achat et à son utilisation. Selon le trajet emprunté, des frais de péages peuvent également s'ajouter.

2.2.2 Facteurs temporels

L'usage des différents modes est également influencé par certains facteurs temporels. En effet, certains cycles d'usage se répètent à intervalles réguliers, soit de façon journalière, hebdomadaire et annuelle. Tout d'abord, les cycles journaliers présentent généralement deux périodes de pointe pour les jours de semaine, soit une pointe en matinée et une pointe en fin d'après-midi. Durant la fin de semaine, l'achalandage est habituellement plus constant durant la journée (Nishiuchi, King et Todoroki, 2013; Nosal et Miranda-Moreno, 2014). Ensuite, des cycles d'usage d'une période d'une semaine sont aussi observés. Les jours de semaine comptent généralement plus de déplacements que les jours de fin de semaine. De plus, certaines études montrent que le lundi et le vendredi ont des usages différents des autres jours de semaine (Miranda-Moreno et Nosal, 2011; Sabir, 2011). Par ailleurs, un cycle annuel est également observé et celui-ci varie selon le mode. Par exemple à Montréal durant l'été, l'usage du vélo augmente de façon significative alors que l'usage du taxi diminue, ce qui laisse présager un transfert modal (Morency, Trépanier, Saunier, Verreault et Bourdeau, 2016). Finalement, les jours de congé, les fêtes et les périodes de vacances peuvent également introduire une variation de l'achalandage.

2.2.3 Météorologie

Les conditions météorologiques affectent chaque mode différemment. Malgré cela, Sabir (2011) observe que le nombre total de déplacements est très peu affecté par la météorologie. Cette étude se base sur l'enquête Néerlandaise sur les déplacements de la population pour conclure que les conditions météorologiques introduisent un transfert modal important, sans toutefois changer significativement le nombre total de déplacements effectués. Par exemple, des précipitations introduisent un transfert modal du vélo vers le transport en commun et la voiture, sans influencer significativement la demande totale. Par ailleurs, des températures très élevées (plus de 25ºC) impliquent une augmentation de 22% de l'usage du vélo, une diminution de 15% pour la voiture et une diminution de 20% de l'usage du transport en commun. Sabir conclut que la seule condition météorologique ayant un impact significatif sur la demande en déplacement globale est les températures très élevées (plus de 25ºC), qui impliquent une diminution de la demande globale d'environ 5%.

Les cyclistes sont particulièrement sensibles aux conditions météorologiques, c'est pourquoi le vélo est le mode le plus affecté par la météorologie (Brandenburg, Matzarakis et Arnberger, 2007; Sabir, 2011). L'impact de la météorologie sur l'usage du vélo varie selon la région étudiée. Les régions comptant davantage de jours de gel rapportent une part modale plus faible que les autres (Winters, Friesen, Koehoorn et Teschke, 2007). L'usage du vélo diminue généralement en présence de précipitations et augmente avec la température (Brandenburg et al., 2007; Miranda-Moreno et Nosal, 2011; Niemeier, 1996; Sabir, 2011; Thomas, Jaarsma et Tutert, 2008). Sabir explique ce changement par un transfert modal de la voiture et du transport en commun. Cependant, plusieurs études montrent que l'impact de la température est non linéaire (Miranda-Moreno et Nosal, 2011; Niemeier, 1996; Thomas et al., 2008). En effet, les cyclistes sont moins susceptibles d'utiliser le vélo lors de températures extrêmement élevées ou faibles, étant donné que le confort est réduit (Miranda-Moreno et Nosal, 2011; Richardson, 2000). Lors de températures de 28ºC et plus, l'usage du vélo diminue significativement à Montréal (Miranda-Moreno et Nosal, 2011). Pour continuer, les cyclistes semblent baser leur choix modal sur la météo actuelle plutôt que sur les prévisions météorologiques (Gallop, Zhao et Tse, 2012). Miranda-Moreno précise que des précipitations durant la matinée ou durant les 3 heures précédant un déplacement impliquent une baisse significative de l'usage du vélo. Par ailleurs, l'humidité s'est révélée significative pour quelques études, impliquant une diminution de l'usage du vélo en présence d'humidité très élevée (Miranda-

Moreno et Nosal, 2011). Le vent s'est avéré significatif pour Thomas (2008), Sabir (2011) et Guo (2007), mais n'était pas significatif pour Miranda-Moreno (2011). Finalement, les différentes études s'accordent pour énoncer que la demande récréative est plus sensible aux conditions météorologiques que la demande utilitaire (Brandenburg et al., 2007; Richardson, 2000; Thomas et al., 2008).

Le transport en commun est peu affecté par les conditions météorologiques, en particulier les modes sur rail (Guo, Wilson et Rahbee, 2007). Selon Sabir (2011), la demande utilitaire n'est pas affectée significativement par la météo, contrairement aux déplacements récréatifs. Lors de conditions météorologiques défavorables (pluie, vent et températures extrêmes), la demande pour le métro de New York diminue, mais de façon différente selon l'emplacement et le temps (Singhal, Kamga et Yazici, 2014). Aux Pays-Bas, la demande de transport en commun tend à diminuer légèrement lorsque la température augmente, sauf pour des températures très faibles lors desquelles la demande diminue (Sabir, 2011). À l'inverse, la demande semble augmenter avec la température à Chicago (Guo et al., 2007). Stover (2012) observe une diminution de l'usage du bus dans le comté de Pierce (États-Unis) durant l'hiver seulement, alors que l'attente à l'extérieur peut être moins confortable (Stover et McCormack, 2012). Par ailleurs, des précipitations entrainent une diminution d'usage d'environ 3% selon Kashfi (2013), alors que Sabir (2011) identifie plutôt une légère augmentation de 0,2%. De son côté, Guo (2007) observe une réduction de l'usage des modes sur rail plus importante que celle du bus. Cette réduction est plus importante la fin de semaine alors que la demande est majoritairement récréative (Guo et al., 2007). La neige semble également diminuer la demande du bus et du métro (Guo et al., 2007; Stover et McCormack, 2012). Stover (2012) estime cette diminution à 11% pour le bus. Ensuite, la vitesse du vent diminue l'usage du bus selon Guo (2007) et Stover (2012), pouvant être attribué à une réduction du confort lors du temps d'accès et d'attente, ainsi qu'une réduction de la vitesse d'opération (Sabir, 2011). Cependant, Guo (2007) relève que le vent a un impact négligeable sur les modes sur rail. Finalement, Guo (2007) suggère que le brouillard augmenterait la demande des modes sur rail jusqu'à 10% à Chicago, mais cette variable n'a pas été considérée dans les autres études consultées.

La demande du taxi est moins discutée dans la littérature que le vélo ou le transport en commun. Une étude de Kamga (2013) révèle que les conditions météorologiques ont un impact sur les revenus du taxi à New York (États-Unis). En présence de pluie, le nombre de courses augmente et la distance moyenne des trajets diminue. Cependant, la neige ne semble pas avoir d'impact

significative sur la demande de taxi. Il est important de mentionner que les chauffeurs de taxi sont généralement des travailleurs autonomes qui n'ont aucune obligation d'offrir un service minimum à un temps donné. Par conséquent, l'offre de transport subit des variations causées par les décisions individuelles de chauffeurs de taxi (Kamga, Yazici et Singhal, 2013).

L'usage de la voiture varie également selon les conditions météorologiques, ce qui peut être expliqué par un transfert modal (Sabir, 2011). En Belgique, l'achalandage sur les routes diminue lors de précipitations de pluie et de neige, ainsi que lors de forts vents (Cools, Moons et Wets, 2010). L'étude de Sabir (2011) observe plutôt une augmentation de son usage en présence de précipitations.

2.2.4 Activités

L'impact de différentes activités sur la demande en transport semble très peu avoir été étudié dans la littérature. La disponibilité de données historiques d'activités est probablement un frein à ce type d'analyse. Tao (2014) analyse l'achalandage journalier du métro de Xi'an (Chine) pour mieux comprendre l'impact des activités majeures et des congés fériés sur la demande. Lors des jours de festivals traditionnels, l'achalandage du métro augmente d'environ 19%. La demande maximale a lieu lors des congés fériés principaux, lors desquels la demande augmente en moyenne de 24%.

Kuppam (2013) a analysé l'impact d'activités spéciales sur le système léger sur rail de la région métropolitaine de Phoenix, aux États-Unis. Une enquête a été menée en collectant des données sur le motif du déplacement et sur les modes empruntés. Un modèle a été développé pour prévoir la demande en transport lors de différents types d'évènements.

2.2.5 Économie

L'augmentation du prix de l'essence peut encourager des utilisateurs de la voiture à se tourner vers d'autres alternatives modales. En effet, l'étude de Stover et McCormack (2012) conclut que le prix de l'essence fait augmenter l'achalandage de l'autobus de façon significative.

Mattson (2008) a mené une étude portant sur 11 systèmes de transport en commun dans l'état du Dakota du Nord sur une période de 10 ans. Il conclut que la variation du prix de l'essence ne produit pas une réaction immédiate sur la demande de transport en commun, ce qui pourrait être expliqué par les hypothèses suivantes. Tout d'abord, suite à une hausse de prix, certains automobilistes

s'attendent à ce que le prix redescende au niveau original, c'est pourquoi ils peuvent prendre du temps à décider de changer leurs habitudes de déplacements (Mattson, 2008). Ensuite, les automobilistes développent un certain attachement à leur voiture lié à son confort et à sa liberté, ce qui rend le changement d'habitudes plus lent (Mattson, 2008). Ainsi, Mattson observe que les répercussions sur l'utilisation du transport en commun sont perceptibles jusqu'à 7 mois suivant l'augmentation du prix de l'essence. À long terme, l'élasticité du transport en commun est généralement située entre 0,08 et 0,16, ce qui signifie qu'une augmentation de 1% du prix de l'essence mène à une augmentation de l'utilisation du transport en commun d'environ 0,08% à 0,16%. Cependant, les revenus générés par l'augmentation d'achalandage ne sont pas suffisants pour couvrir les coûts d'opération supplémentaires dus à l'essence.

2.3 Modélisation de la demande

Avec la disponibilité croissante de données opérationnelles, les analyses portant sur la modélisation de l'usage des différents modes sont de plus en plus nombreuses. Différentes méthodes sont utilisées dans la littérature afin d'analyser les variations de la demande selon certaines variables explicatives.

Différents types de données sont utilisées, tel que des données transactionnelles (Kamga et al., 2013; Morency et al., 2016; Sabir, 2011; Stover et McCormack, 2012), des compteurs automatiques (Gallop, Zhao et Tse, 2012; Miranda-Moreno et Nosal, 2011; Thomas et al., 2008), des compteurs manuels (Niemeier, 1996), des enquêtes (Lin, 2017; Richardson, 2000; Sabir, 2011) et des enregistrements vidéos (Brandenburg et al., 2007). Par ailleurs, l'unité temporelle d'analyse diffère, allant d'une heure (Gallop et al., 2012; Miranda-Moreno et Nosal, 2011; Thomas et al., 2008) à un jour (Guo et al., 2007; Stover et McCormack, 2012).

Différents types de modèles sont utilisés, souvent avec la demande comme variable dépendante. Plusieurs études utilisent un modèle linéaire (Brandenburg et al., 2007; Guo et al., 2007; Stover et McCormack, 2012), alors que d'autres utilisent des régressions logistiques (Sabir, 2011), des modèle de Poisson (Niemeier, 1996), des modèles de séries temporelles (Gallop et al., 2012) et des modèles multiniveaux (Thomas et al., 2008; Winters et al., 2007) et même un modèle économétrique (Sabir, 2011).

Cette section présente plusieurs méthodes de modélisation unimodales, ainsi que quelques méthodes multimodales.

2.3.1 Modèles unimodaux

L'étude de Lorimier et El-Geneidy (2013) modélise l'utilisation et la disponibilité des véhicules du système d'autopartage Communauto à Montréal, en utilisant une année de données transactionnelles. Un modèle multi-niveaux estime l'usage mensuelle du service en fonction de diverses variables explicatives. Un second modèle estime la disponibilité des véhicules en fonction de caractéristiques liées au véhicule, à la station et au moment de l'emprunt.

L'étude de Miranda-Moreno et Nosal (2011) s'intéresse à l'impact de la météo sur l'utilisation du vélo en tenant compte des tendance temporelles. Elle utilise 3 ans de données de comptage de vélo à Montréal, agrégées par heure. L'analyse des données à l'échelle horaire permet d'observer la variation de comportement durant la journée. Deux approches sont utilisées, soit une approche absolue et une approche relative. L'approche absolue cherche à expliquer le volume de cyclistes horaire selon les variables météo absolues. Pour cette approche, un modèle binomial négatif et un modèle log-linéaire sont calibrés. Ensuite, l'approche relative explique le volume de cyclistes horaire en fonction de la déviation des conditions météo à la moyennes saisonnière.

De leur côté, Gallop et al. (2012) s'intéressent au trafic à vélo en fonction de la météo pour la ville de Vancouver, en utilisant près de 2 ans de données de comptage horaire. L'étude cherche à comprendre l'impact des variables météorologiques telles que la température, les précipitations, l'humidité, la vitesse du vent et le brouillard. Un modèle ARIMA (modèle autorégressif à moyenne mobile intégrée) est utilisé pour capturer la tendance, la saisonnalité et la corrélation des termes d'erreur. La cycles journaliers et hebdomadaires sont tous deux pris en charge par le modèle. La performance du modèle est évaluée par validation croisée, soit en utilisant 75% des données pour entrainer le modèle et les 25% restants pour le tester. Des variables sont ajoutées pour tenir compte des précipitations dans les 3 heures précédant chaque heure étudiée.

Gallop et al. (2012) développent également un second modèle sous la forme double-log. Celui-ci considère la quantité de service offerte, la tarification, ainsi que des facteurs socio-économiques. De plus, ce modèle prend en compte la tendance temporelle non expliquée par les autres variables.

Par ailleurs, l'étude de Mattson (2008) cherche à expliquer la variation d'achalandage de l'autobus en fonction du prix de l'essence, en utilisant les données transactionnelles de 12 agences de transport américaines sur une période de 10 ans. Afin de quantifier à la fois les effets du prix de l'essence actuel et des prix passés, un modèle à retards échelonnés Almon est utilisé. Les données historiques du prix de l'essence moyen et du nombre de déplacements effectués sont agrégées par mois. La forme fonctionnelle double-log est utilisée. Or, ce modèle néglige les autres variables telles que la quantité de service offert, la qualité du service, la tarification ainsi que d'autres variables externes.

2.3.2 Modèles multimodaux

Les recherches intégrant des flux de données continus provenant de plusieurs modes se font plus rares dans la littérature. Le défi de ces analyses est l'utilisation de techniques de fusion de données portant sur divers objets d'analyse. De plus, la demande peut également être d'ordres de grandeur différents selon le mode. La disponibilité des données est un frein à l'intégration de plusieurs modes dans ces modèles.

Une étude menée en Suède par Bergström et Magnusson (2003) s'intéresse au transfert modal entre le vélo et la voiture en fonction des saisons. Un sondage a été répondu par mille travailleurs afin de mieux comprendre leurs choix modaux selon les saisons, la température et les précipitations. L'étude conclut que le nombre de déplacements effectués à vélo diminue de 47% en hiver, alors que l'usage de la voiture augmente de 27%. D'autre part, les cyclistes qui continuent d'utiliser le vélo en hiver déclarent être moins affectés par les variations météorologiques.

De son côté, Sabir (2011) estime les conséquences de la météo sur le choix modal des utilisateurs, en utilisant 10 ans de données de l'enquête de transport Néerlandaise. L'indicateur utilisé pour estimer l'usage de chaque mode est le nombre de déplacements agrégé par jour. Un modèle binomial est utilisé pour expliquer le nombre de déplacements journalier, incluant des variables météorologiques discrétisées et des variables identifiant la saison. Trois analyses sont effectuées, soit une portant sur la demande totale, une sur la demande par mode, et une sur la demande par motif de déplacement. L'étude conclut que les déplacements ayant comme motif le travail sont moins affectés par la météo que les autres motifs. Par ailleurs, un autre modèle de type Tobit estime la demande journalière avec la distance totale parcourue, afin de déterminer la relation entre la

distance parcourue et les variables météo. Un troisième modèle de type multinomial logit est utilisé afin d'estimer les variations de choix modal selon la température de façon désagrégée.

L'étude de Diab et Shalaby (2017) évalue l'impact d'une interruption de service du métro de Toronto sur les services de bus et de tramway. Pour ce faire, l'indicateur utilisé pour estimer la performance des services de bus et de tramway est la vitesse moyenne des véhicules. Une régression linéaire est développée afin d'estimer la réduction de vitesse selon le temps depuis lequel l'interruption de service du métro a débuté. L'impact sur les différentes lignes est identifié indépendamment. L'analyse utilise les données d'interruption de service du métro et les positions GPS des bus et des tramways pour les jours de semaine du mois de mai 2013, durant lequel 388 incidents ont eu lieu. Un rayon de 200 mètres est considéré autour des stations de métro pour identifier les lignes de bus et de tramway qui étaient susceptibles de voir leur achalandage augmenter.

Morency et al. (2016) utilisent 5 flux de données passifs sur une période d'un an pour analyser la demande de divers modes. Les modes considérés sont la voiture, l'autobus, le métro, le vélopartage, le taxi et l'autopartage. Pour comparer l'usage de systèmes de transport ayant des utilisations d'ordre de grandeur très différents, des indicateurs d'usage normalisés sont calculés indépendamment pour chaque mode. Un premier indicateur s'intéresse à la proportion des déplacements effectués à chaque heure d'une journée. Un second indicateur dont la valeur est comprise entre 0 et 1 est défini comme une intensité d'usage journalière. Cette intensité d'usage indique si l'utilisation journalière d'un mode est importante en la comparant à l'usage maximal parmi tous les jours analysés. À l'aide de ces indicateurs, une première analyse unimodale est menée afin d'identifier les patrons d'usage journaliers typiques à l'aide de méthodes de classification. Une seconde analyse cette fois multimodale classifie les jours selon l'intensité d'usage, puis selon le patron journalier d'usage de tous les modes. Finalement, l'impact de certaines variables externes telles que la température et la quantité de précipitations est évaluée en utilisant un arbre de décision.

La présente étude combine plusieurs idées ayant été discutées dans cette revue de littérature. Elle propose d'analyser la demande de 4 différents modes en utilisant des techniques de regroupement, des modèles additifs généralisés (GAM) et des modèles de séries temporelles.

2.4 Zone d'attraction des stations de métro

La zone d'attraction peut être définie comme le territoire aux environs d'une station de transport en commun pour lequel cette station est facilement accessible. La majorité des usagers empruntant cette station proviennent ou se destinent dans cette zone d'attraction (Andersen et Landex, 2009). Un temps de marche à pied est généralement défini afin de tracer les frontières de la zone d'attraction. Dans le cadre de cette étude, la définition d'une zone d'attraction autour des stations de métro sert de zone d'analyse des interactions entre les différents modes de transport.

La technique la plus simple pour définir une zone d'attraction consiste à spécifier un rayon fixe autour de la station de métro. La distance euclidienne est utilisée, soit la distance à vol d'oiseau. Ce rayon peut être choisi à partir d'un temps de marche maximal et d'une vitesse de marche moyenne. Par exemple, l'étude de Farber et Marino (2017) portant sur le métro de Toronto estime la zone d'attraction par un rayon fixe de 800 mètres basé sur un temps maximal de marche de 10 minutes. Similairement, l'étude de Canepa (2007) menée en Oregon (États-Unis) conclut qu'une distance d'environ 0,5 mille (805 mètres) est ciblée pour les stations importantes, tout en précisant que cette distance peut fluctuer selon le contexte de chaque ville. De son côté, l'étude de Sohn et Shim (2010) portant sur le métro de la région métropolitaine de Séoul propose un rayon fixe de 500 mètres, soit une distance de marche raisonnable, tel que le proposait une étude menée en Corée.

Cependant, une limitation importante de cette méthode est que la distance euclidienne ne considère pas la connectivité du réseau. La présence de barrières géographiques telles que des édifices, des voies ferrées ou des cours d'eau causent une distance de marche réelle plus élevée que la distance euclidienne. Une solution possible consiste à ajouter un facteur de détour réduisant le rayon d'attraction pour compenser la distance de marche plus élevée (Andersen et Landex, 2009).

Une méthode encore plus rigoureuse consiste à utiliser la distance réseau. Cette technique calcule la distance réelle de la station en empruntant le réseau routier et même parfois le réseau piétonnier. Un polygone est tracé afin d'inclure tous les segments du réseau accessibles en une certaine distance de la station. Les zones d'attraction obtenues par cette méthode sont donc davantage réalistes car elles tiennent compte du contexte géographique et de la connectivité du réseau (Andersen et Landex, 2009). Ainsi, une partie de la zone d'attraction générée par la méthode du rayon fixe est exclue en utilisant plutôt la distance réseau. Anderson (2009) démontre que la

superficie de cette zone peut être réduite jusqu’à 60% en utilisant cette technique, en se basant sur des stations de transport en commun à Copenhague.

CHAPITRE 3 MÉTHODOLOGIE

Cette section présente tout d'abord le territoire à l'étude. Ensuite, celle-ci présente les données utilisées dans le cadre de cette étude, le pré-traitement de ces données, ainsi que la méthodologie générale du projet.

3.1 Méthodologie générale

La Figure 3-1 présente le diagramme méthodologique général du projet. Les titres sont accompagnés de numéros qui correspondent aux sections du présent mémoire.

Tout d'abord, la revue de littérature présente l'état des connaissances pour différents thèmes abordés dans cette étude. Ensuite, différentes sources de données sont sélectionnées et rassemblées dans une base de données commune. Les données sont traitées de façon à être utilisées dans diverses analyses. Trois différentes analyses sont ensuite complétées.

La première analyse qui porte sur les patrons typiques d'usage a comme objectif de mieux comprendre les cycles d'usage annuels, hebdomadaires et journaliers pour les différents modes.

La deuxième analyse a comme objectif de créer une typologie de stations de métro, soit une classification des stations selon leurs caractéristiques. Cette typologie est ensuite mise en relation avec les patrons journaliers d'usage.

La troisième analyse a comme objectif d'estimer l'impact de différents évènements sur la demande horaire. Tout d'abord, des modèles sur l'ensemble des stations permettent de sélectionner les variables ayant un impact significatif sur la demande et ce, indépendamment pour chaque mode. Ensuite, différents modèles sont calibrés pour chaque type de station identifié par la typologie de stations, en utilisant la sélection de variables portant sur l'ensemble des stations. Finalement, différents modèles sont calibrés pour chaque station afin de permettre leur comparaison.

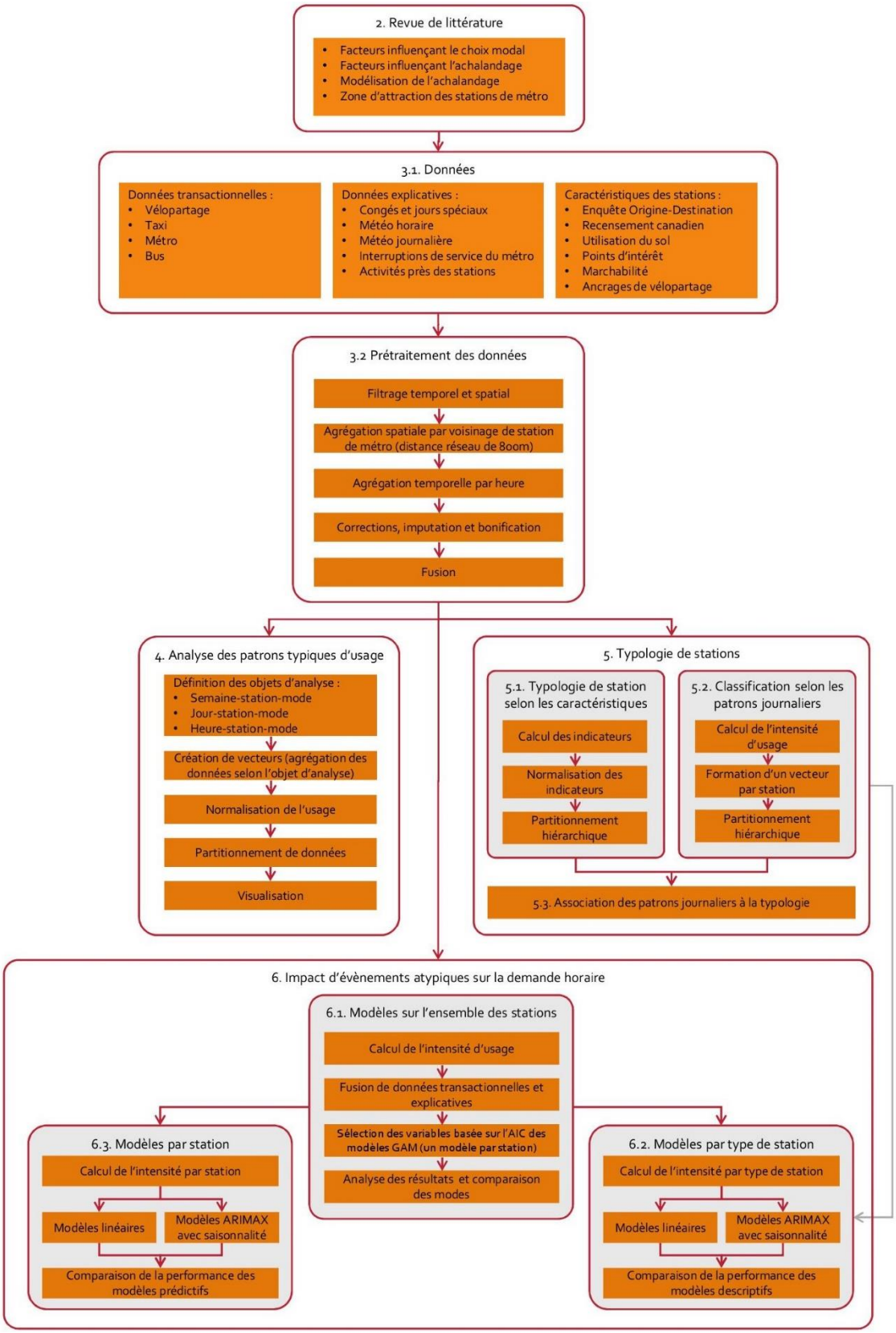

Figure 3-1 : Diagramme méthodologique général

3.2 Territoire à l'étude

Le territoire à l'étude est celui de la région métropolitaine de Montréal. Cette région a été choisie pour son importance, pour la variété de son offre de transport et pour la disponibilité des données. En effet, Montréal est la deuxième région métropolitaine en importance au Canada, avec une population de près de 4,1 M d'habitants en 2017 (Gouvernement du Canada, 2017). L'offre de transport urbaine est variée et celle-ci inclut la voiture personnelle, le transport en commun (métro, bus, train), le taxi, le transport partagé (vélopartage, autopartage), ainsi que différents modes actifs (marche, vélo). Différents types de données transactionnelles sont recueillis par les opérateurs de transport et certaines données sont partagées à des groupes de recherche pour en faire l'analyse.

La Figure 3-2 présente une carte de la région métropolitaine de Montréal.

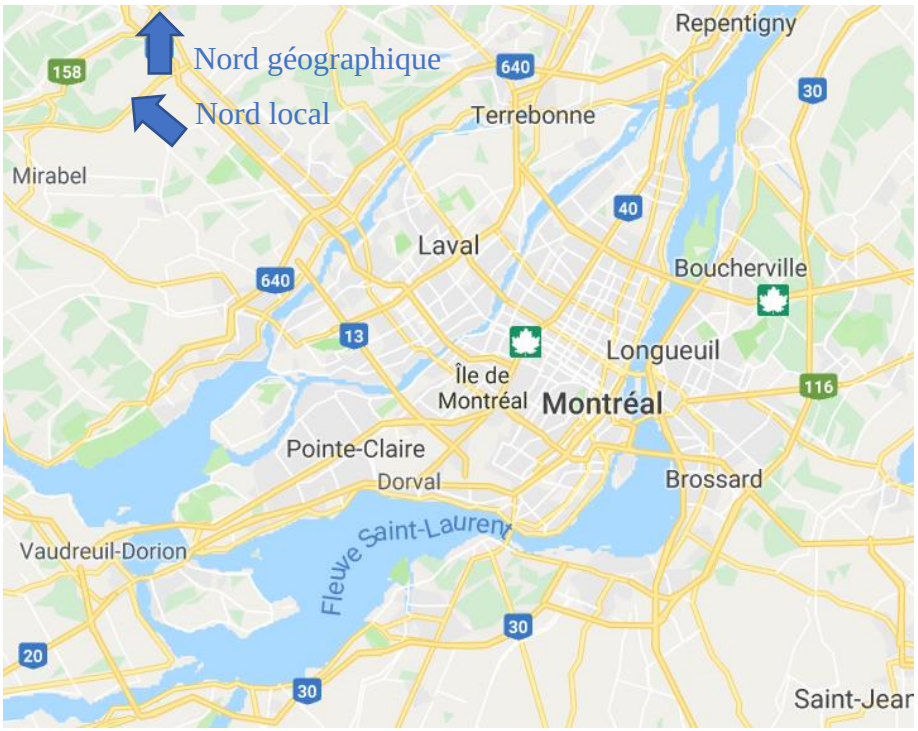

Figure 3-2 : Carte de la région métropolitaine de Montréal (Google, s. d.)

L'appellation locale des points cardinaux à Montréal comporte un certain décalage par rapport aux points cardinaux géographiques. Sur la Figure 3-2 le Nord géographique est orienté vers le haut, alors que l'appellation locale oriente plutôt le Nord avec un décalage anti-horaire d'environ 50 degrés (suivant le plan en damier des rues de la ville). Dans ce mémoire, les points cardinaux seront utilisés selon l'appellation locale.

L'île est bordée au Sud par le fleuve St-Laurent et au Nord par la rivière des Prairies. La Rive-Sud de l'île de Montréal est accessible par l'un des cinq liens routiers traversant le fleuve. De son côté, la Rive-Nord est accessible par une dizaine de ponts traversant la rivière. L'île de Montréal comporte très peu de relief, à l'exception du Mont-Royal situé légèrement au Sud du centre de l'Île. Celui-ci a une hauteur de 232 m au-dessus du niveau de la mer.

La Figure 3-3 présente les différents réseaux de transport qui sont analysés dans cette étude, soit les stations de métro, les arrêts d'autobus, les stations de vélopartage et la zone de taxi A11.

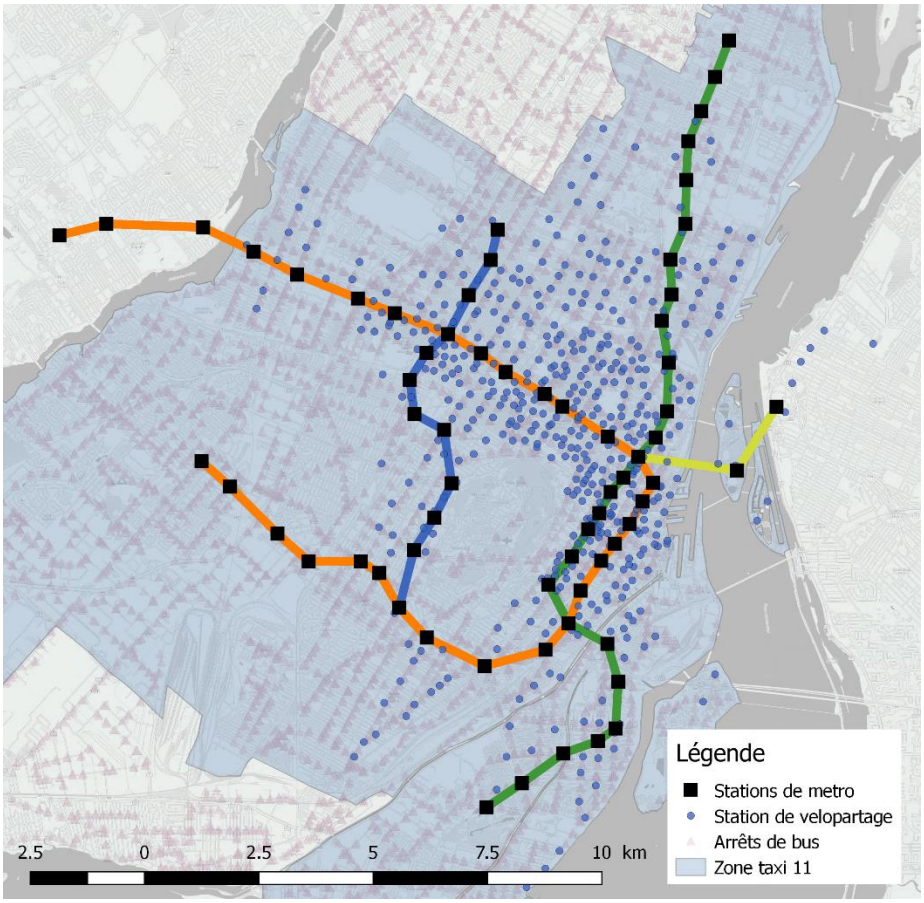

Figure 3-3 : Réseaux de transport à l'étude

Tout d'abord, le réseau de métro comprend 4 lignes distinctes et compte 68 stations (Société de transport de Montréal, 2018). La région où la ligne verte et la ligne orange sont parallèles est désignée comme le centre-ville. La ligne verte relie l'Est et l'Ouest de l'île en passant par le centre-ville. La ligne orange comporte deux branches Nord-Sud, dont l'une qui relie la Rive-Nord, et une section Est-Ouest près du centre-ville. La ligne jaune relie la Rive-Sud au réseau de métro. Finalement la ligne bleue relie les deux branches de la ligne orange selon l'axe Est-Ouest. Le réseau

de métro est opéré par un organisme public nommé la Société de transport de Montréal (STM). Le métro ferme durant la nuit pour une période d'environ quatre heures.

Le service de vélopartage se nomme Bixi et compte près de 7000 vélos répartis dans environ 600 stations (BIXI, 2019). Il s'agit d'un système de vélos en libre-service basé-station qui existe depuis 2009 et dont le nombre de stations augmente chaque année. Les stations et les vélos peuvent être déplacés par l'opérateur selon la demande anticipée. Quelques stations sont présentes sur la Rive-Sud, mais aucune sur la Rive-Nord. Le service est fermé durant la saison froide, soit de la mi-novembre à la mi-avril.

Un réseau de 221 lignes de bus est opéré par la Société de transport de Montréal (STM), dont 23 lignes de bus de nuit (Société de transport de Montréal, 2018). Les réseaux de bus à l'extérieur de l'île de Montréal sont opérés par d'autres sociétés de transport, dont les données n'ont pas été utilisées pour cette étude.

La zone de taxi A11 couve le centre de l'île de Montréal. Les opérateurs de taxi de cette zone ne peuvent pas prendre des clients sur la rue ni à des postes d'attente hors de leur zone. Ils peuvent prendre un client hors zone uniquement si la destination de la course est dans la zone, et que la course a été commandée par téléphone ou via l'application. Les données utilisées pour cette étude proviennent de l'opérateur Taxi Diamond qui opère près de 25% des véhicules dans la zone A11.

3.3 Données

Les données utilisées pour ce projet peuvent être classées en deux catégories, soit les données transactionnelles des systèmes de transport et les données explicatives. Les données transactionnelles correspondent aux données d'achalandage des différents modes de transport. De leur côté, les données explicatives sont des variables externes qui aident à mieux comprendre les variations d'achalandage.

3.3.1 Données transactionnelles

Les modes de transport considérés dans cette étude sont le métro, l'autobus, le vélopartage et le taxi. Ces quatre modes ont été choisis étant donné qu'ils sont tous offerts sur un territoire commun, soit le centre de l'île de Montréal. De plus, des données longitudinales sur l'usage de ces modes

sont disponibles pour la période commune de 2015 à 2017. Les véhicules privés et la marche sont exclus de cette étude, étant donné l'absence de données en continu sur leur usage.

Ces données longitudinales sont collectées par différents systèmes transactionnels pour chaque mode de transport, excepté pour le métro et le bus qui utilisent un système commun. Un déplacement comprenant des transferts modaux sera donc segmenté par mode et ces segments ne pourront pas être regroupés, dû à l'absence d'identifiants communs pour les usagers. Ainsi, cette étude considère chaque segment d'un trajet comme des déplacements indépendants. Cette information est suffisante pour quantifier l'achalandage des différents modes et ainsi permettre de répondre aux objectifs de l'étude.

Différentes données sur les déplacements sont récoltées selon le mode observé. Le tableau suivant présente un résumé des informations collectées par mode, pour la période à l'étude de 2015 à 2017.

Tableau 3-1 : Données transactionnelles collectées selon le mode

Étape du déplacement	Type d'information	Vélopartage	Taxi	Métro	Autobus
Départ	Horodatage	✔	✔	✔	✔
	Localisation	✔	✔	✔	?
En trajet	Trace GPS	✘	✔	✘	✘
Arrivée	Horodatage	✔	✔	✘	✘
	Localisation	✔	✔	✘	✘
Suivi de l'usager	Identifiant unique	✔	✘	✔	✔

✔ : Collecté ✘ : Non collecté ? : Partiellement collecté

Les données du vélopartage sont récoltées par l'opérateur Bixi. Pour chaque déplacement en vélopartage, les stations d'emprunt et de retour sont enregistrées. La position de chaque station étant connue, il est possible d'en déduire les lieux de départ et d'arrivée de chaque trajet. Les usagers réguliers se procurent une clé Bixi qui déverrouille les vélos, ce qui permet d'associer plusieurs déplacements à un même usager.

Du côté du taxi, les données utilisées sont celles de la compagnie Taxi Diamond, qui détient environ 25% de la flotte de taxi opérant dans la zone centrale de l'île de Montréal. Leurs véhicules sont équipés d'un GPS, permettant de connaître la trajectoire précise empruntée pour chaque course de taxi. Les lieux d'origine et de destination des courses peuvent donc être déduites à partir de ces données GPS.

Les données du métro proviennent de la Société de Transport de Montréal (STM). Elles sont collectées via un système de carte à puce nommé Opus. L'usager valide sa carte à puce à l'entrée du réseau de métro. Cependant, la carte à puce n'est pas validée à la sortie du métro, ni lors d'un changement de ligne. Le système de carte à puce permet d'associer les différentes validations de cartes à un même usager. Ces données incluent également les validations de cartes magnétiques jetables, qui sont des supports temporaires qui ne peuvent pas être rechargés contrairement à la carte à puce.

Pour l'autobus, les données proviennent également de la Société de Transport de Montréal (STM) et sont également collectées par le système de carte à puce Opus. La carte à puce est validée à chaque entrée dans un véhicule, mais n'est pas validée à la sortie. Chaque validation de carte à puce est associée à la ligne de bus sur laquelle elle est faite, mais l'arrêt d'embarquement est inconnu. Il n'est donc pas possible d'identifier les lieux de départs et d'arrivée des trajets en bus.

La plage temporelle considérée pour l'analyse est fixée du 1er janvier 2015 au 31 décembre 2017, soit une période continue de 3 ans. Cette période est suffisamment longue pour permettre de bien identifier le cycle annuel d'utilisation des modes et les tendances à plus long terme. Or, cette période est suffisamment courte pour supposer que les comportements de mobilité sont similaires durant la période et qu'il n'y a pas eu de changement majeur dans l'organisation des réseaux de transport à Montréal durant cette période.

Le Tableau 3-2 présente le nombre de jours avec des observations disponibles sur la période de 3 ans, pour chaque mode de transport. Le vélopartage n'a que 58,9% des jours où des données d'utilisation sont disponibles, étant donné que le service est ouvert du 15 avril au 15 novembre seulement. Pour le taxi, il n'y a que 587 jours de données disponibles, puisque les données de 2017 n'étaient pas disponibles au moment où a été complétée cette analyse. Par ailleurs, plusieurs jours de 2015 comportent des données manquantes.

Tableau 3-2 : Jours avec données transactionnelles disponibles

Mode	Jours avec données disponibles	Proportion de la période d'observation
Métro	1093	99,7 %
Autobus	1093	99,7 %
Vélopartage	645	58,9 %
Taxi	587	53,6 %

3.3.2 Données explicatives

Les variables explicatives choisies sont des variables externes pouvant aider à expliquer les variations d'achalandage dans le temps et dans l'espace. Ces données portent sur des objets différents et comportent des contextes spatio-temporels différents. Ces données portent sur les congés fériés, la météorologie horaire, la météorologie journalière, les interruptions de service du métro et les activités à proximité des stations.

3.3.2.1 Congés fériés et période des fêtes

Tout d'abord, les jours fériés au Québec sont listés, étant donné qu'il est attendu que les habitudes de déplacement diffèrent durant ceux-ci. Les jours fériés sont tirés du site de la Commission des normes, de l'équité, de la santé et de la sécurité du travail (CNESST). Les jours compris durant la période des fêtes sont également identifiés. Cette période s'étend du 24 décembre au 2 janvier inclusivement, durant laquelle les habitudes de déplacements risquent également de différer.

3.3.2.2 Météorologie horaire

Les données climatiques proviennent d'une station météorologique de la société Nav Canada située à l'aéroport Pierre-Elliott Trudeau, au Sud-Ouest de l'île de Montréal. Il s'agit de la station météorologique la plus près du centre de l'île de Montréal pour laquelle l'historique des données météo horaires est disponible. Ces données publiques sont publiées sur le site du Gouvernement du Canada. Le Tableau 3-3 liste les variables numériques accompagnées de quelques statistiques descriptives. À ces variables numériques s'ajoute une variable discrète qui est le temps observé. Il

s'agit du phénomène météorologique principal identifié visuellement à la station, parmi une liste d'une trentaine de phénomènes possibles (nuageux, pluie, grêle, brouillard, etc.).

Tableau 3-3 : Statistiques descriptives sur les variables météo horaires

Variable	Moyenne	Médiane	Écart-type	Minimum	Maximum	Données manquantes
Température (°C)	7,8	8,5	12,3	-27,5	33,7	0,05%
Point de rosée (°C)	2,0	2,9	12,1	-33,7	23,8	0,05%
Humidité relative (%)	68,7	70,0	17,1	11,0	99,0	0,05%
Direction du vent (degrés)	196,0	220,0	100,0	0,0	360,0	0,11%
Vitesse du vent (km/h)	16,7	16,0	8,9	0,0	67,0	0,11%
Visibilité (km)	27,4	24,1	12,6	0,2	80,5	0,05%
Pression (kPa)	101,1	101,1	0,9	97,2	104,1	0,05%

Les figures suivantes présentent la distribution de la température et celle de la vitesse du vent, soit deux variables qui risquent d'avoir une influence sur le volume de déplacements. La ligne pointillée verticale indique la valeur moyenne des variables, soit une température de 7,8 °C et une vitesse de vent moyenne de 16,7 km/h. La ligne pointillée grise indique la distribution des variables pour la période de disponibilité du vélopartage uniquement.

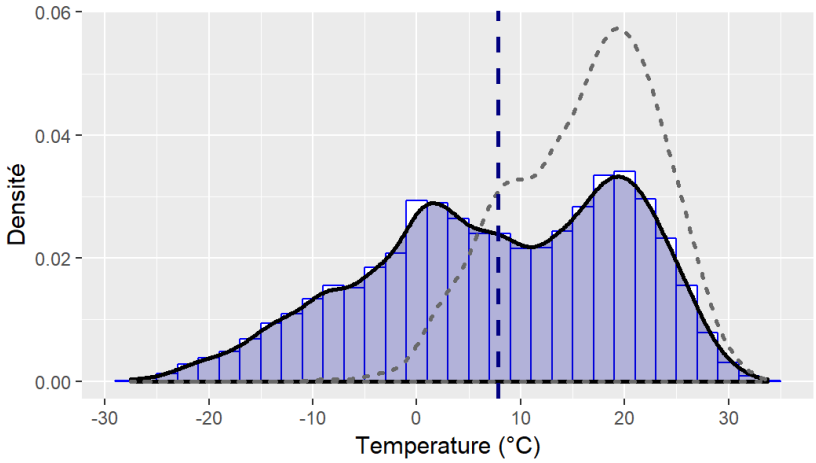

Figure 3-4 : Distribution de la température observée

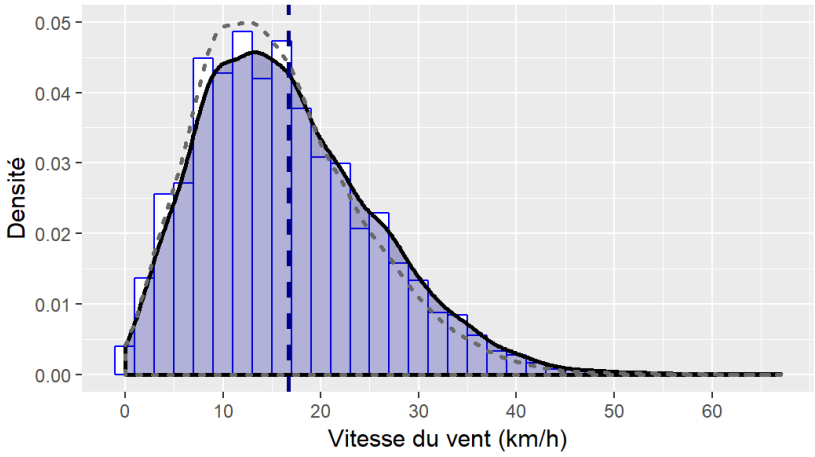

Figure 3-5 : Distribution de la vitesse du vent

3.3.2.3 Météorologie journalière

Les données météorologiques journalières comprennent des informations différentes de celles horaires. Ces données proviennent de la même station météorologique que les données horaires. Le Tableau 3-4 présente la liste des variables journalières ainsi que quelques statistiques descriptives.

Tableau 3-4 : Statistiques descriptives sur les variables météo journalières

Variable	Moyenne	Médiane	Écart-type	Minimum	Maximum	Données manquantes
Température max (°C)	12,3	13,2	12,6	-20,5	33,8	0,36%
Température min (°C)	3,0	3,7	11,8	-27,6	22,2	0,00%
Température moyenne (°C)	7,6	8,3	12,1	-23,7	27,4	0,36%
djc (Degré-jour de chauffage)	11,3	9,8	11,0	0,0	41,7	0,36%
djr (Degré-jour de refroidissement)	1,0	0,0	1,9	0,0	9,4	0,36%
Quantité de pluie (mm)	2,5	0,0	6,5	0,0	67,0	0,00%
Quantité de neige (cm)	0,6	0,0	2,4	0,0	39,2	0,00%
Précipitations totales (cm d'eau)	3,1	0,0	6,9	0,0	67,0	0,00%
Neige au sol (cm)	7,0	5,0	6,6	0,0	29,0	68,06%
Rafale maximale (km/h)	43,2	41,0	12,1	30,0	98,0	0,00%

Les figures suivantes présentent la distribution de la quantité de pluie et de neige. La ligne pointillée verticale indique la valeur moyenne de 2,5 mm de pluie et de 0,6 cm de neige par jour.

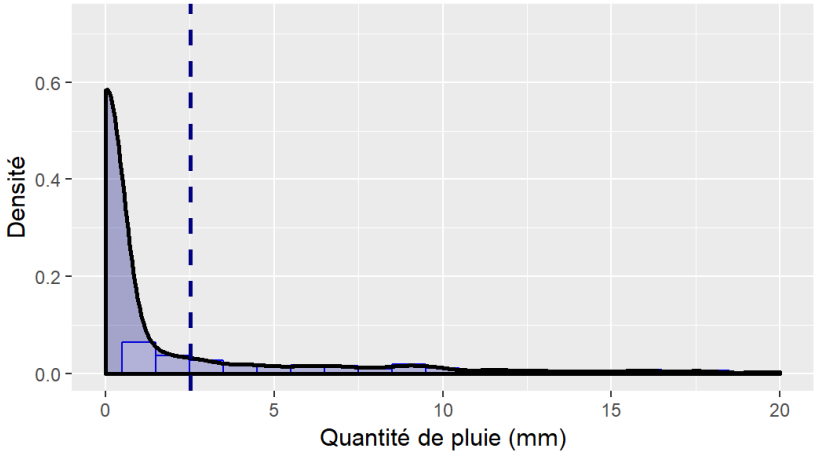

Figure 3-6 : Distribution de la quantité de pluie journalière

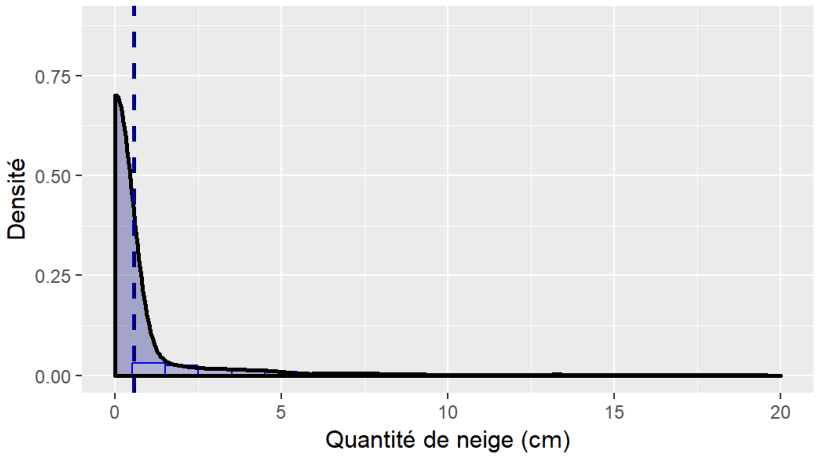

Figure 3-7 : Distribution de la quantité de neige journalière

3.3.2.4 Interruptions de service du métro

Une table répertorie toutes les interruptions de service du métro de Montréal entre 2015 et 2017 d'une durée minimale de cinq minutes. Pour chaque interruption, les informations suivantes sont recueillies : l'horodatage de début, l'horodatage de fin, la durée, la ligne de métro, le lieu d'origine de la panne et la cause. Le lieu d'origine de la panne indique la station ou l'inter-stations où la panne a été initiée.

Le lieu d'origine de la panne est donc connu, mais toutes les stations en arrêt de service dû à cette panne sont inconnues. Par ailleurs, la période durant laquelle chaque station est en arrêt est inconnue, excepté pour la station d'origine.

Un total de 3051 interruptions de service sont survenues pour la période observée. Le Tableau 3-5 présente le nombre d'arrêts de service par ligne de métro. Dans certains cas, un arrêt de service peut concerner plusieurs lignes.

Tableau 3-5 : Nombre d'arrêts de service du métro par ligne

Ligne	Nombre
Verte (1)	1146
Orange (2)	1527
Jaune (4)	131
Bleue (5)	300

La Figure 3-8 présente le nombre d'interruptions par station d'origine, avec une moyenne de 42 pannes par station. Il est possible d'observer que les stations en extrémité de ligne sont plus souvent touchées par des pannes que les autres, ce qui est probablement dû aux manœuvres de retournement en fin de ligne, aux changements de chauffeurs et à l'achalandage généralement élevé de ces stations.

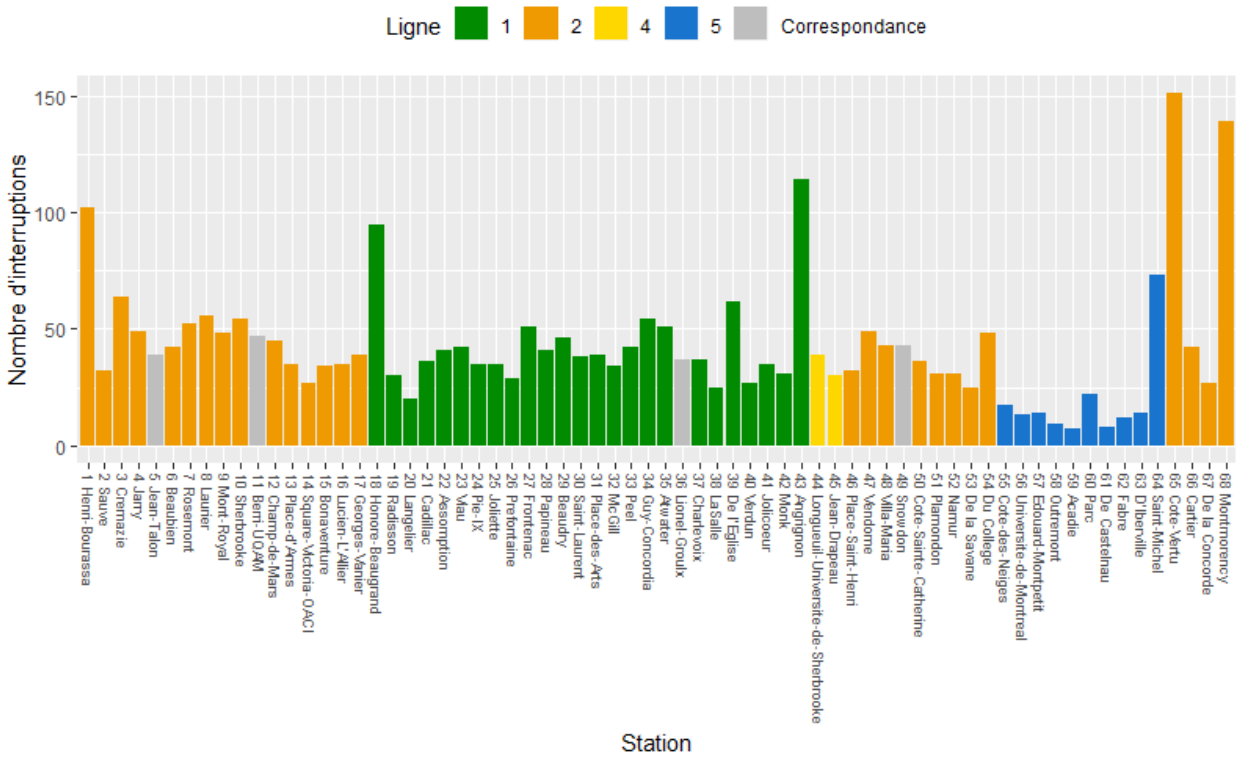

Figure 3-8 : Nombre d'interruptions par station de métro

La Figure 3-9 présente la distribution de la durée des interruptions de service du métro de 5 minutes ou plus. La durée moyenne est de 11,8 minutes et la durée maximale est de 683 min (11h 23min).

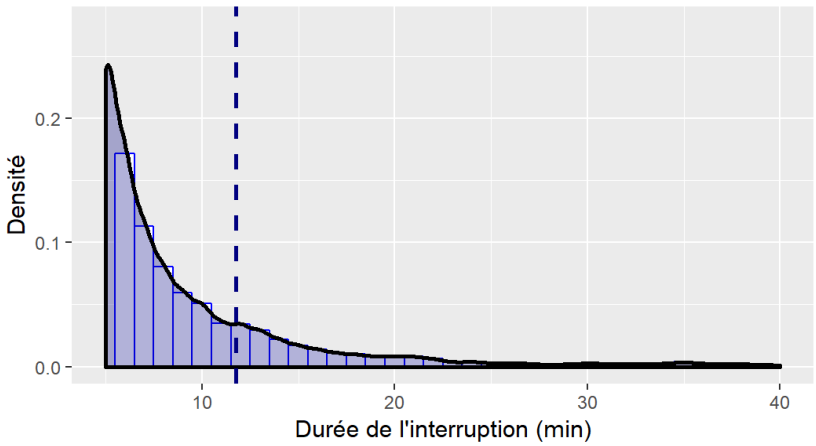

Figure 3-9 : Distribution de la durée des interruptions de service du métro

Lors d'une interruption de courte durée, le service peut se poursuivre sur la majorité de la ligne en ne causant qu'un ralentissement local en amont de l'incident. Le ralentissement est causé par l'obstruction d'un ou plusieurs véhicules qui ne peuvent être dépassés étant donné l'absence de

voie de contournement. Lorsque la durée de la panne augmente et que les véhicules ne sont plus uniformément répartis sur la ligne, les opérateurs peuvent décider d'effectuer un service provisoire. Ce type de service est opéré sur une portion de la ligne qui n'est pas touchée par l'interruption de service et utilise des aiguillages au centre de la ligne pour permettre aux véhicules de changer de direction. Le tableau suivant présente l'ensemble des services provisoires pouvant être effectués. Or, il n'y a aucun historique de l'emploi de chacun de ces services provisoires.

Tableau 3-6 : Services provisoires de métro

Ligne	Station de début	Station de fin
Ligne verte (1)	Lionel-Groulx	Honoré-Beaugrand
	Berri-UQAM	Honoré-Beaugrand
	Angrignon	Lionel-Groulx
	Angrignon	Atwater
	Angrignon	Berri-UQAM
	Angrignon	Pie-IX
Ligne orange (2)	Côte-Vertu	Henri-Bourassa
	Côte-Vertu	Berri-UQAM
	Côte-Vertu	Crémazie
	Côte-Vertu	Beaubien
	Côte-Vertu	Lionel-Groulx
	Henri-Bourassa	Montmorency
	Berri-UQAM	Montmorency
	Lionel-Groulx	Montmorency
Ligne bleue (5)	St-Michel	Parc

3.3.2.5 Activités

Une base de données regroupe 1772 activités importantes (telles que des évènements sportifs, des festivals ou des concerts) ayant lieu autours des stations de métro entre 2015 et 2017. Ces données sont compilées par la Société de Transport de Montréal (STM) pour les aider à adapter l'offre de transport à la demande élevée lors d'activités. Les champs de cette table sont le nom de l'activité, l'horodatage de début, l'horodatage de fin et la station de métro la plus proche du lieu d'activité. Les activités ont été catégorisées selon la méthode de Toqué (2019), soit une méthode semi-automatisée d'identification à partir de mots-clés compris dans le titre des activités. La Figure 3-10 présente le nombre d'activités par catégorie.

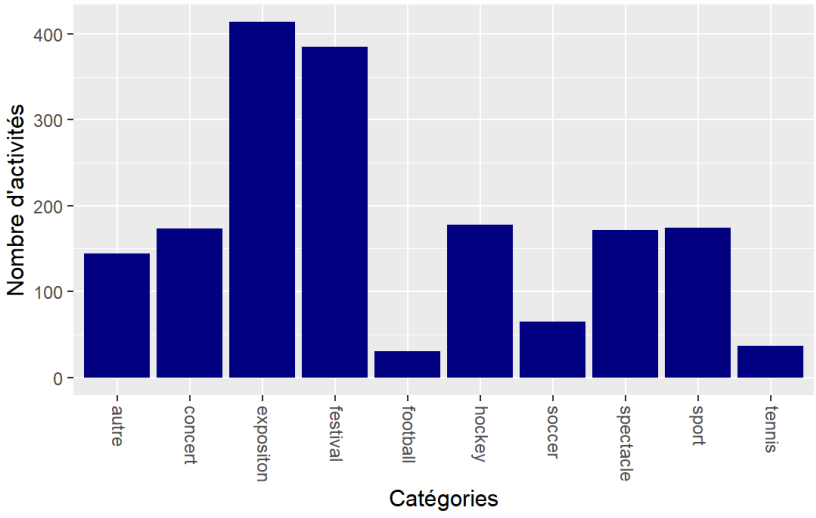

Figure 3-10 : Nombre d'activités par catégorie

Pour continuer, la Figure 3-11 présente le nombre d'activités par station de métro. La station Lucien-L'Allier est celle qui compte le plus d'activités, avec 469 activités dont plusieurs ont lieu au Centre-Bell.

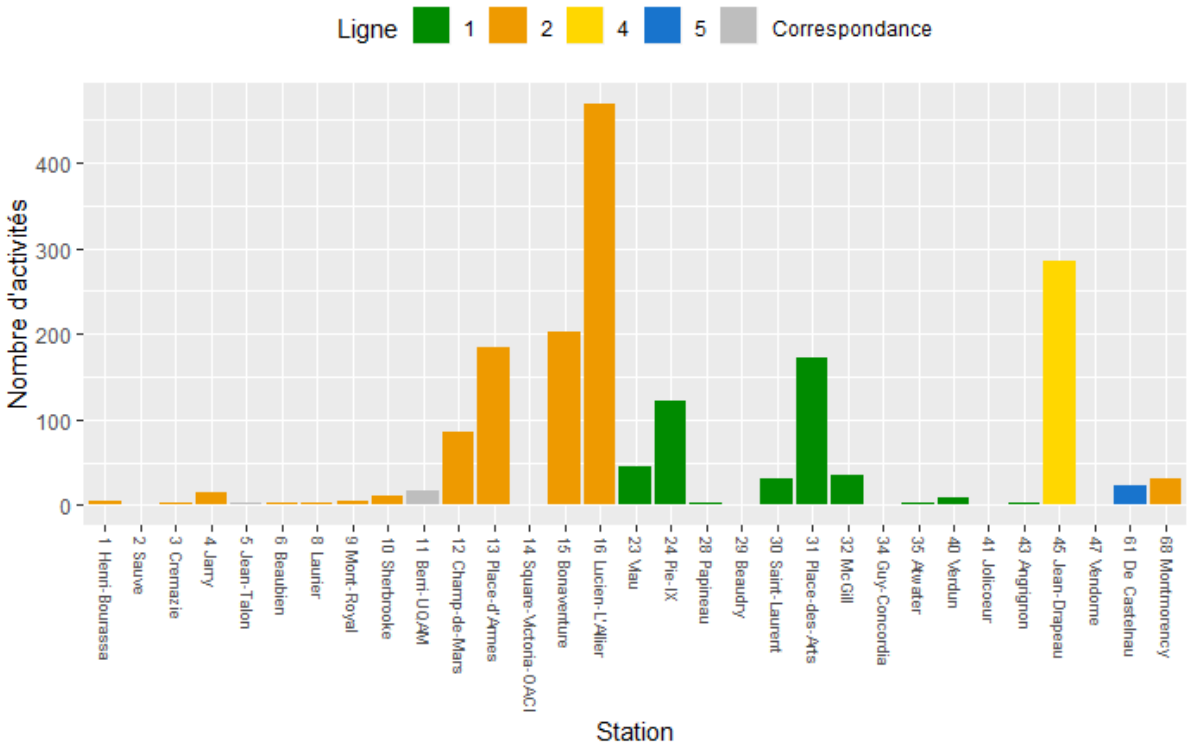

Figure 3-11 : Nombre d'activités par station

3.3.3 Caractéristiques du voisinage des stations

Les caractéristiques du voisinage des stations de métro peuvent être utiles afin de mieux comprendre les facteurs qui influencent la demande aux différentes stations. Plusieurs sources de données sont utilisées tel que l'enquête Origine-Destination, le recensement canadien, l'utilisation du sol et les points d'intérêt. Le Tableau 3-5 liste les caractéristiques ayant été sélectionnées dans l'objectif de distinguer les stations l'une de l'autre.

Tableau 3-7 : Caractéristiques du voisinage des stations

Variable	**Type**	**Source**	**Description**
population	continue	Enquête OD	Nombre de résidents
homme	continue	Enquête OD	Proportion d'hommes (%)
age_moy	continue	Enquête OD	Âge moyen des résidents
age0_14	continue	Enquête OD	Proportion de personnes âgées de 0 à 14 ans (%)
age15_24	continue	Enquête OD	Proportion de personnes âgées de 15 à 24 ans (%)
age25_39	continue	Enquête OD	Proportion de personnes âgées de 25 à 39 ans (%)
age40_64	continue	Enquête OD	Proportion de personnes âgées de 40 à 64 ans (%)
age65_plus	continue	Enquête OD	Proportion de personnes âgées de 65 ans et plus (%)
revenu_med	continue	Recensement	Revenu médian selon les aires de diffusion
prop_inf20k	continue	Recensement	Proportion de la population âgée de 15 ans et plus avec un revenu après impôt inférieur à 20 000$ (%)
depl_travail	continue	Enquête OD	Nombre de déplacements avec motif travail dont la destination est dans le voisinage
marchabilite	continue	Lefebvre-Ropars et Morency	Indicateur de marchabilité prenant en compte la densité résidentielle, la densité d'intersections, la mixité des usages et la densité d'occupation du sol des commerces. Une valeur élevée indique un voisinage de station propice à la marche.
acces_auto	continue	Enquête OD	Taux moyen d'accès à une automobile (ratio entre le nombre d'automobiles et le nombre de personnes possédant un permis de conduire dans le ménage)
nb_ancrages_bixi	continue	Bixi	Nombre total d'ancrages de vélopartage
aire_residentielle	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type résidentielle (code 100)
aire_commerciale	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type commerciale (code 200)
aire_bureau	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type bureau (code 300)
aire_institutionnelle	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type institutionnelle (code 500)
aire_parc	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type parc (code 600)
aire_utilite_publique	continue	CMM	Proportion de l'aire dont l'utilisation du sol est de type utilité publique (code 700)
nb_ecole_publique	continue	DMTI	Nombre d'écoles primaires et secondaires publiques
cegep	binaire	DMTI	Présence d'au moins un cégep
université	binaire	DMTI	Présence d'au moins une université
hopital	binaire	DMTI	Présence d'au moins un hôpital
temps_cv_metro	continue	Google Maps	Temps du déplacement en métro vers le centre-ville (station Bonaventure ou Peel) à l'heure de pointe
nb_activites	continue	STM	Nombre d'activités totales associées à la station de métro par la STM

Le Tableau 3-8 présente quelques statistiques descriptives sur les différentes variables identifiées pour effectuer le partitionnement de données.

Tableau 3-8 : Statistiques descriptives sur les caractéristiques des stations

Variables	Moyenne	Médiane	Écart-type	Minimum	Maximum
population	16193	15012	7750	3307	36720
homme	50,28	49,30	3,86	43,82	60,32
age_moy	39,16	38,55	3,22	33,04	52,81
age0_14	12,06	12,86	4,26	4,18	22,05
age15_24	13,57	12,25	5,40	5,44	35,27
age25_39	29,77	30,49	6,58	14,83	43,27
age40_64	30,12	30,28	4,57	20,40	40,67
age65_plus	14,48	13,03	5,31	7,71	36,69
revenu_med	26747	26908	4879	13374	39853
prop_inf20k	35,69	35,91	6,61	15,71	48,21
depl_travail	25531	8557	42358	1021	173724
marchabilite	17,63	14,70	9,47	5,76	38,87
acces_auto	0,50	0,49	0,10	0,30	0,81
nb_ancrages_bixi	1522	897	1803	0	7158
aire_residentielle	33,56	34,15	12,44	8,33	54,89
aire_commerciale	6,98	5,37	5,62	0,20	33,73
aire_bureau	3,89	1,39	5,70	0,02	24,93
aire_institutionnelle	11,45	8,67	9,51	1,52	57,91
aire_parc	4,92	4,12	5,27	0,16	41,27
aire_utilite_publique	33,12	33,76	7,23	14,29	49,36
nb_ecole_publique	3,27	3	2,43	0	10
cegep	0,13	0	0,34	0	1
universite	0,15	0	0,36	0	1
hopital	0,27	0	0,45	0	1
temps_cv_metro	11,96	12	6,87	0	26
nb_activites	22,19	0	69,71	0	469

Les sous-sections suivantes fournissent de l'information supplémentaire sur les différentes sources de données pour les caractéristiques des stations

3.3.3.1 Enquête Origine-Destination

L'enquête Origine-destination 2013 menée dans la région de Montréal renseigne sur la composition de la population et sur les déplacements effectués lors d'un jour typique de semaine d'automne. L'enquête de 2013 est celle la plus récente disponible au moment de la publication de cette étude.

C’est pourquoi celle-ci a été utilisée malgré le fait qu’elle ait été menée à l’extérieur de la période d’intérêt de 2015 à 2017. Les facteurs de pondération du Ministère des Transports du Québec sont utilisés afin d’étendre les données de l’échantillon sur l’ensemble de la population de référence, en se basant sur les caractéristiques socio-démographiques.

3.3.3.2 Recensement canadien

Le recensement canadien de 2016 permet d’obtenir de l’information sur le revenu des ménages. Des statistiques sur le revenu après impôt sont disponibles pour chaque aire de diffusion. Le revenu par aire de diffusion peut être estimé en pondérant les revenus des différentes aires de diffusion à l’intérieur de la zone de voisinage, proportionnellement à la surface.

3.3.3.3 Utilisation du sol

Des données de la Communauté métropolitaine de Montréal (CMM) attribuent à chaque parcelle du sol un type d’utilisation parmi 11 classes prédéterminées (CMM, 2016).

3.3.3.4 Points d’intérêt

La base de données des points d’intérêts EPOI CanMap de la compagnie DMTI permet d’identifier plusieurs informations sur les points d’intérêt sur l’ensemble du territoire canadien. Ces données permettent d’identifier le nombre d’opportunités de certains types autour de chaque station de métro.

3.3.3.5 Marchabilité

Le *Pedestrian Index of the Environment* (PIE) est un indicateur qui évalue l’aisance à se déplacer à la marche sur un territoire donné. Cet indicateur de marchabilité a été développé par Frank et al. (2010) et a été adapté au contexte montréalais par Lefebvre-Ropars et Morency (2018). Cet indicateur prend en compte la densité résidentielle, la densité d’intersections, la mixité des usages et la densité d’occupation du sol des commerces. Plus la valeur de cet indicateur est élevée, plus le voisinage de la station est propice à la marche.

3.4 Prétraitement des données

L'objectif du prétraitement des données est de préparer les données pour afin de faciliter leur analyse. Les différentes bases de données sont donc filtrées, agrégées, corrigées, imputées, bonifiées et finalement fusionnées, tel que décrit ci-dessous.

3.4.1 Filtrage

Premièrement, les données sont sélectionnées selon leur cadre spatio-temporel. Un filtre temporel est appliqué afin de ne conserver que les données entre 2015 et 2017 inclusivement.

Un filtre spatial est aussi appliqué afin de ne conserver que les données situées dans la zone d'attraction des stations de métro, soit le territoire sur lequel cette station est jugée facilement accessible.

3.4.2 Agrégation spatiale

Deuxièmement, les déplacements sont agrégés par voisinage des stations de métro selon leur lieu de départ. Le voisinage des stations de métro est défini selon la zone la plus probable de laquelle proviennent les usagers. À la suite de la revue de littérature, cette zone a été fixée à une distance réseau de 800 mètres à partir des édicules des stations de métro. La distance considérée est calculée sur le réseau routier et non à vol d'oiseau, de façon à prendre en compte la tortuosité des rues et les obstacles physiques. Les zones d'influence des stations ont été tracées sur la Figure 3-12.

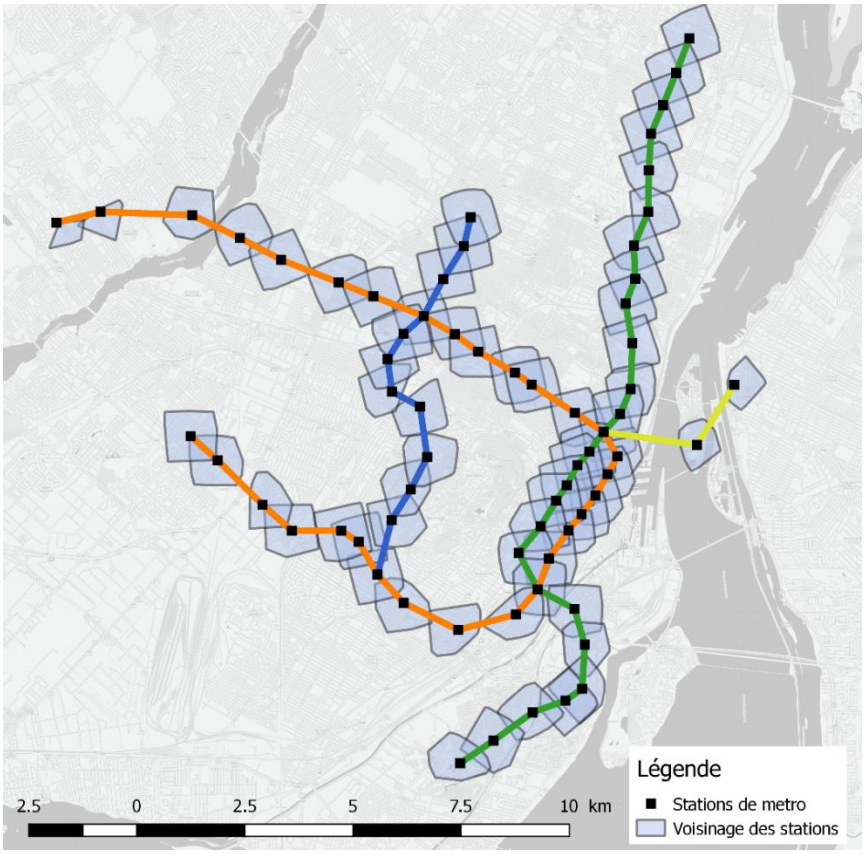

Figure 3-12 : Zones d'influence des stations de métro

Tel qu'illustré sur cette figure, les zones d'influence sont non-exclusives (elles peuvent se chevaucher) ce qui signifie qu'un déplacement peut être compté dans plusieurs zones. Les zones qui se chevauchent de façon importante risquent de voir leurs usages fortement corrélés. La somme des déplacements des différentes zones n'est donc pas pertinente, puisque certains déplacements seraient comptés plus d'une fois. Cependant, la moyenne des déplacements aux stations pourrait être pertinente pour donner un portrait global de l'île de Montréal.

Une intersection spatiale est donc effectuée entre les lieux de départ des déplacements et les zones d'influence des stations. La Figure 3-13 présente un exemple de zone d'influence pour la station Édouard-Montpetit en bleu foncé. Pour le vélopartage, le taxi et le bus, tous les départs effectués dans une zone sont additionnés. Pour le métro, les déplacements ne sont pas agrégés selon cette zone d'influence, mais sont directement attribués à la zone les entourant.

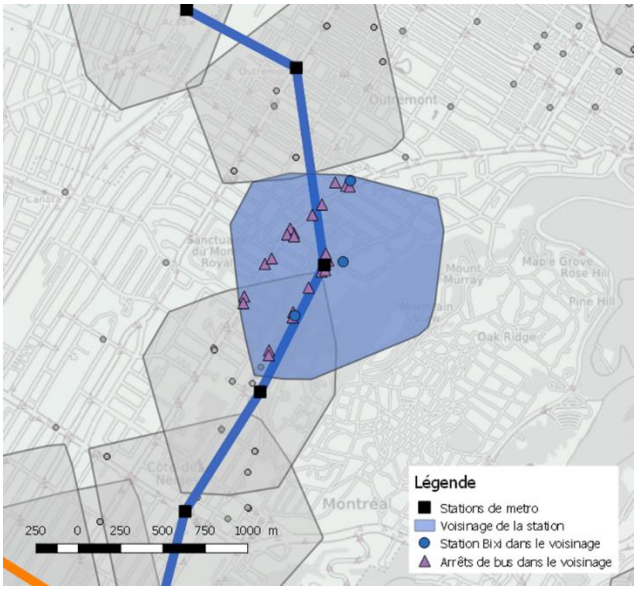

Figure 3-13 : Voisinage de la station de métro Édouard-Montpetit

3.4.3 Agrégation temporelle

Troisièmement, les données transactionnelles sont agrégées en période d'une heure de façon à obtenir le nombre total de départs par période horaire pour chaque combinaison de station et de mode. L'intervalle d'une heure a été sélectionné de façon à être suffisamment court pour capturer les changements d'utilisation dus aux évènements atypiques, sans toutefois être trop court pour éviter de surcharger la base de données.

3.4.4 Corrections

Quatrièmement, il est nécessaire de nettoyer et d'apporter quelques corrections aux différentes tables, afin de standardiser l'information et de corriger certaines informations erronées.

Une première correction concerne la nomenclature des stations de métro. Le nom des stations de métro n'est pas constant pour les différentes bases de données utilisées. Les différentes désignations des stations ont donc été remplacées par les numéros de station standards de la STM allant de 1 à 68, de façon à les rendre uniques.

Les données d'utilisation du métro ont également dû être corrigées pour le mois de juillet 2015. Pour chaque jour de juillet 2015, les déplacements de 8h ont été combinés par erreurs avec ceux de 7h. Il est donc nécessaire de distribuer les validations entre ces 2 heures. Pour ce faire, la proportion

des validations faits pour chaque heure a été calculée pour les mois précédents et suivants celui erroné. Cette même proportion a été estimée indépendamment pour chaque station et pour le type de jour (semaine ou fin de semaine). Ces proportions ont ensuite été utilisées pour distribuer les validations entre 7h et 8h.

Les données Bixi à l'extérieur de la période d'ouverture du 15 avril au 15 novembre sont retirées, puisqu'il est supposé qu'elles correspondent à des tests menés par l'opérateur.

3.4.5 Imputation

Cinquièmement, quelques données manquantes sont imputées afin d'assurer la continuité des données et afin d'éviter d'exclure un grand nombre de données incomplètes.

3.4.5.1 Météorologie

Tout d'abord, quelques données météorologiques journalières manquantes ont été comblées par des données d'une autre station météorologique, soir celle de l'Aéroport de St-Hubert, situé à environ 15km à vol d'oiseau du réseau de métro. Par ailleurs, plusieurs descriptions qualitatives manquantes pour la météo horaire ont été imputées. Ce champ comportait au maximum deux heures consécutives de données manquantes. Il a donc été décidé de remplacer une donnée manquante par celle de l'heure la plus près avec une donnée disponible. La description imputée provient donc de l'heure précédente ou suivante.

3.4.5.2 Activités

Au sein de la table de données d'activités, plusieurs heures de début et de fin sont manquantes. Or, ces informations sont essentielles afin d'analyser la demande de transport à l'échelle horaire, car la date à elle seule n'est pas suffisante pour estimer l'impact réel des activités sur la demande. En effet, une activité a un impact très concentré autour de sa plage horaire, mais impacte généralement peu le reste de la journée. Au total, 538 activités ont des heures de début ou de fin incomplètes, ce qui représente 30% des activités. La méthode d'imputation suivante permet donc d'augmenter le nombre d'activités qui pourront être utilisées dans cette étude.

Pour commencer, les heures de début d'activités sont imputées. La méthode d'imputation se base sur la catégorie des activités, en supposant que les activités d'une même catégorie ont une heure de début similaire. L'heure de début médiane est déterminée pour chaque catégorie à l'aide des

activités dont l'heure de début est connue. La distribution des heures de début pour chaque catégorie est présentée à la Figure 3-14. Sur cette figure, l'heure médiane est identifiée par un point.

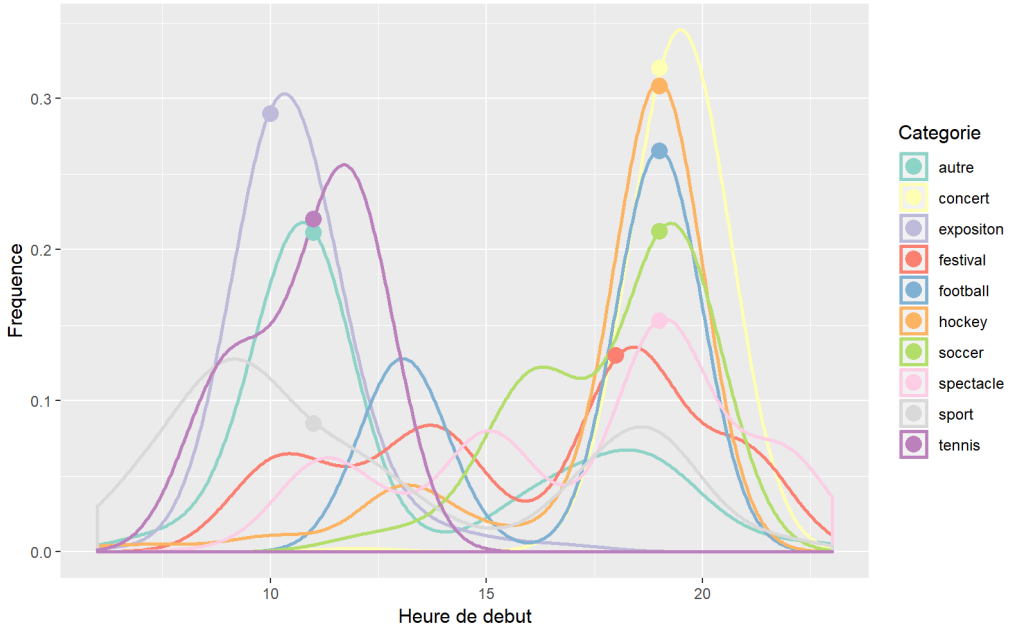

Figure 3-14 : Distribution des heures de début d'activité par catégorie

Il est possible d'observer que l'heure médiane est représentative pour certaines catégories alors qu'elle l'est moins pour d'autres. Pour la catégorie des concerts par exemple, l'heure de début médiane de 19h semble bien représenter l'ensemble des événements, étant donné que la distribution est concentrée autour de cette heure. L'heure médiane pourrait donc être utilisée afin d'imputer les heures de début manquantes pour cette catégorie. De son côté, la catégorie des festivals présente une distribution très étendue durant la journée. L'heure de début médiane de 18h ne semble donc pas représentative de cette catégorie.

Afin de déterminer si l'heure de début d'une catégorie d'activités peut être imputée, il est nécessaire d'évaluer si l'heure médiane est représentative de la distribution d'heures. Une marge d'erreur d'une heure est tolérée, étant donné qu'une telle estimation permet tout de même d'obtenir une bonne approximation de l'étendue d'une activité, permettant d'estimer son impact sur la demande en transport. Un taux d'erreur de prédiction est estimé par la proportion des heures de début à plus d'une heure de la médiane et ce, pour chaque catégorie d'activité. Un taux d'erreur maximal de 20% est toléré pour procéder à l'imputation des données. Le Tableau 3-9 présente la proportion des heures de début manquantes et la décision d'imputer ou non les heures pour chaque catégorie.

Tableau 3-9 : Heures de début d'activité imputées par catégorie

Catégorie	Nombre d'activités	Heures manquantes	Heure médiane	Taux d'erreur de prédiction (tolérance de ±1h)	Heures imputées
Autre	144	6,9 %	11	38,8 %	Non
Concert	173	0,6 %	19	1,2 %	Oui
Exposition	414	5,3 %	10	12,8 %	Oui
Festival	385	31,4 %	18	62,5 %	Non
Football	30	0 %	19	33,3 %	-
Hockey	178	0 %	19	21,9 %	-
Soccer	65	0 %	19	40,0 %	-
Spectacle	172	0,6 %	19	60,2 %	Non
Sport	174	21,3 %	11	73,7 %	Non
Tennis	37	0 %	11	29,8 %	-

Un total de 192 activités présente une heure de début manquante et ces activités sont réparties dans 6 catégories. Les échantillons d'heures valides contiennent suffisamment d'observations afin de prédire les heures manquantes. Les taux d'erreur de prédiction varient de façon importante selon la catégorie d'activité. Le taux d'erreur le plus faible est de 1,2% pour la catégorie concert, ce qui signifie que 98,8% des activités ont une heure de début comprise entre 18h et 20h (l'heure médiane étant de 19h). Les catégories respectant l'erreur maximale de 20% sont les catégories concert et exposition. L'heure de début des concerts est donc imputée à la médiane de 19h, alors que celle des expositions est imputée à 10h. Les activités des autres catégories qui présentent des heures de début manquantes qui ne peuvent être imputées sont donc supprimées.

Pour continuer, une démarche similaire est complétée afin d'imputer les heures de fin des activités. La durée des activités est d'abord imputée, puis l'heure de fin est calculée à partir de l'heure de début et de la durée. Cela permet une meilleure prédiction de l'heure de fin en considérant que celle-ci varie davantage que la durée. La Figure 3-15 présente la distribution des durées d'activité par catégorie.

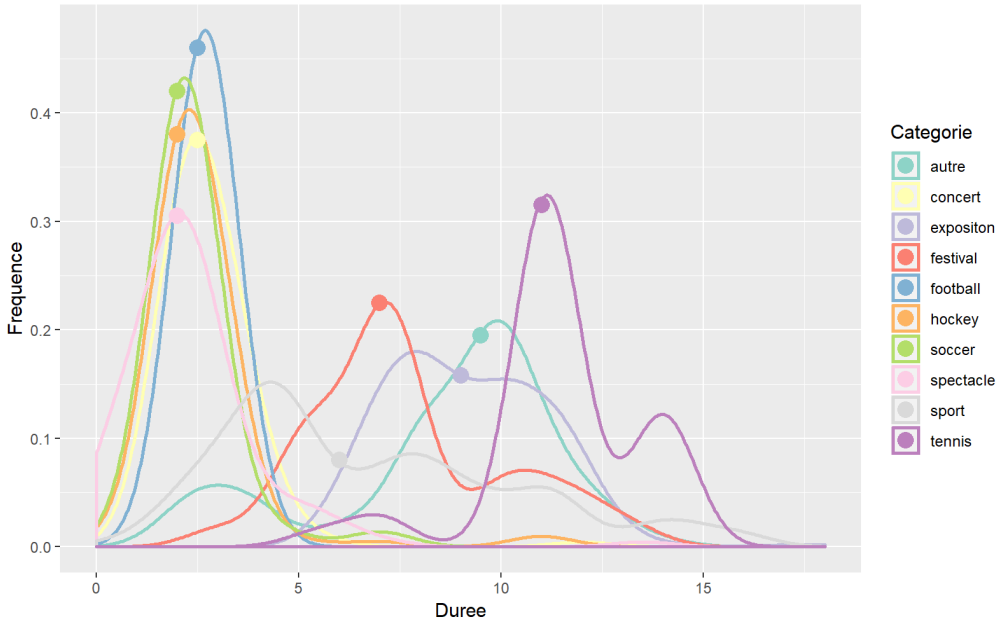

Figure 3-15 : Distribution des durées d'activité par catégorie

Plusieurs catégories d'activités ont une distribution très concentrée autour de durées de 2h à 2h30, telles que les catégories concert, football, hockey et soccer. Le taux d'erreur de prédiction est calculé pour chaque catégorie en considérant une tolérance d'une heure. Le Tableau 3-10 présente les durées manquantes et la décision considérant l'imputation de la durée et ce, pour chaque catégorie.

Tableau 3-10 : Durées d'activité imputées par catégorie

Catégorie	Nombre d'activités	Durées manquantes	Durée médiane (h)	Taux d'erreur de prédiction (tolérance de ±1h)	Durées imputées
Autre	144	15,7 %	9,5	58,4 %	Non
Concert	173	33,5 %	2,5	11,3 %	Oui
Exposition	414	5,3 %	9	50,5 %	Non
Festival	385	14,8 %	7	48,4 %	Non
Football	30	30,0 %	2,5	0,0 %	Oui
Hockey	178	42,1 %	2	4,9 %	Oui
Soccer	65	44,6 %	2	5,6 %	Oui
Spectacle	172	35,7 %	2	29,1 %	Non
Sport	174	19,0 %	6	76,6 %	Non
Tennis	37	0 %	11	32,4 %	-

Un total de 367 activités présente une durée manquante. Seule la catégorie tennis n'a aucune activité dont la durée est manquante. La durée médiane des activités varie de façon importante selon la catégorie, d'où l'importance d'imputer les durées individuellement pour chaque catégorie. En effet, la durée médiane varie de 2h à 11h. Les catégories d'activité pouvant être imputées sont les catégories concert, football, hockey et soccer. Ces catégories ont toutes des durées médianes entre 2h et 2,5h, laissant croire que les activités plus longues semblent présenter une durée plus difficilement prévisible. Ainsi, les durées de 171 activités sont imputées. L'heure de fin de ces activités est calculée en additionnant la durée à l'heure de début. Les activités des autres catégories dont la durée n'a pu être imputée sont donc supprimées.

Bref, l'imputation des heures de début et de fin permet de conserver 192 activités dont au moins l'une des deux heures était manquante. La table des activités comprend un total de 1456 activités valides dont le lieu et les heures sont connus.

3.4.6 Bonification

Sixièmement, quelques tables sont bonifiées en ajoutant des variables dérivées. Certaines variables météorologiques dérivées sont calculées, tel que le nombre d'heures de pluie par jour, le nombre d'heures de neige par jour, le nombre d'heures de pluie en matinée, le nombre d'heures de pluie dans les 3 heures précédentes, ou la variation de température par rapport à la moyenne saisonnière. Les données d'interruption de service du métro sont bonifiées en ajoutant une variable pour la durée de chaque interruption basée sur ses heures de début et de fin. Ces variables seront utilisées dans des analyses afin d'identifier les facteurs ayant un impact significatif sur l'usage de chaque mode.

Les déplacements ayant été effectués peu après minuit sont généralement liés à des activités débutées la journée précédente. Il est donc décidé que les déplacements ayant été effectués jusqu'à 3 heure du matin seront affectés à la journée précédente. L'horodatage des bases de données transactionnelles est donc remplacé par un champ de date et d'heure corrigés. Les heures entre 0h et 3h du matin seront nommées 24h à 27h afin d'indiquer qu'elles sont associées à la journée précédente.

3.4.7 Fusion

Septièmement, les nombres de départs pour chaque mode sont fusionnés dans une table commune afin de simplifier leur comparaison. Trois dimensions sont présentes, soit le temps, la station et le mode. La jointure des tables est effectuée sur les champs communs date, heure et station. Un extrait de la table résultante de la fusion est présenté au Tableau 3-11. Les modes de transport sont en colonne de façon à faciliter leur comparaison. Cette table a comme clé primaire composée les champs date, heure et station.

Tableau 3-11 : Extrait de la base de données transactionnelle

station	date	heure	metro	autobus	vélopartage	taxi
15	2015-01-01	22	248	37	0	31
15	2015-01-01	23	198	47	0	35
15	2015-01-01	24	79	11	0	35
15	2015-01-01	25	0	3	0	51
15	2015-01-01	26	0	0	0	60
15	2015-01-01	27	0	0	0	54
15	2015-01-02	4	0	0	0	9
15	2015-01-02	5	31	20	0	11

Les variables explicatives ne sont pas fusionnées dans cette même table pour éviter la redondance. En effet, les données transactionnelles dépendent du temps et de l'espace, alors que les variables explicatives dépendent majoritairement soit de l'un ou de l'autre. Les fusionner nécessiterait donc de répéter les mêmes valeurs des variables explicatives sur plusieurs lignes de la table transactionnelle, ce qui surchargerait inutilement la table. Seules les variables explicatives pertinentes seront ajoutées aux données transactionnelles selon l'analyse effectuée.

Toutes les tables de données traitées sont stockées dans une base de données sur un serveur local PostgreSQL, de façon à pouvoir accéder aux données à partir de plusieurs logiciels. Ainsi, les logiciels utilisés pour consulter cette base de données sont PgAdmin pour la gestion de la base de données, R pour les statistiques et QGIS pour la visualisation de données spatiales. Patrons typiques d'usage

L'achalandage des divers modes de transport fluctue à travers le temps en suivant certains cycles. En effet, les cycles annuels, hebdomadaires et journaliers introduisent des variations d'achalandage qui se répètent à intervalles réguliers. Les différents modes de transport ne sont pas tous affectés de la même façon par ces cycles. Par ailleurs, ces fluctuations peuvent aussi différer selon le cadre spatial.

CHAPITRE 4 ANALYSE DES PATRONS TYPIQUES D'USAGE

Cette section s'intéresse à l'analyse des patrons types d'utilisation, en ayant recours à du partitionnement de données. L'objectif est de mieux comprendre la dynamique présente entre les modes de transport et d'identifier les différences spatiales. Dans cette section, les patrons annuels seront étudiés, suivi des patrons hebdomadaires et des patrons journaliers.

4.1 Patrons annuels

Le cycle annuel a une période de 365 jours ou 366 jours pour les années bissextiles. Combiner trois années de données permet d'obtenir une année moyenne. L'unité temporelle utilisée est la semaine, soit une durée suffisamment courte pour observer les variations ponctuelles, et suffisamment longue pour éviter de capturer les cycles hebdomadaires. L'unité d'analyse est la combinaison d'une semaine, du mode et de la station. Seuls les jours ouvrables sont considérés pour cette analyse, soit les jours où l'usage est le plus uniforme en raison d'un usage majoritairement utilitaire.

4.1.1 Méthodologie

Les données utilisées sont celles des déplacements horaires par mode et par station. Les données sont tout d'abord agrégées par jour, de manière à obtenir le nombre total de déplacements par jour-station-mode. Les données sont filtrées pour ne conserver que les jours ouvrables. Les données de taxi sont filtrées pour ne conserver que les trajets dont l'origine est dans la zone A11, soit la zone de taxi couvrant le centre de l'île de Montréal. Les quelques trajets dont l'origine est sur la Rive-Sud ou la Rive-Nord sont retirés étant donné qu'ils sont peu nombreux. Plusieurs jours d'opération sont exclus pour le taxi dû au manque de données pour certaines heures de la journée.

Ensuite, chaque jour se fait attribuer un numéro de semaine entre 1 et 53 selon le système de numérotation de la norme ISO 8601. Selon cette norme, les semaines débutent le lundi et la première semaine de l'année est celle qui inclut le premier jeudi de l'année civile. La moyenne du nombre de déplacements par numéro de semaine est calculée.

Le nombre de déplacements est ensuite normalisé afin de simplifier la comparaison entre les stations et les modes ayant des nombres de déplacements d'ordres de grandeur différents. La

normalisation se fait par rapport au nombre annuel de déplacements pour une combinaison de station et de mode, tel qu'exprimé par la formule suivante.

$$i_{t,m,s} = \frac{e_{t,m,s}}{r_{t,m}}$$

où

i : Intensité des déplacements

e : Nombre de déplacements

s : Numérotation de la semaine ISO 8601

t : Identifiant de la station

m : mode de transport

r : Nombre de déplacements annuels de référence

$$r_{t,m} = \frac{53}{n_m} \sum_{s=1}^{53} e_{t,m,s}$$

n : Nombre de semaines de service disponible par année

Une fois le nombre de déplacements normalisé, celui-ci est désigné comme l'intensité des déplacements. L'intensité d'usage moyen par mode est présentée à la Figure 4-1 pour chacune des 53 semaines de l'année. Sur cette figure, la ligne pointillée indique l'intensité moyenne hebdomadaire.

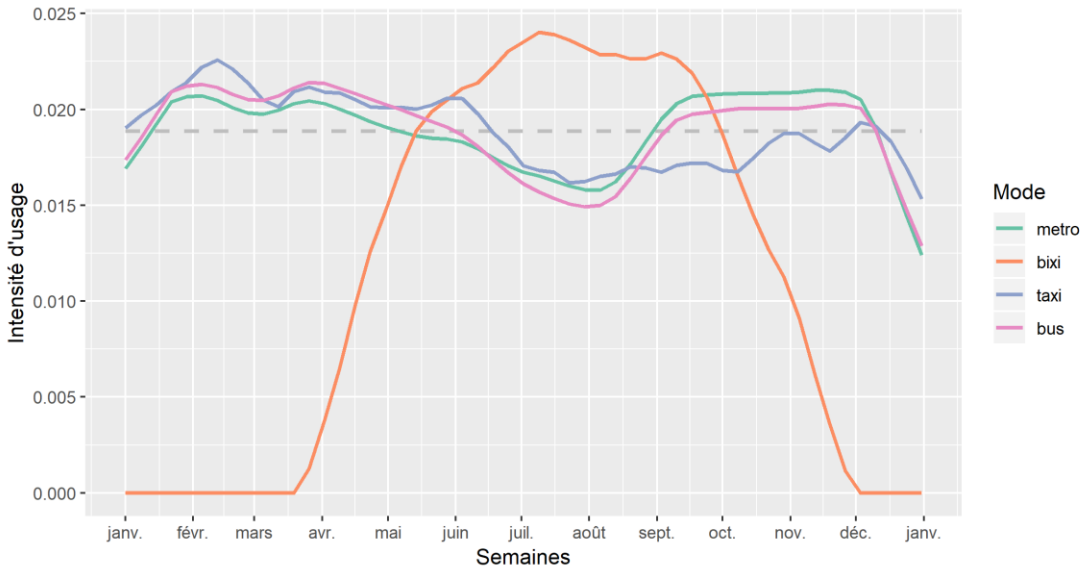

Figure 4-1 : Intensité d'usage moyenne par mode

Les courbes du métro et du bus sont similaires. Elles ont une intensité élevée en début d'année qui diminue progressivement pour atteindre un minimum durant les vacances d'été. L'intensité la plus faible pour ces modes survient durant la période des Fêtes. Pour le vélopartage, l'intensité d'usage est nulle durant la période où le service est indisponible, soit de la mi-novembre et la mi-avril. L'intensité d'usage augmente progressivement en début de saison pour atteindre un maximum en juillet, puis diminue jusqu'à la fermeture. Du côté du taxi, l'intensité est élevée en début d'année et diminue durant la période estivale, pour ne remonter que légèrement à l'automne. Durant l'été, un transfert modal vers le vélopartage pourrait avoir lieu, étant donné que son usage augmente alors que celui des autres modes diminue.

Pour continuer, des vecteurs contenant l'intensité d'usage par semaine sont créés pour chaque combinaison de station et de mode, pour un total de 272 vecteurs. Le Tableau 4-1 présente un extrait de la table résultante, où chaque ligne correspond à un vecteur.

Tableau 4-1 : Extrait de la table de données contenant les vecteurs

		Semaines				
station	mode	semaine 1	semaine 2	semaine 3	semaine 4	semaine 5
1	metro	0,01236	0,01843	0,01955	0,01996	0,02055
1	bixi	0,00000	0,00000	0,00000	0,00000	0,00000
1	taxi	0,02005	0,02074	0,02014	0,01913	0,01986
1	bus	0,01645	0,01765	0,01903	0,02045	0,02094
2	metro	0,01266	0,01885	0,01952	0,01988	0,02034
2	bixi	0,00000	0,00000	0,00000	0,00000	0,00000
2	taxi	0,01608	0,02346	0,02119	0,02039	0,01911
2	bus	0,01707	0,01806	0,01922	0,02042	0,02074

Certains vecteurs présentent des irrégularités importantes, entre autres des semaines consécutives avec des intensités très différentes. C'est le cas de plusieurs vecteurs de taxi, qui sont très sensibles aux activités ponctuelles autour d'une station.

Ce problème est illustré à gauche de la Figure 4-2, où deux exemples de vecteurs de taxi sont présentés. Malgré le fait qu'il y ait une différence importante entre les deux patrons, ils suivent tous deux une tendance annuelle similaire. Pour réduire ces irrégularités et lisser la courbe, la moyenne mobile sur 4 semaines est calculée pour chaque vecteur. Le résultat du lissage pour les deux semaines prises en exemple est présenté à droite de la Figure 4-2. Le lissage permet facilement d'identifier que les deux stations ont des patrons annuels très similaires. Une fois le lissage effectué, ces deux stations ont plus de chance d'être classées dans le même groupe.

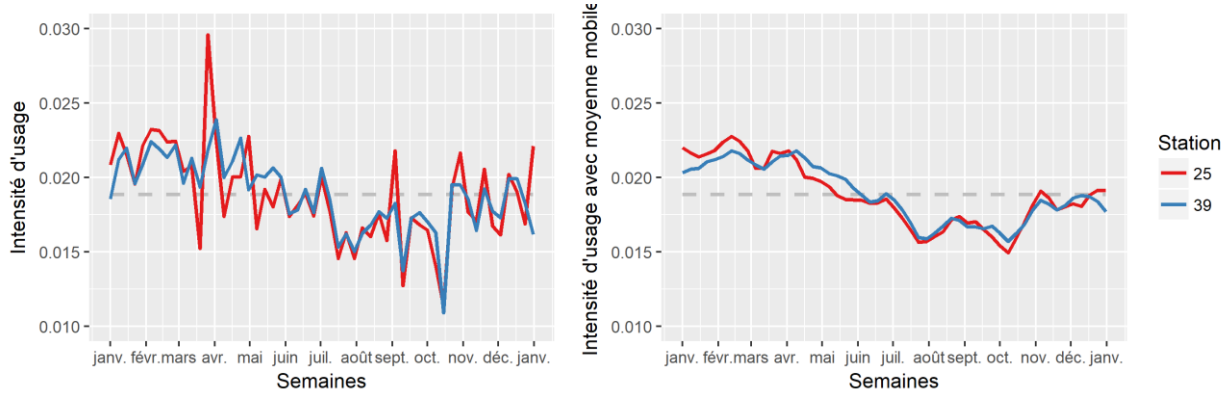

Figure 4-2 : Exemple de vecteurs pour le taxi, avant et après le lissage

Le partitionnement des vecteurs est effectué avec l'algorithme des k-moyennes. Quatre vecteurs sont exclus du partitionnement étant donné qu'ils ont des patrons qui diffèrent grandement des

autres et nécessiteraient leur groupe à eux seuls. Cela permet d'obtenir des groupes de tailles relativement semblables et évite d'obtenir des groupes avec une seule observation. Les 4 vecteurs exclus sont présentés à la Figure 4-3.

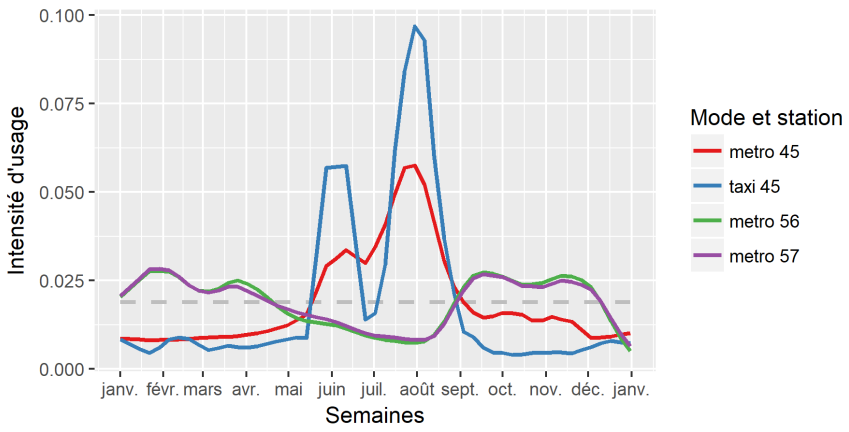

Figure 4-3 : Vecteurs exclus aberrants

La courbe rouge représente le patron d'usage typique pour le métro à la station Parc-Jean-Drapeau (station 45). Cette station enclavée est populaire pour ses activités extérieures et accueille plusieurs évènements durant l'été comme des festivals ou des événements ponctuels. Similairement, la courbe en bleu présente le patron typique pour le taxi à cette même station, dont la fluctuation d'usage est d'autant plus prononcée durant certains événements. Le maximum en juin est attribuable principalement au Grand Prix de formule 1, alors que le maximum en août correspond aux festivals de musique Osheaga et ÎleSoniq. Pour continuer, les courbes verte et mauve sont celles du métro pour les stations Université-De-Montréal et Édouard-Montpetit. Ces deux stations desservent plusieurs universités et écoles, c'est pourquoi leurs patrons annuels sont grandement liés au calendrier scolaire. En effet, l'achalandage est très faible durant l'été et les vacances des fêtes. Les semaines de relâche au printemps et à l'hiver sont également visibles, même si elles sont atténuées par le lissage.

Le partitionnement de données est effectué avec l'algorithme des k-moyennes avec un total de 251 vecteurs. Le nombre de groupes est fixé de manière à avoir suffisamment de groupes pour maximiser l'homogénéité intra-groupe. La méthode du coude (*Elbow*) est utilisée pour fixer le nombre de groupes à 5, tel que démontré sur la Figure 4-4.

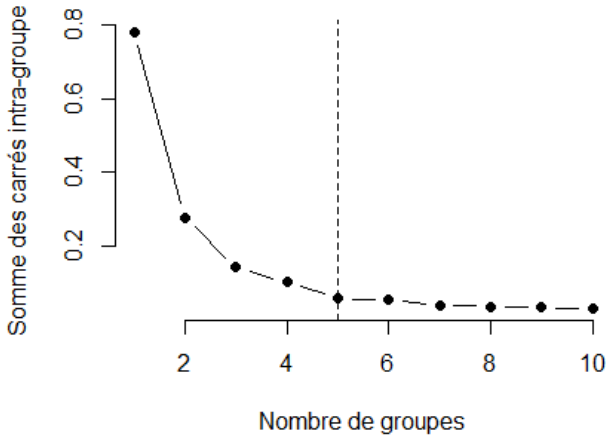

Figure 4-4 : Nombre de groupes optimal par la méthode du coude

4.1.2 Résultats

La répartition des vecteurs parmi les 5 groupes est présentée au Tableau 4-2. La taille des groupes varie de 20 à 95 observations, soit de 8,0 % à 37,8 % des observations.

Tableau 4-2 : Répartition des vecteurs parmi les groupes

Groupe	Nombre de vecteurs	Proportion
1	95	37,8 %
2	37	14,7 %
3	20	8,0 %
4	39	15,5 %
5	60	23,9 %
Total	251	100,0 %

La Figure 4-5 présente la répartition en groupes pour chaque mode. Le métro et le bus semblent partager des patrons annuels similaires étant donné qu'ils sont représentés en grande majorité par les deux mêmes groupes. Les groupes dominants pour le vélopartage lui sont propres.

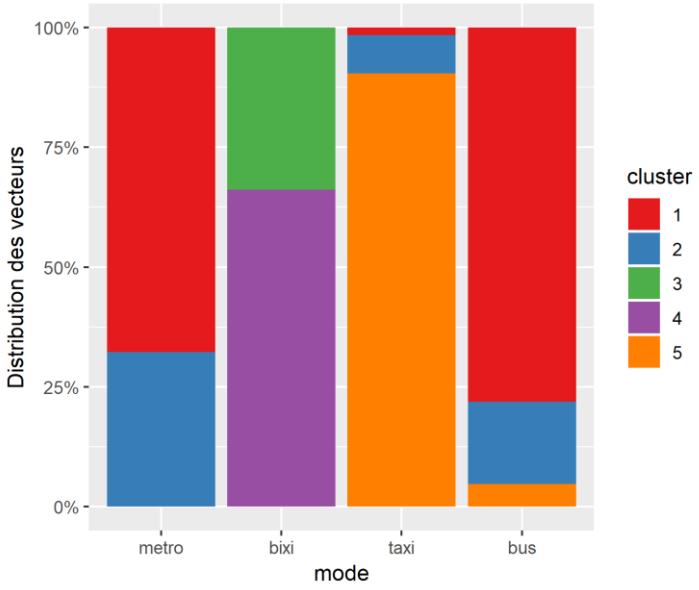

Figure 4-5 : Répartition des groupes par mode

Les centres des groupes sont tracés à la Figure 4-6. La ligne grise pointillée démarque l'intensité annuelle moyenne en période de disponibilité de service. Les lignes grises continues sont les quatre vecteurs aberrants ayant été exclus du partitionnement en raison de leurs patrons uniques (détaillés à la Figure 4-3).

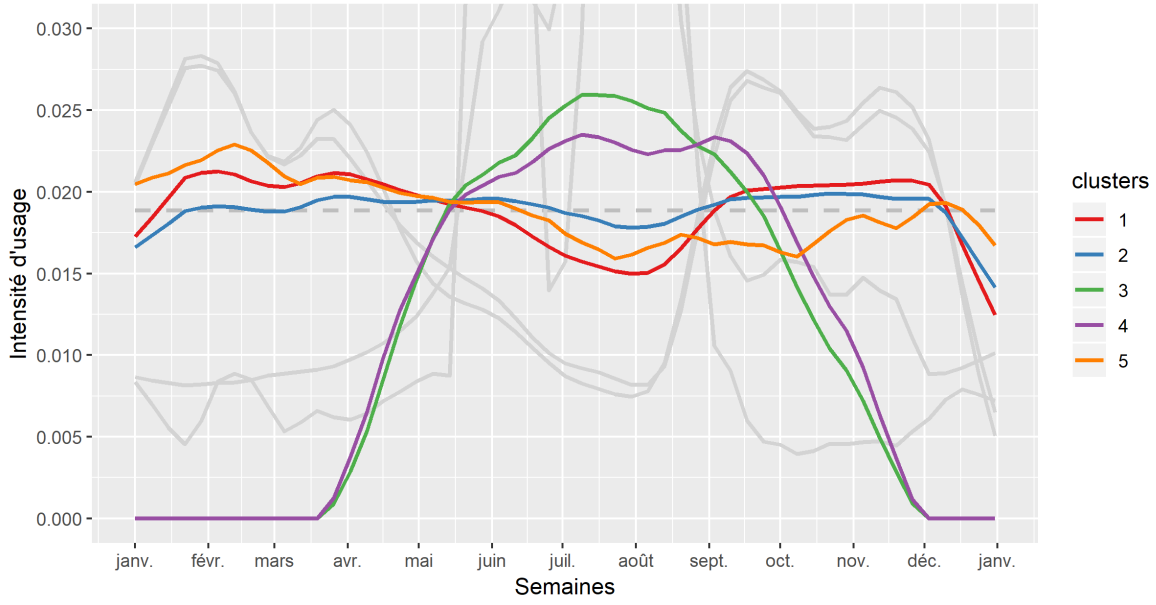

Figure 4-6 : Centres des groupes des patrons annuels

Les vecteurs du métro et du bus sont représentés dans les groupes 1 et 2. Le groupe 1 présente un usage plus faible durant la période de vacances comparé au groupe 2 dont l'usage est très constant durant l'année. Les groupes 3 et 4 ne contiennent que des vecteurs de vélopartage, ce qui est attendu étant donné qu'ils ont une intensité nulle durant l'hiver, correspondant à la période de fermeture du service. Le groupe 4 présente une intensité d'usage plus constante durant l'été que le groupe 3. Les vecteurs de taxi sont majoritairement classés dans le groupe 5, soit celui ayant une intensité décroissante durant l'année. Bref, les patrons d'usage du vélopartage et du taxi se distinguent de façon importante de ceux du transport en commun.

Il est également possible de se questionner sur l'appartenance des stations aux différents groupes. La Figure 4-7 identifie sur une carte le groupe de chaque station pour le bus. Les résultats sont très similaires pour le métro. Sur cette figure, les quatre stations sans remplissage correspondent à des stations à l'extérieur de l'île de Montréal qui ne sont pas desservies par la Société de transport de Montréal.

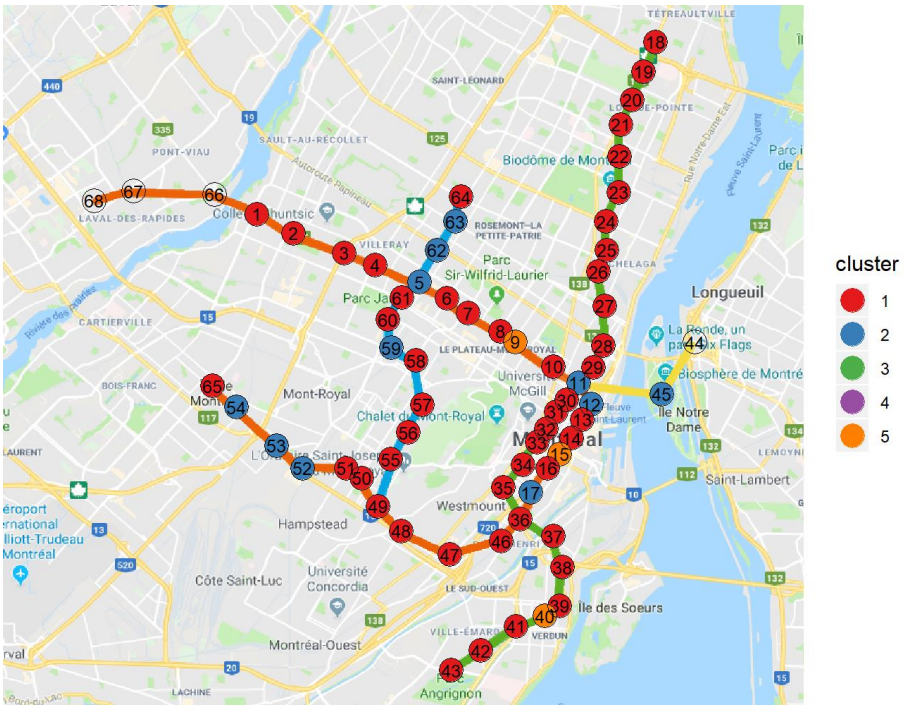

Figure 4-7 : Attribution des groupes aux stations, pour le bus

Les stations faisant partie du groupe 2 ont une intensité d'usage qui demeure élevée en période estivale. Plusieurs d'entre elles correspondent à des stations où les activités de loisirs ou de tourismes sont populaires.

Pour continuer, la Figure 4-8 présente la distribution spatiale des groupes pour le vélopartage. Les stations sans remplissage correspondent à des stations de métro sans stations de vélopartage dans le voisinage.

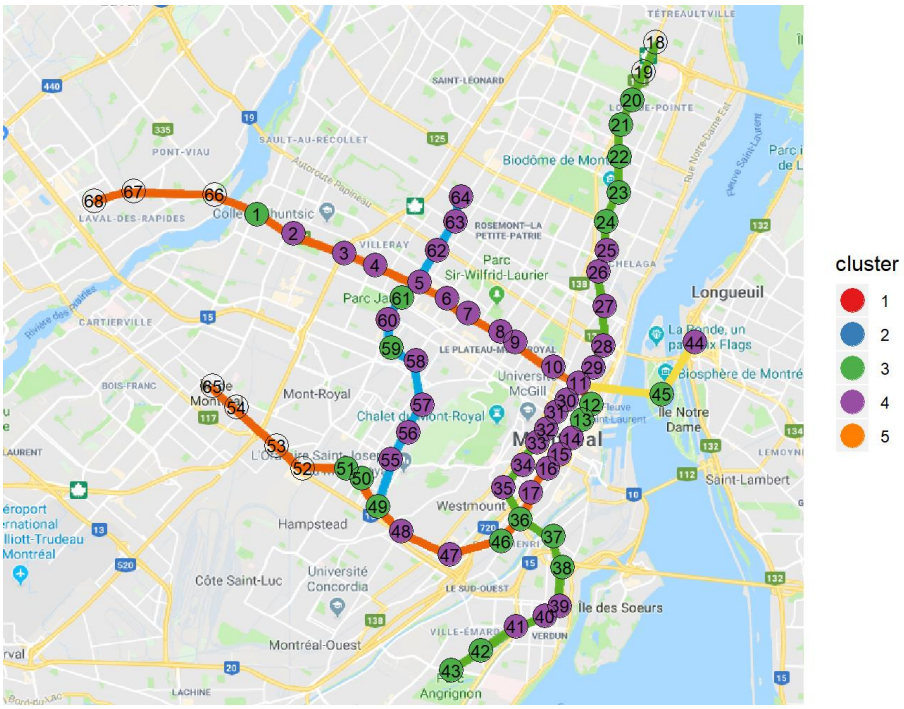

Figure 4-8 : Attribution des groupes aux stations, pour le vélopartage

Les stations au centre de l'île font majoritairement partie du groupe 4, soit celui ayant une intensité d'usage plus constante durant la période estivale. Cette zone est plus dense et les déplacements utilitaires sont probablement plus fréquents. De leur côté, plusieurs stations éloignées du centre de l'île font partie du groupe 3, soit celui avec une pointe d'usage importante au milieu de l'été, soit durant la période des vacances. Les stations Champ-de-Mars et Place-d'Armes (stations 12 et 13) font également partie du groupe 3 étant donné qu'elles bordent le Vieux-Port, un lieu touristique important durant la période estivale. Similairement, la station Jean-Drapeau (station 45) est aussi un lieu touristique populaire en période estivale.

En ce qui concerne la distribution spatiale des groupes pour le taxi, elle demeure très uniforme étant donné que 58 stations sur les 63 étudiées font partie du groupe 5. Les stations Champ-de-Mars, Place-d'Armes et Viau (stations 12, 13 et 23) font partie du groupe 2, étant donné que ces stations touristiques ont une demande importante en été et à l'automne.

4.2 Patrons hebdomadaires

Le cycle hebdomadaire a une période d'une durée de sept jours. Celui-ci varie selon le mode étudié et la période de l'année. Les trois années de données sont combinées pour obtenir une intensité d'usage moyenne pour chaque jour de la semaine et pour chaque semaine de l'année. L'unité pour cette analyse est la combinaison d'un jour de la semaine, pour un certain mode et à une certaine station. Il a été défini qu'un jour débute à 4h du matin, de façon à attribuer les déplacements effectués au début de la nuit au jour précédent.

4.2.1 Méthodologie

La méthode est similaire à celle utilisée pour les patrons annuels. Les données utilisées en entrée sont celles des déplacements horaires par mode et par station. Pour commencer, les données sont agrégées par jour, de façon à obtenir le nombre total de déplacements par jour-station-mode. Les données sont ensuite filtrées pour ne conserver que les déplacements de Taxi Diamond dont l'origine est dans la zone autorisée par le règlement sur le transport par taxi de Montréal. Quelques jours avec des données manquantes à certaines heures sont exclus.

Ensuite, chaque jour se fait attribuer un numéro de semaine entre 1 et 53 selon le système de numérotation de la norme ISO 8601. Le nombre moyen de déplacements est calculé pour chaque semaine et jour de la semaine, en combinant les trois années de données. Par exemple, les lundis de la semaine 3 sont agrégés pour les trois années de manière à définir le lundi typique de la semaine 3.

Pour continuer, le nombre de déplacements est normalisé afin de simplifier la comparaison entre les stations et les modes ayant des nombres de déplacements d'ordre de grandeur différents. La normalisation est faite par rapport au nombre hebdomadaire de déplacements pour une combinaison de station et de mode, tel qu'indiqué par la formule suivante.

$$i_{t,m,s,j} = \frac{e_{t,m,s,j}}{r_{t,m}}$$

où

i : Intensité des déplacements

e : Nombre de déplacements

t : Numéro de la station

m : Mode de transport

s : Numérotation de la semaine ISO 8601

j : Jour de la semaine

r : Nombre de déplacements hebdomadaires de référence

$$r_{t,m} = \frac{1}{n_m} \sum_{s=1}^{53} \sum_{j=1}^{7} e_{t,m,s,j}$$

n : Nombre de semaines de service disponible par année

L'intensité moyenne d'usage par mode est présentée à la Figure 4-9. Sur cette figure, la ligne pointillée indique l'intensité journalière moyenne de 0,143, identique pour chaque mode.

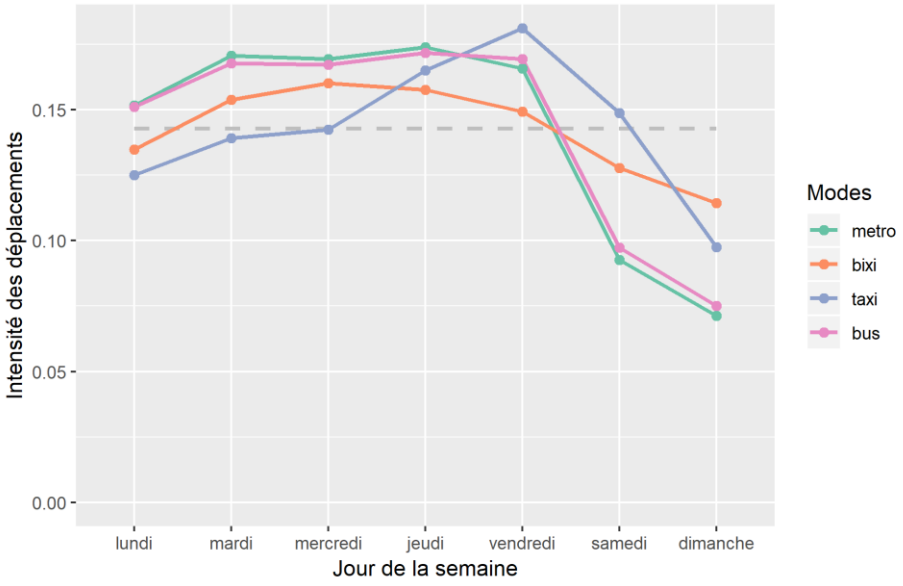

Figure 4-9 : Intensité moyenne d'usage par mode

Il est possible d'observer que les modes ont des patrons d'usage qui diffèrent. Le métro et le bus ont des patrons similaires qui présentent une intensité élevée pour les jours en semaine et une intensité plus faible la fin de semaine, particulièrement le dimanche. Du côté du vélopartage, l'intensité moyenne en semaine est à peine plus élevée que celle durant la fin de semaine. Pour le taxi, le jour où l'intensité est la plus élevée est le vendredi, suivi du jeudi puis du samedi, soit les jours présentant plusieurs activités sociales.

Pour continuer, des vecteurs contenant l'intensité d'usage pour chaque jour de semaine sont créés. Un total de 13 423 vecteurs sont créés, pour chaque combinaison de semaine, de station et de mode. Le Tableau 4-3 présente un extrait de la table résultante, où chaque ligne correspond à un vecteur.

Tableau 4-3 : Extrait de la table de données contenant les vecteurs

			Jours de la semaine						
station	mode	semaine	lundi	mardi	mercredi	jeudi	vendredi	samedi	dimanche
1	metro	1	0,0884	0,1246	0,1395	0,1162	0,1229	0,0771	0,0599
1	bixi	1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	taxi	1	0,0995	0,1574	0,1389	0,1227	0,1713	0,1342	0,0937
1	bus	1	0,0873	0,1255	0,1409	0,1163	0,1281	0,0842	0,0605
2	metro	1	0,0937	0,1301	0,1481	0,1165	0,1273	0,0733	0,0573
2	bixi	1	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	taxi	1	0,1120	0,1655	0,1193	0,1315	0,1412	0,1558	0,1010
2	bus	1	0,0973	0,1385	0,1568	0,1228	0,1377	0,0765	0,0594

Les vecteurs comprenant moins de 25 déplacements sont retirés, étant donné que leur patron peut s'avérer atypique dû au nombre très faible de déplacements. La majorité des vecteurs retirés sont des vecteurs de vélopartage dont les stations sont peu occupées en début où en fin de saison. Ces vecteurs supprimés ne représentent que 0.6 % de l'ensemble des vecteurs.

Le partitionnement de données est complété avec l'algorithme des k-moyennes avec un total de 13 343 vecteurs. Le nombre de groupes est fixé de façon à avoir suffisamment de groupes pour garantir l'homogénéité au sein d'un groupe, sans toutefois avoir plusieurs groupes similaires. La méthode du coude est utilisée, tel qu'illustré à la Figure 4-10. Le premier creux observé est pour deux groupes. Or, deux groupes semblent insuffisants pour distinguer les différents patrons hebdomadaires, c'est pourquoi le second creux est choisi, soit celui à six groupes.

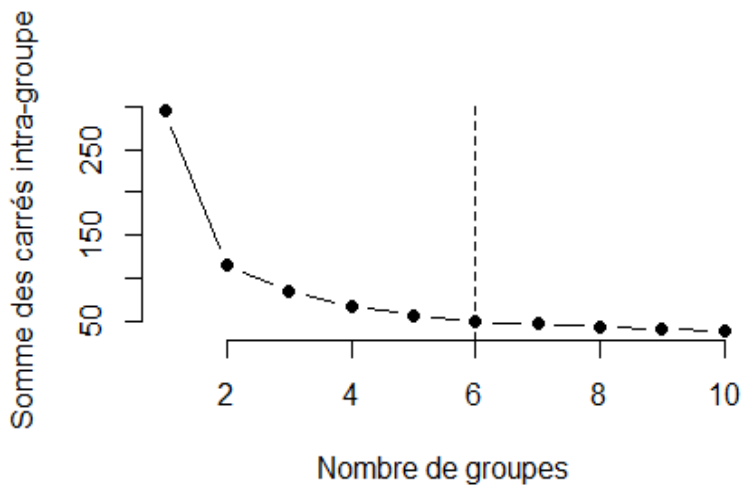

Figure 4-10 : Nombre de groupes optimal par la méthode du coude

4.2.2 Résultats

La répartition des vecteurs parmi les six groupes est présentée au Tableau 4-4.

Tableau 4-4 : Répartition des vecteurs parmi les groupes

Groupe	Nombre de vecteurs	Proportion
1	1 318	9,8 %
2	1 790	13,3 %
3	990	7,4 %
4	912	6,8 %
5	4 881	36,3 %
6	3 543	26,4 %
Total	13 434	100,0 %

La taille des groupes n'est pas homogène. En effet, le groupe le plus petit contient 6,8 % des observations alors que le groupe le plus important contient 36,3 % des observations. Le groupe le plus important sera donc partitionné de nouveau afin d'identifier s'il dissimule différents patrons distincts.

Pour continuer, la Figure 4-11 présente les centres des groupes et la Figure 4-12 présente la répartition des modes pour chaque groupe.

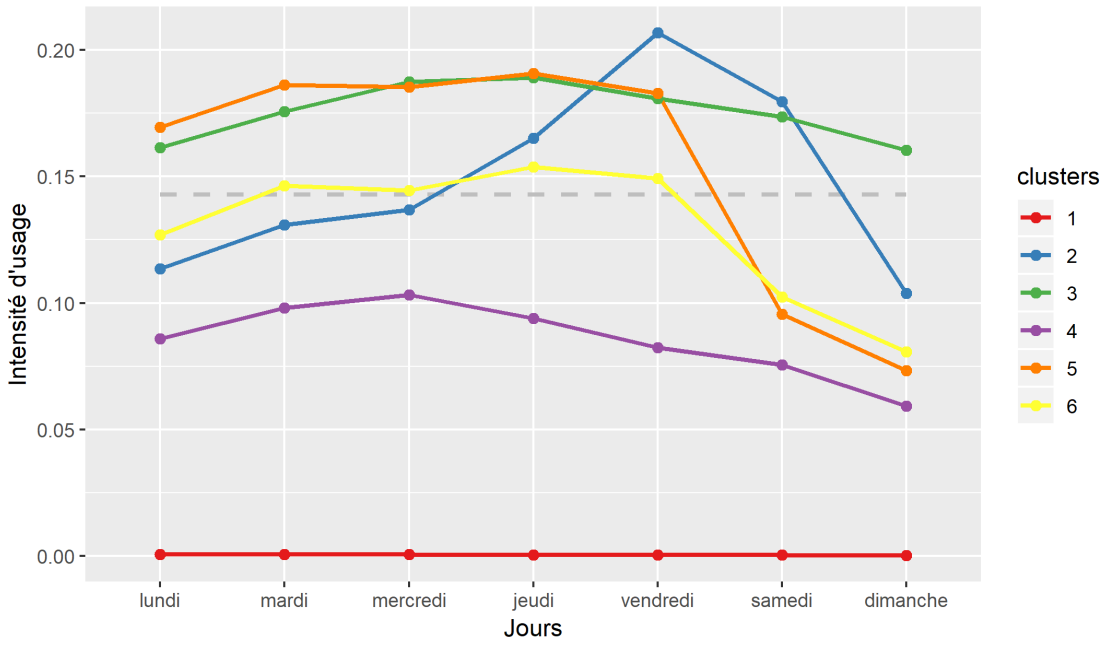

Figure 4-11 : Centres des groupes des patrons hebdomadaires

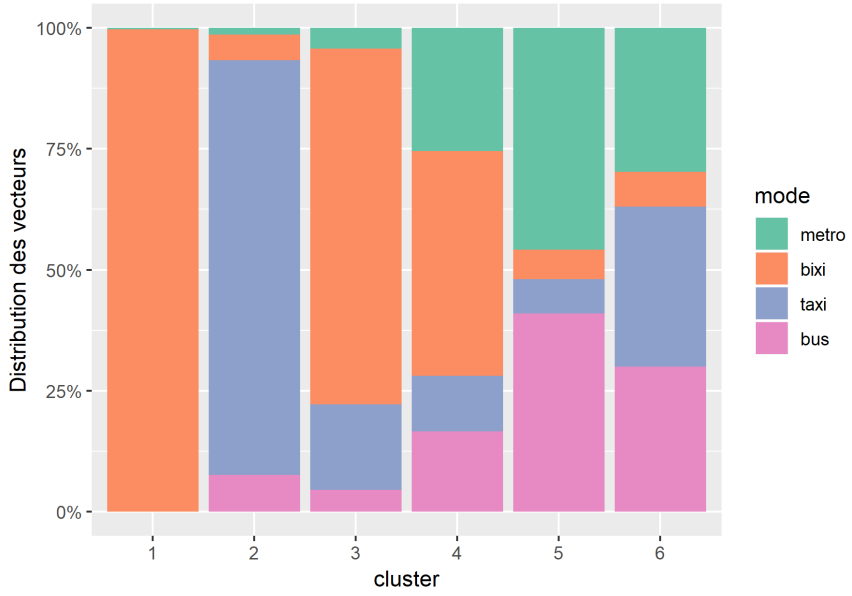

Figure 4-12 : Répartition des modes par groupe

Le groupe 1 a une intensité d'usage nulle. Il ne contient que des vecteurs de vélopartage, étant donné qu'il s'agit de semaines où le service est fermé pour la période froide. De son côté, le groupe 2 est unique avec son intensité d'usage particulièrement élevée du jeudi au samedi. Il s'agit des

jours où les activités sociales sont les plus fréquentes en soirée. Plusieurs activités sociales se terminent tard, même après l'heure de fermeture du métro, ce qui réduit les alternatives possibles de transport vers le domicile. Ainsi, le taxi semble être une alternative populaire ces jours-là. Voici pourquoi ce groupe comprend à 86% des vecteurs du mode taxi.

Le groupe 3 a une intensité d'usage très élevée et relativement constante pour chaque jour de la semaine. La majorité des vecteurs de ce groupe proviennent du vélopartage (76%). Ensuite, le groupe 4 a une intensité d'usage généralement faible et qui diminue vers la fin de la semaine. Ce groupe est composé de vecteurs du mode vélopartage à 46%, mais aussi de métro à 25% et de bus à 17%.

Le groupe 5 présente une intensité élevée pour tous les jours de la semaine, mais beaucoup plus faible la fin de semaine. Celui-ci est composé en grande majorité du mode métro (41%) et du bus (46%). Finalement, le groupe 6 a une intensité moyenne durant la semaine et celle-ci diminue la fin de semaine.

Il est possible de remarquer que même si les groupes 4, 5 et 6 ont des intensités bien différentes en semaine, leur intensité la fin de semaine est similaire.

En isolant les observations classées dans le groupe le plus important, soit le groupe 6 et en procédant au partitionnement de ce sous-ensemble de données, il est possible de conclure que le groupe est en grande partie homogène. Seul un patron se distingue des autres, soit un patron dont les jours de semaine ont une intensité encore plus élevée, soit une intensité d'environ 0,26. Ce patron est attribuable en grande majorité au métro, pour les stations Université-De-Montréal et Édouard-Montpetit (stations 56 et 57). Ces deux stations ont des fluctuations d'achalandage marquées qui concordent avec le calendrier scolaire, introduisant des intensités d'usage très importantes.

La Figure 4-13 présente la distribution temporelle des groupes pour chaque mode. Pour chaque combinaison de semaine et de mode, le groupe le plus fréquent parmi toutes les stations est retenu.

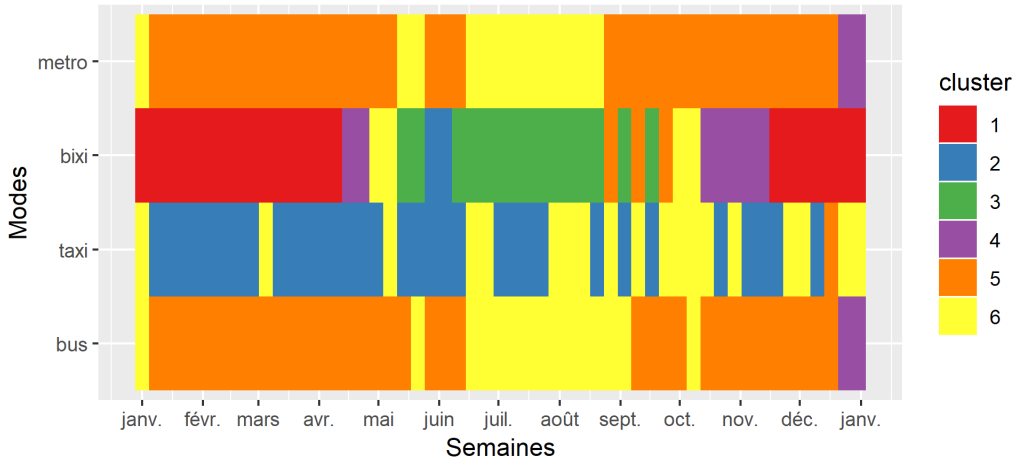

Figure 4-13 : Distribution temporelle des groupes les plus fréquents par mode

Pour le métro, la majorité des semaines font partie du groupe 5, ce qui est peu étonnant étant donné qu'il s'agit du groupe où l'intensité est élevée la semaine et faible la fin de semaine. La période estivale est mieux représentée par le groupe 6, soit une intensité 23% plus faible en semaine que le groupe 5, mais une intensité similaire pour la fin de semaine. Cette même semaine d'intensité plus faible est présente pour la première semaine type de l'année. Les deux dernières semaines de l'année font quant à elles partie du groupe 4, soit d'intensité encore plus faible en semaine, soit environ 44% plus faible en semaine que la semaine type jaune.

Du côté du vélopartage, les 15 premières semaines de l'année et les 7 dernières font partie du groupe d'intensité nulle étant donné que le service Bixi est fermé pour cette période. Le début et la fin de la saison de vélopartage sont classés dans les groupes 4 et 6, soit de faible intensité. Au milieu de l'été, le groupe le plus fréquent est le groupe 3, soit celui avec l'intensité la plus élevée. Il est ainsi possible d'observer une augmentation progressive de l'achalandage en début de saison, puis une diminution progressive en fin de saison. Au milieu de l'été, quelques semaines du groupe 2 indiquent un fort achalandage entre le jeudi et le samedi, ce qui est similaire au patron type du taxi. Cela laisse croire que le vélopartage peut se substituer au taxi en période chaude, soit la période la plus confortable pour faire du vélo. Quelques semaines de septembre sont décrites par le groupe 5, qui a une intensité très forte en semaine, mais plus faible la fin de semaine. Même si la température demeure propice au vélo durant septembre, l'utilisation chute durant la fin de semaine, possiblement dû à la reprise de la routine et à la diminution de temps consacré aux activités récréatives.

Pour continuer, les vecteurs du taxi sont principalement distribués dans le groupe 2 (62 %) et dans le groupe 6 (36 %). Les semaines durant l'hiver et le printemps font majoritairement partie du groupe 2. Les groupes 2 et 6 ont des intensités similaires le lundi, mardi, mercredi et dimanche, mais le groupe 2 présente une intensité plus élevée du jeudi au samedi. Cela laisse croire que le taxi est davantage utilisé durant l'hiver et le printemps pour des activités sociales. Par ailleurs, l'intensité du taxi diminue en juin à la même période où l'intensité du vélopartage augmente. Cela laisse croire une certaine compétitivité entre les deux modes.

La Figure 4-14 présente la distribution typique des semaines pour la station Université-de-Montréal (station 56), soit une station dont les patrons hebdomadaires se distinguent des autres stations.

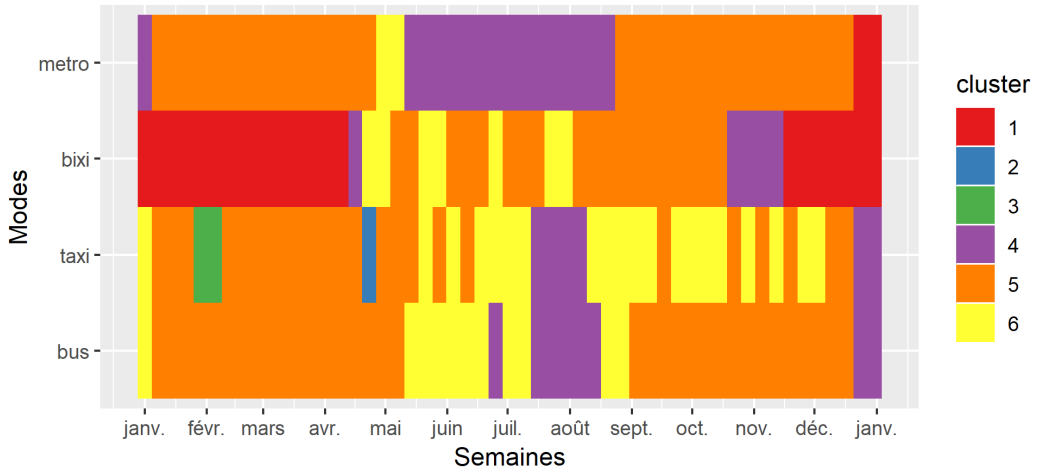

Figure 4-14 : Distribution temporelle des groupes par mode pour la station Université-de-Montréal

Cette station se distingue des autres par un usage particulièrement faible du métro et du bus durant la période de congé universitaire. En effet, la période de cours se distingue clairement par des semaines du groupe 5 et la période des vacances d'été se classe dans le groupe 4, soit celui ayant l'intensité d'usage la plus faible. Durant les vacances des fêtes, le nombre de déplacements est tellement faible pour le métro que deux semaines font partie du groupe 1, soit celui dont l'intensité est pratiquement nulle. Par ailleurs, la présence dominante du groupe 5 pour le vélopartage et le taxi indique que leur utilisation est principalement faite les jours de semaine.

Il est également possible de se questionner sur l'appartenance des stations aux différents groupes. Pour ce faire, le groupe le plus fréquent parmi les 53 semaines est affiché sur une carte, pour chaque mode. La Figure 4-15 présente les groupes les plus fréquents pour le métro.

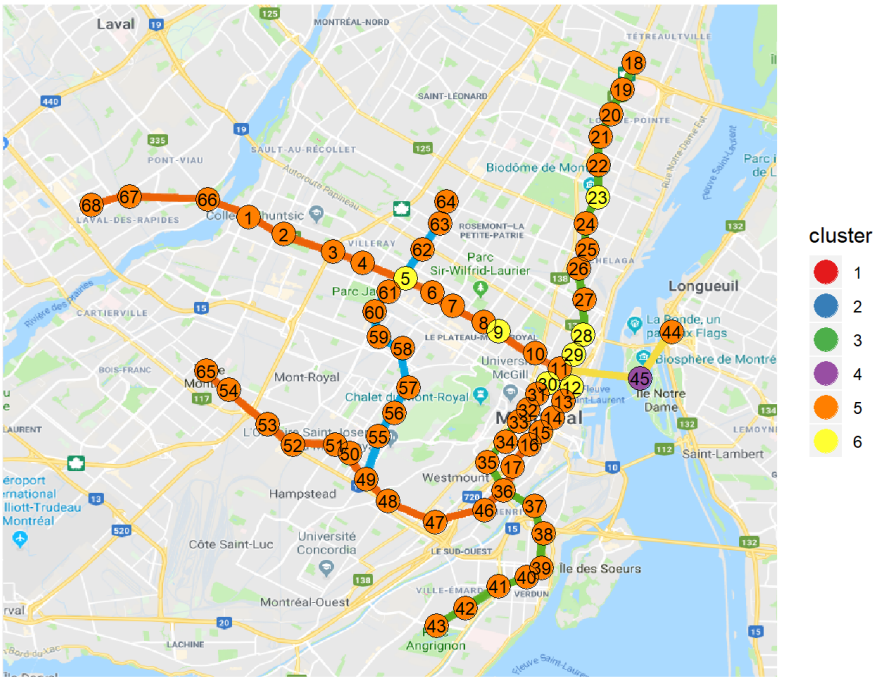

Figure 4-15 : Groupes le plus fréquent à chaque station, pour le métro

La majorité des stations appartiennent au groupe 5, soit celui avec une forte intensité en semaine et une faible intensité la fin de semaine. Sept stations appartiennent au groupe 6 et une au groupe 4, soit des groupes avec la période de semaine de plus faible intensité. Plusieurs de ces stations présentent des attraits touristiques et points d'intérêt de loisir à proximité, ce qui explique les semaines de plus faible intensité. Or, l'interprétation de cette carte est limitée au groupe le plus fréquent, signifiant qu'il n'est pas possible de savoir quelle proportion de l'année typique est représentée par ce groupe. Une solution possible est de tracer des diagrammes circulaires pour chaque station pour représenter la répartition en groupes de chaque station. La Figure 4-16 présente cette répartition des groupes pour le métro.

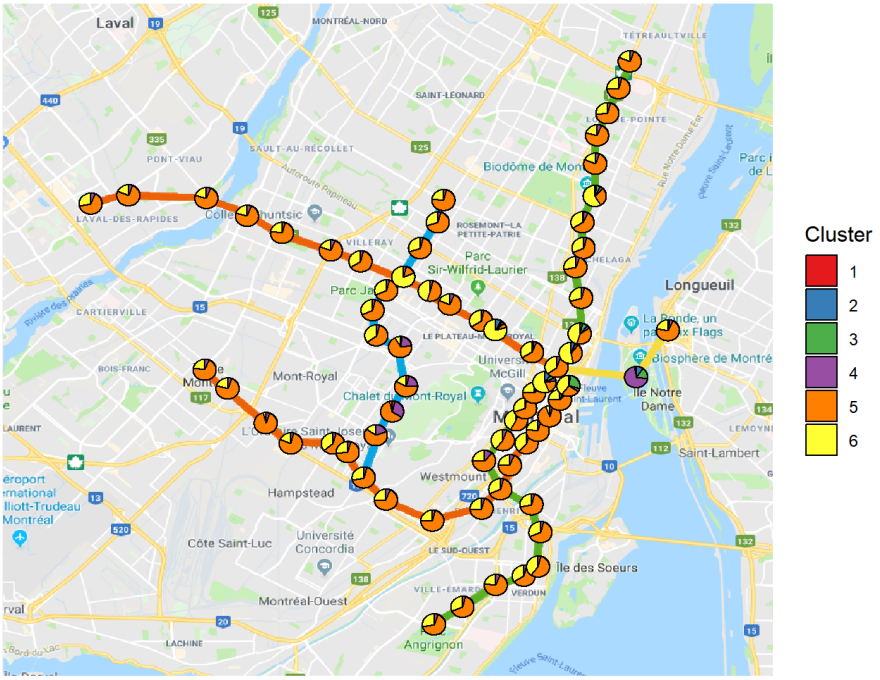

Figure 4-16 : Répartition des groupes à chaque station, pour le métro

Il est possible d'observer que pour la majorité des stations, une proportion appréciable des semaines est attribuable au groupe 6. Cette proportion est plus élevée autour de l'intersection entre la branche Est de la ligne orange et la ligne verte. Les stations Champ-de-Mars et Jean-Drapeau (stations 12 et 45) ont également des proportions non négligeables de 26% et de 15% respectivement de groupe 3, soit d'intensité élevée la fin de semaine, dû à leurs nombreuses activités de tourisme à proximité durant la période estivale. Il est également possible d'observer que les stations 55 à 58 sur la ligne bleue ont une proportion importante de vecteurs du groupes 4, soit des intensités très faibles. Cela peut être expliqué par la présence de plusieurs écoles dans cette zone, ce qui entraine une réduction de l'achalandage durant la période de vacances estivales.

Une carte semblable peut être tracée pour les déplacements en vélopartage, tel qu'illustré à la Figure 4-16. Pour cette figure, seules les semaines où le service Bixi est disponibles ont été utilisées pour tracer les diagrammes circulaires. Les stations transparentes correspondent à des stations de métro sans station Bixi dans leur voisinage.

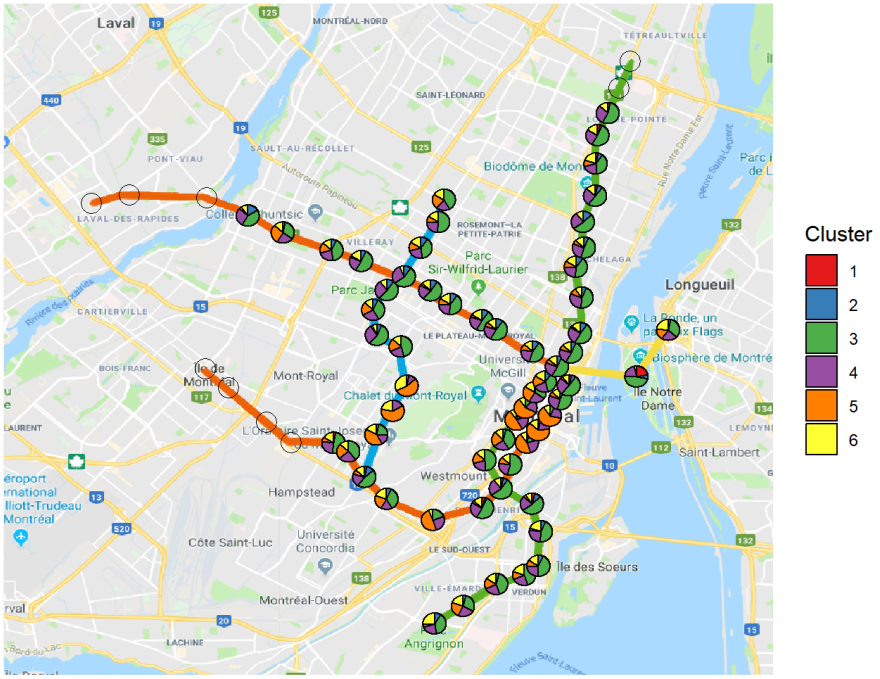

Figure 4-17 : Répartition des groupes à chaque station, pour le vélopartage

Le groupe 3 est le plus fréquent, soit celui avec une intensité d'usage élevée même la fin de semaine. Autour du centre-ville et à l'Ouest de la ligne bleu près des institutions, le groupe 5 est davantage présent, indiquant des fins de semaine de moins grande intensité. Les groupes 4 et 6 sont aussi présents pour la majorité des stations, indiquant principalement des intensités plus faibles en début et en fin de saison.

Enfin, la répartition spatiale pour les groupes du taxi est présente à la Figure 4-16. Les stations sans diagramme circulaire sont situées à l'extérieur de la zone de service de Taxi Diamond.

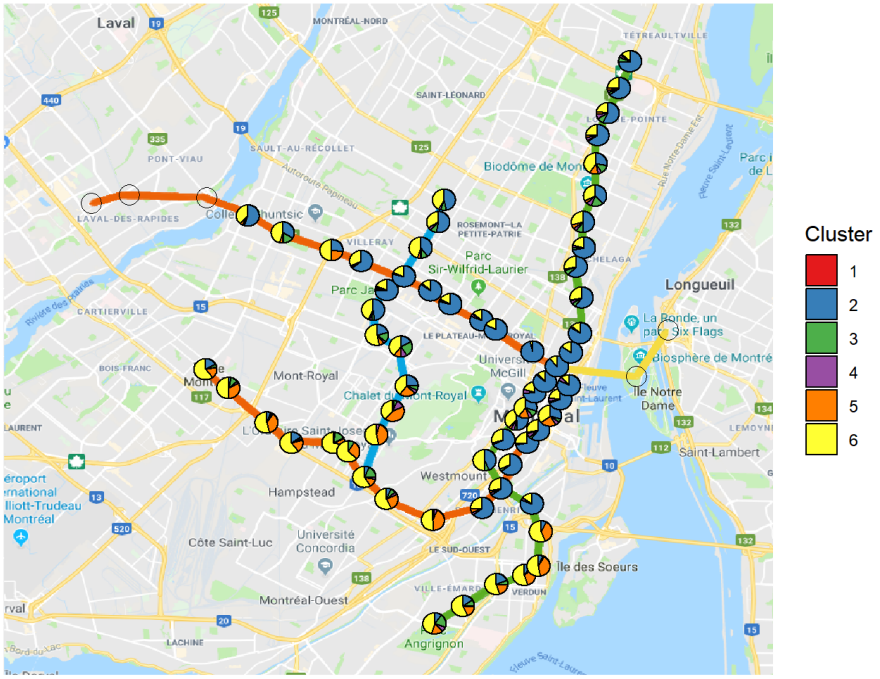

Figure 4-18 : Répartition des groupes à chaque station, pour le taxi

Pour le taxi, le centre-ville et l'Est du territoire sont représentés en majorité par le groupe 2, suivi du groupe 6. Le groupe 6 a une intensité moyenne pour tous les jours de la semaine, alors que le groupe 2 a une intensité plus élevée du jeudi au samedi. Près du croisement de la ligne verte avec la branche Est de la ligne orange, il s'agit d'un quartier très animé en soirée, ce qui peut expliquer l'usage important du taxi. Autour du centre-ville, la proportion de semaines du groupes 2 est très élevée et atteint 96 % pour la station Sherbrooke (station 10). Vers l'Ouest, le groupe le mieux représenté est le groupe 6, suivi du groupe 5.

En ce qui concerne le bus, la répartition des groupes à chaque station est très similaire à celle du métro, c'est pourquoi elle ne sera pas illustrée.

4.3 Patrons journaliers

Le cycle journalier a une période d'une durée de 24 heures. Les patrons types d'utilisation des systèmes de transport varient selon le jour de la semaine et la période de l'année. Généralement, la distribution des déplacements en semaine présente deux périodes de pointe distinctes, alors que pour la fin de semaine les déplacements sont répartis plus uniformément durant la journée. Or, l'amplitude de ces périodes de pointe varie selon le mode et la localisation.

Pour cette analyse, l'unité est la combinaison d'une heure, pour un certain mode et à une certaine station. Les trois années de données sont combinées pour obtenir une année type de 365 jours. Il a été défini qu'un jour débute à 4h du matin, de façon à attribuer les déplacements effectués au début de la nuit au jour précédent. Les heures identifiées 25h, 26h et 27h représentent donc les heures 1h, 2h et 3h respectivement, pour le jour précédent.

4.3.1 Méthodologie

La méthode est similaire à celles présentées précédemment pour les patrons annuels et hebdomadaires. Les données utilisées sont celles des déplacements horaires par mode et par station. Les quelques jours comprenant des données manquantes à certaines heures sont exclus.

Chaque jour se fait attribuer un numéro de semaine entre 1 et 53 selon le système de numérotation de la norme ISO 8601. Les données des trois années sont agrégées en calculant la moyenne des patrons pour les jours correspondants. Par exemple, les lundis de la semaine 1, pour chacune des 3 années, sont agrégés.

Par la suite, le nombre de déplacements est normalisé afin de simplifier la comparaison entre les stations et les modes ayant des nombres de déplacements d'ordres de grandeur différents. La normalisation est faite par rapport au nombre de déplacements journaliers pour une combinaison de station et de mode, tel qu'indiqué par la formule suivante.

$$i_{t,m,s,j,h} = \frac{e_{t,m,s,j,h}}{r_{t,m}}$$

où

i : Intensité des déplacements

e : Nombre de déplacements

t : Numéro de la station

m : Mode de transport

s : Numérotation de la semaine ISO 8601

j : Jour de la semaine

h : Heure

r : Nombre de déplacements hebdomadaire de référence

$$r_{t,m} = \frac{1}{n_m}\frac{1}{7}\sum_{s=1}^{53}\sum_{j=1}^{7}\sum_{h=4}^{27} e_{t,m,s,j,h}$$

n : Nombre de semaines de service disponible par année

La Figure 4-19 présente l'intensité d'usage moyenne par mode pour les jours de semaine. La Figure 4-20 présente aussi l'intensité d'usage moyenne par mode, mais pour les jours de fin de semaine. Pour ces figures, seule la période d'ouverture du service Bixi est considérée. Il est important de rappeler que l'horodatage d'un déplacement est arrondi à l'heure inférieure, c'est-à-dire qu'un déplacement qui est effectué entre 8h et 8h59 est attribué à 8h.

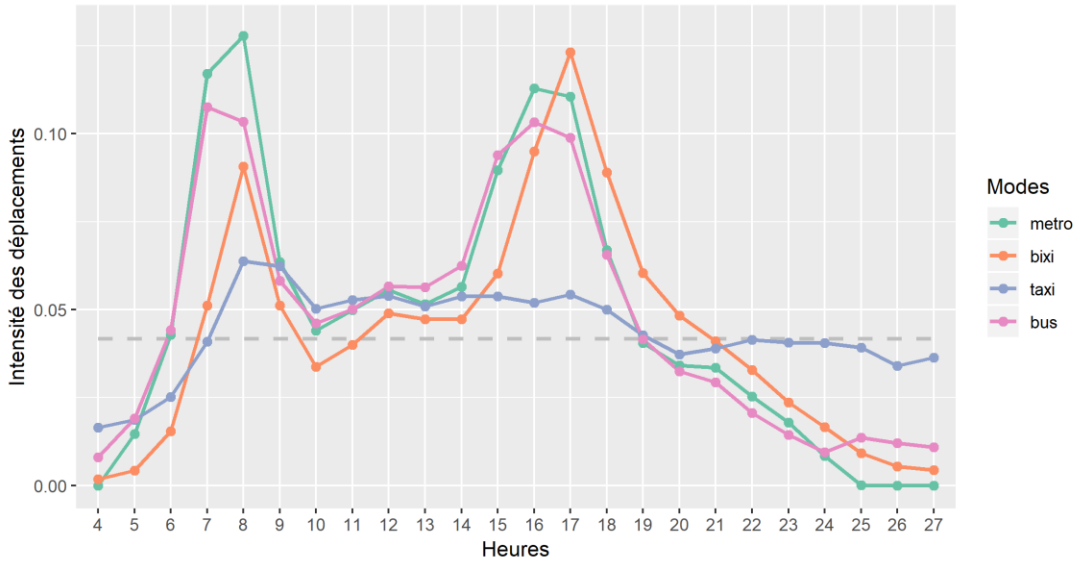

Figure 4-19 : Intensité moyenne d'usage par mode, pour les jours de semaine

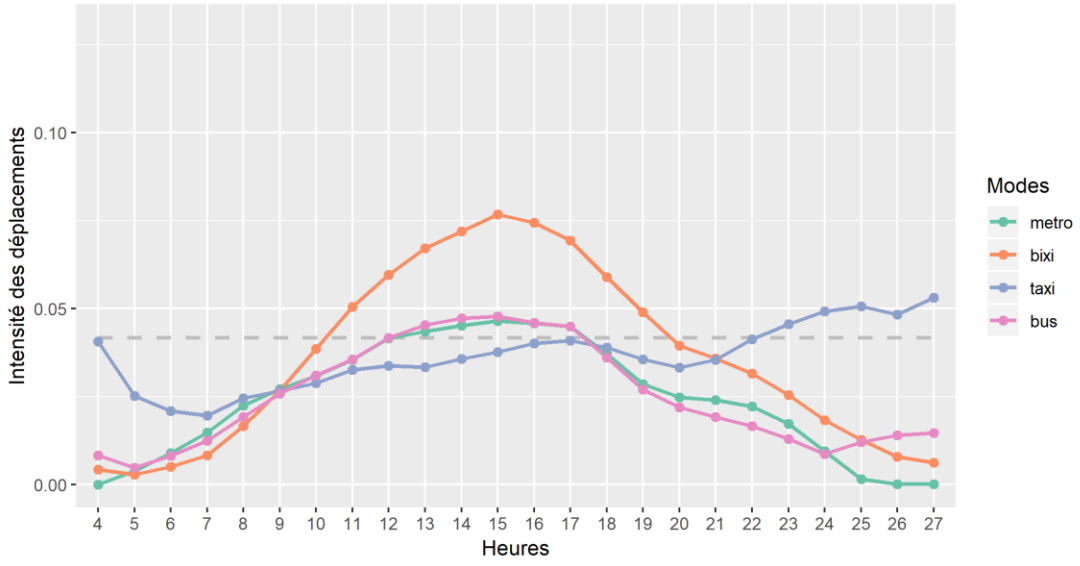

Figure 4-20 : Intensité moyenne d'usage par mode, pour les jours de fin de semaine

Durant les jours de semaine, deux périodes de pointe sont visibles pour chacun des modes. La période de pointe du matin débute entre 6h et 7h et se poursuit jusqu'à 9h inclusivement. Pour la pointe du soir, l'heure de pointe du métro et du bus est entre 15h et 18h, celle du vélopartage semble se poursuivre jusqu'à 19h, alors que celle du taxi est pratiquement inexistante. La pointe maximale pour le métro, le taxi et le bus est celle du matin, alors que pour le vélopartage il s'agit de celle du soir. Un intervalle entre les données plus court qu'une heure permettrait une meilleure précision quant à la pointe maximale et à l'heure à laquelle elle survient. Entre les deux périodes de pointe,

l'intensité des déplacements est plutôt constante. Après l'heure de pointe du soir, l'intensité diminue progressivement pour le métro, le vélopartage et le bus. Le métro est fermé entre 2h et 5h du matin, c'est pourquoi les heures 26, 27 et 4 ont des intensités nulles. Du côté du taxi, son usage demeure constant durant la soirée. L'usage du taxi est le plus faible entre 4h et 6h, mais demeure plus élevée que pour les autres modes.

Pour les jours de fin de semaine, les patrons moyens diffèrent de ceux en semaine. Pour le métro et le vélopartage, l'intensité des déplacements augmente progressivement en matinée pour atteindre un maximum autour de 15h, puis diminuer progressivement pour le reste de la journée. Pour le vélopartage, l'intensité est plus élevée que celle du métro pour la majorité de la journée. Le bus a un patron très similaire à celui du métro, mais avec une légère augmentation d'intensité durant les heures de fermeture du métro. Du côté du taxi, le patron se distingue par une croissance du nombre de déplacements au cours de la journée. La période la plus occupée est entre 24h et 27h, avec une intensité maximale atteinte autour de 27h, soit 3h du matin, expliquée par l'absence de service du métro et de la majorité des lignes d'autobus durant la nuit, réduisant les alternatives de déplacement possibles. Pour cette raison, l'intensité du taxi est relativement élevée tout au long de la journée.

Pour continuer, des vecteurs sont créés pour stocker les intensités d'usage pour les 24 heures de la journée. Un total de 92 616 vecteurs sont créés, pour chaque combinaison de semaine, de jour, de station et de mode. Le Tableau 4-5 présente un extrait de la table résultante, où chaque ligne correspond à un vecteur.

Tableau 4-5 : Extrait de la table de données contenant les vecteurs

				Heure				
station	mode	semaine	jour	4h	5h	6h	7h	8h
1	bixi	1	jeudi	0,00000	0,00000	0,00000	0,00000	0,00000
1	metro	1	jeudi	0,00000	0,00673	0,01199	0,00764	0,01010
1	taxi	1	jeudi	0,17174	0,12490	0,01561	0,01561	0,03123
1	bus	1	jeudi	0,0039	0,0044	0,0065	0,0068	0,0077
2	bixi	1	jeudi	0,00000	0,00000	0,00000	0,00000	0,00000
2	metro	1	jeudi	0,00000	0,00573	0,00794	0,00694	0,00734
2	taxi	1	jeudi	0,06579	0,08223	0,06579	0,06579	0,01645
2	bus	1	jeudi	0,0012	0,0031	0,0051	0,0056	0,0058

Le partitionnement de données est complété avec l'algorithme des k-moyennes, afin de regrouper les vecteurs similaires. Le nombre de groupes est fixé de façon à avoir suffisamment de groupes pour garantir l'homogénéité au sein d'un groupe, sans toutefois avoir plusieurs groupes similaires. La méthode du coude aide à fixer le nombre de groupes, avec un léger creux à six groupes tel que démontré à la Figure 4-21.

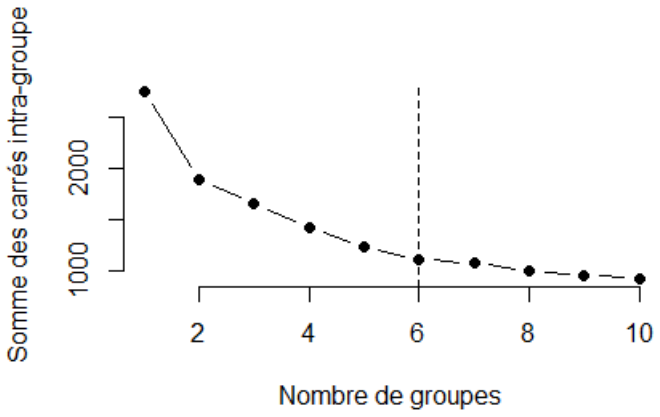

Figure 4-21 : Nombre de groupes optimal par la méthode du coude

4.3.2 Résultats

La répartition des vecteurs parmi les six groupes est présentée au Tableau 4-6.

Tableau 4-6 : Répartition des vecteurs parmi les groupes

Groupe	Nombre de vecteurs	Proportion
1	10 483	11,3 %
2	28 463	30,7 %
3	6 974	7,5 %
4	20 965	22,6 %
5	20 746	22,4 %
6	4 985	5,4 %
Total	92 616	100,0 %

La taille des groupes n'est pas homogène, puisque les groupes contiennent entre 5,4 % et 30,7 % des observations. Il pourrait être pertinent d'isoler le groupe le plus important et de partitionner à nouveau ce sous-ensemble de données afin d'identifier si ce groupe dissimule plusieurs patrons distincts.

La Figure 4-22 présente les centres des groupes et la Figure 4-23 présente la répartition des groupes pour chaque mode.

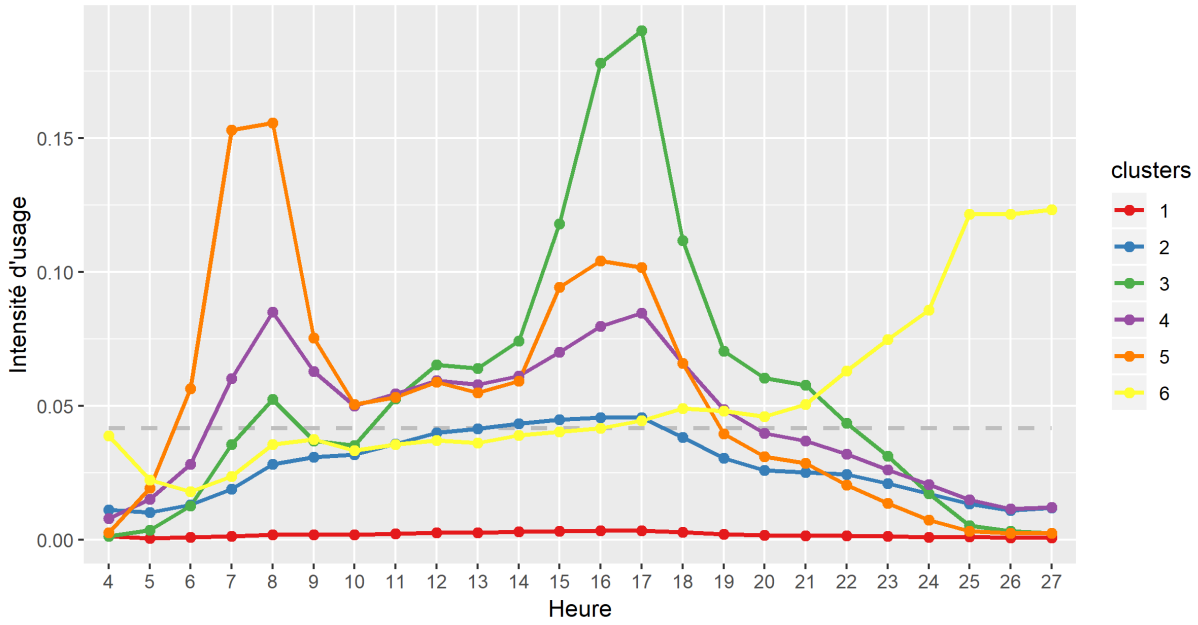

Figure 4-22 : Centres des groupes des patrons hebdomadaires

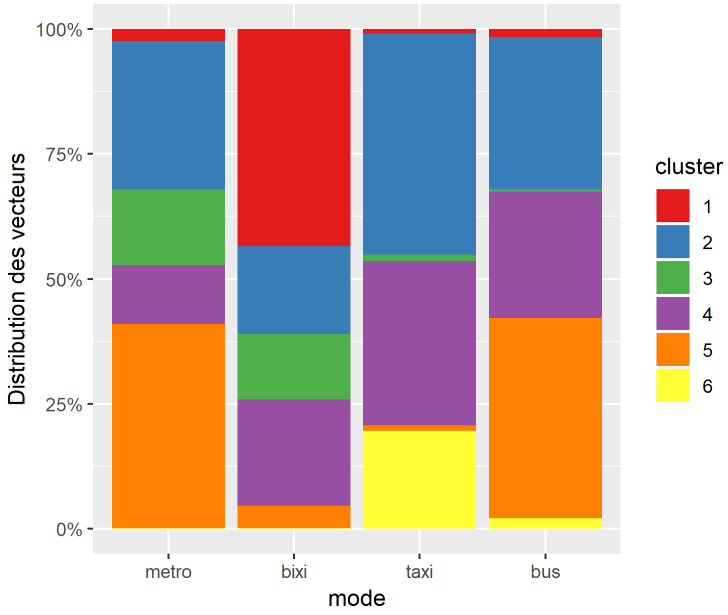

Figure 4-23 : Répartition des groupes par mode

Le groupe 1 a une intensité d'achalandage quasi-nulle. Ce groupe contient majoritairement des vecteurs de vélopartage, dû à la fermeture du service Bixi 5 mois par an.

Le groupe 2 présente une intensité relativement faible toute au long de la journée, atteignant un maximum vers 17h puis diminuant doucement en soirée. Ce groupe contient un nombre important de vecteurs de chacun des quatre modes. Il s'agit du groupe le plus important pour le taxi, contenant près de 44% des vecteurs de taxi.

Du côté du groupe 3, celui-ci présente une forte intensité en pointe du soir, en opposition au groupe 5. La pointe du soir est la plus importante entre 15h et 18h environ, atteignant une intensité maximale de 0.19 vers 17h. En soirée, l'intensité diminue progressivement pour atteindre une intensité quasi-nulle vers 2h. La pointe du matin est peu importante pour ce groupe. Le groupe 3 est principalement présent pour le métro et le Bixi.

Le groupe 4 montre deux périodes de pointe distinctes, mais de moins grande intensité que pour les groupes 3 et 5. Les heures où l'intensité est maximale pour les deux pointes correspondent à celles des groupes 3 et 5, soit à 8h pour la pointe du matin et à 17h pour la pointe du soir.

Pour continuer, le groupe 5 présente une période de pointe matinale marquée entre 6h et 9h environ, en atteignant une intensité de 0,16 à la 8ème heure. La période de pointe du soir est également visible, mais beaucoup moins importante avec une intensité maximale de 0,11. L'intensité diminue durant la soirée pour atteindre une intensité quasi-nulle à partir de 1h (25h). Environ 49% des vecteurs classés dans ce groupe sont du mode métro et 45% sont du bus.

Le groupe 6 présente une intensité croissante durant la journée, qui augmente rapidement en soirée pour atteindre un maximum vers 3h (27h). Le patron journalier de ce groupe ressemble au patron moyen du taxi les jours de fin de semaine illustré à la Figure 4-20. Ce groupe est presque uniquement composé de vecteurs du mode taxi.

La Figure 4-24 présente la distribution temporelle des groupes pour chaque mode. Étant donné que la combinaison d'un jour et d'un mode inclut plusieurs stations, seul le groupe le plus fréquent parmi toutes les stations est indiqué sur cette figure.

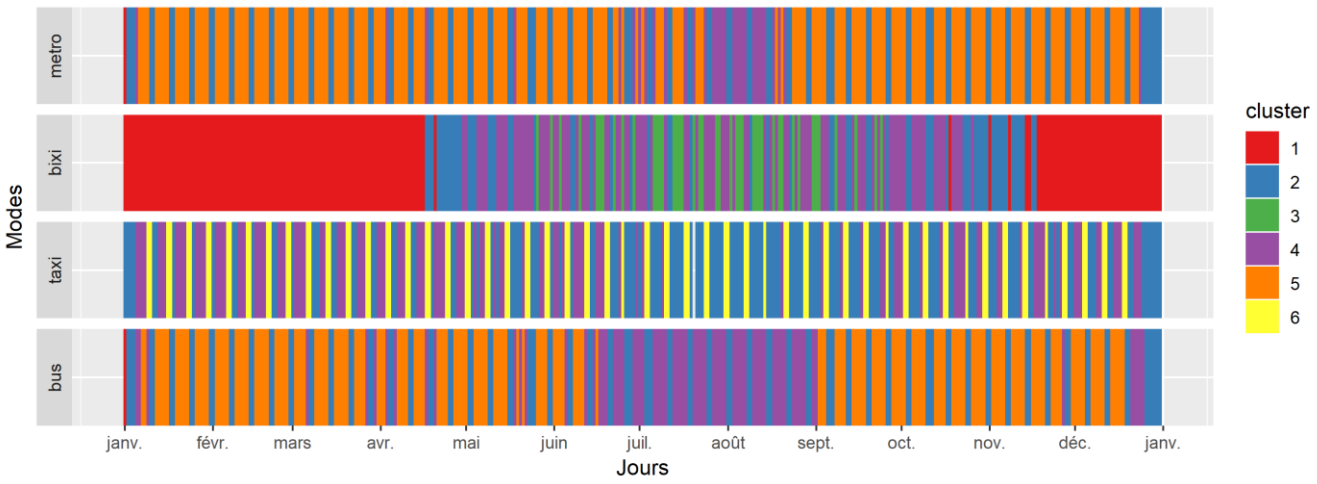

Figure 4-24 : Distribution temporelle des groupes les plus fréquents par mode

Pour le métro et le bus, il est clairement possible de distinguer les jours de semaine des jours de fin de semaine, se répétant tous les 7 jours. Pour ces deux modes, les jours de semaine font majoritairement partie du groupe 5, soit du groupe avec une pointe du matin élevée. Cependant, il est important de mentionner que le groupe 3 est presque aussi fréquent que le groupe 5, mais que cette figure ne présente que le groupe le plus fréquent. Le groupe 5 est plus fréquent étant donné qu'il y a une majorité de stations de métro qui ont une pointe matinale plus forte que celle du soir. Les jours de semaine d'été font partie du groupe 4, soit celui avec deux périodes de pointe plus faibles. Les jours de fin de semaine font partie du groupe 2. Les journées de congé font partie du groupe 2 ou 4, soit avec des intensités plus faibles. Pour les jours fériés, certains peuvent changer de semaine selon l'année, comme c'est le cas du vendredi et du lundi de Pâques qui se déplacent parmi les semaines 13, 14 et 16 pour les trois années observées. Dans ce cas, l'effet du congé férié est plus difficilement observé.

Du côté du vélopartage, la distinction entre les jours de semaine et de fin de semaine n'est pas aussi claire que pour le métro et le bus. Les jours sont répartis parmi les groupes 2, 3 et 4 en période de disponibilité de service. Les fins de semaines font généralement partie du groupe 2 alors que les jours de semaine font majoritairement partie du groupe 4. Or, l'usage du Bixi semble dépendre d'autres facteurs que le jour de la semaine, comme probablement les conditions météorologiques ainsi que divers évènements. Les jours classés dans le groupe 3, soit avec une pointe du soir élevée, sont plus fréquents au milieu de l'été, probablement dû à la période chaude et propice au Bixi.

Pour le taxi, le groupe le plus fréquent est le groupe 2, soit celui qui présente une faible intensité durant la journée. Le groupe 4 est aussi fréquent en début d'année et en fin d'année, où l'intensité d'usage est plus importante. Les jours de la fin de la semaine font souvent partie du groupe 6, soit celui avec une intensité croissante en soirée, ce qui correspond aux jours où les personnes font des activités de loisir. La Figure 4-25 confirme cette affirmation, en démontrant que le groupe 6 est majoritairement présent du jeudi au samedi pour le taxi, soit les soirées populaires pour sortir.

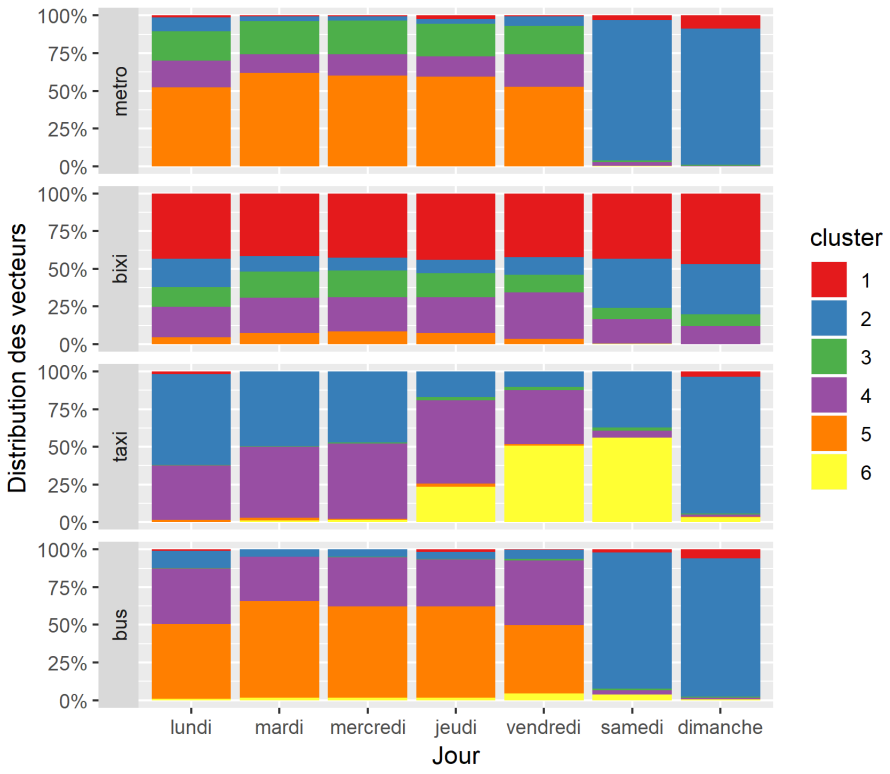

Figure 4-25 : Distribution des vecteurs par jour de la semaine et par mode

Sur cette figure, deux distributions se distinguent pour le métro, soit une distribution pour la semaine et une autre pour la fin de semaine. La distribution en semaine est principalement composée de vecteurs des groupes 3, 4 et 5. Cette distribution est très similaire pour les différents jours en semaine. Le groupe 5 est le plus fréquent avec environ 52% des vecteurs. Quelques vecteurs sont classés dans le groupe 2 le lundi et le vendredi, ce qui est probablement associé à des jours fériés qui présentent une plus faible intensité d'usage. Les vecteurs des jours de fin de semaine sont aussi classés dans le groupe 2.

Pour le vélopartage, la distribution des vecteurs parmi les groupes est similaire pour tous les jours de la semaine. Les jours de fin de semaine ont une proportion plus élevée de vecteurs du groupe 2, soit d'intensité plus faible et constante.

De leur côté, les vecteurs du taxi font majoritairement partie des groupes 2 et 4. Tel que mentionné plus tôt, le groupe 6 est présent du jeudi au samedi, atteignant 59% des vecteurs pour le samedi.

Pour continuer, la Figure 4-26 présente la distribution des groupes en semaine pour le métro. Seuls les jours de semaine sont considérés, étant donné que les jours de fin de semaine diffèrent. Pour les jours de fin de semaine, les vecteurs font presque tous partie du groupe 2 tel que démontré à la Figure 4-25, alors leur répartition spatiale est peu pertinente.

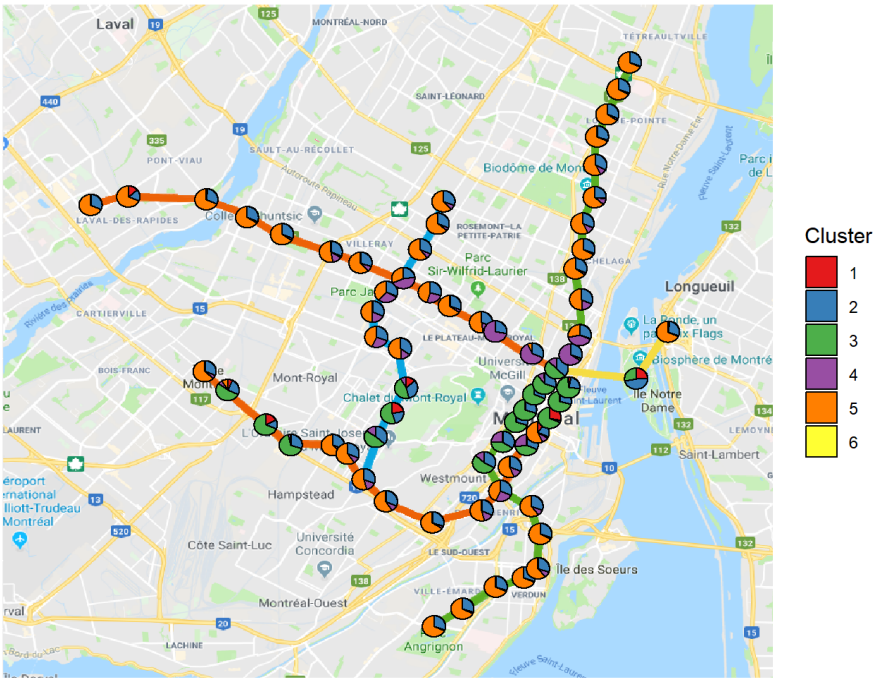

Figure 4-26 : Répartition des groupes à chaque station pour les jours de semaine, pour le métro

La majorité des stations sont décrites par un seul groupe qui domine sur les autres. La zone autour du centre-ville est dominée par le groupe 3 ayant une période de pointe du soir importante. Ce résultat est attendu étant donné qu'il s'agit de déplacements qui sont initiés au centre-ville en fin de journée pour le retour au domicile. Les stations 52 à 57 font également partie du groupe 3, dû à de multiples lieux d'activités à proximité tel que des institutions scolaires et centres d'achat. Par ailleurs, la majorité des stations plus éloignées du centre-ville font partie du groupe 5, soit avec une forte pointe du matin. Autour de la branche Est de la ligne orange, quelques stations du groupe 4

sont présentes, soit des stations ayant un nombre similaire de déplacements effectués en pointe du matin et du soir. Finalement, la station Jean-Drapeau (station 45) est unique, avec de nombreux jours classés dans les groupes 1 et 2, soit d'intensité faible ou quasi-nulle.

Pour continuer, la Figure 4-27 présente la répartition des groupes pour le vélopartage en période de disponibilité du service. Les jours de semaine et de fin de semaine sont combinés, étant donné que la répartition spatiale des groupes est similaire pour ces deux périodes.

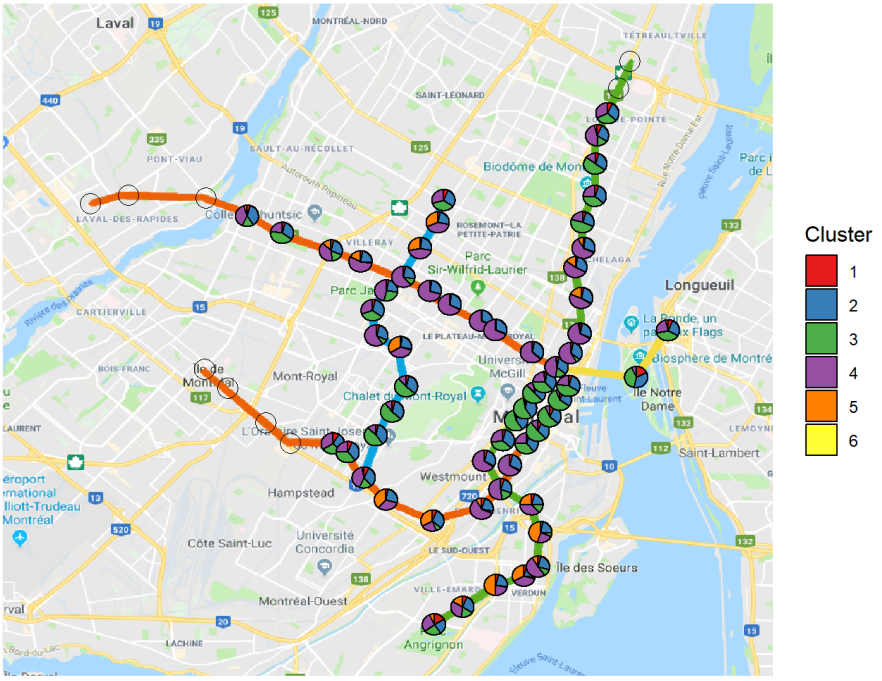

Figure 4-27 : Répartition des groupes à chaque station, pour le vélopartage

Les stations autour du centre-ville et à l'Ouest de la ligne bleue font majoritairement partie du groupe 3, comme c'est le cas pour le métro (voir Figure 4-26). Les autres stations font principalement partie du groupe 4. La proportion non-négligeable de jours du groupe 2 est expliquée par les périodes moins achalandées au début et à la fin de la saison de vélopartage. Il y a très peu de stations dans le groupe 5 qui a une pointe matinale importante. Cela pourrait être attribué à l'utilisation du Bixi généralement plus importante en après-midi qu'en matinée, tel qu'observé sur la Figure 4-19 et la Figure 4-20.

Similairement, la Figure 4-28 présente la répartition des groupes par station pour le taxi.

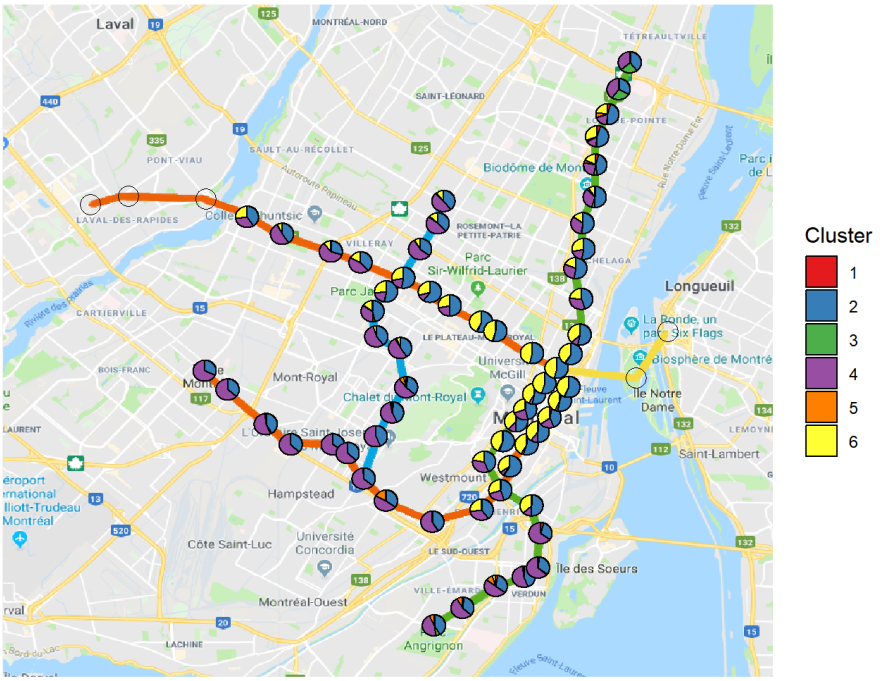

Figure 4-28 : Répartition des groupes à chaque station, pour le taxi

Le taxi se distingue des autres modes en présentant plusieurs observations classées dans le groupe 6, principalement autour du centre-ville et sur la branche Est de la ligne orange. Les autres stations sont dominées par le groupe 2 ou le groupe 4 qui présentent des intensités d'usage modérées le jour et faibles en soirée.

La Figure 4-29 présente la répartition des groupes pour le bus. Les stations sans diagramme circulaire sont des stations non desservies par l'opérateur de bus STM.

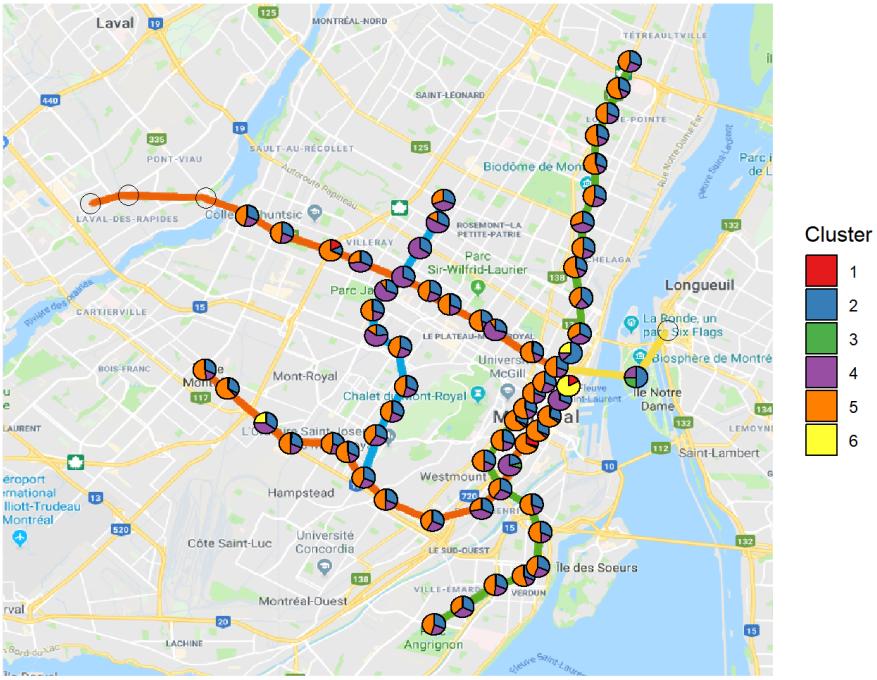

Figure 4-29 : Répartition des groupes à chaque station, pour le bus

Les différentes stations se différencient peu comparément aux autres modes, ce qui est probablement dû à une limitation de la méthode. En effet, les arrêts d'embarquement sont inconnus pour le bus, alors l'ensemble des embarquements sur une même ligne sont comptabilisés ce qui réduit la variabilité inter-stations.

CHAPITRE 5 TYPOLOGIE DE STATIONS

L'objectif de cette section est de développer une typologie de stations de métro en regroupant les stations similaires. Tout d'abord, une typologie des stations est créée sur la base de multiples caractéristiques du voisinage des stations de métro. Ensuite, une classification des stations est faite sur la base de la demande journalière des différents modes autour des stations. Finalement, ces deux classifications sont mises en relation, afin de déterminer s'il est possible d'expliquer la demande journalière des stations à partir des caractéristiques de leur voisinage.

Une telle typologie peut être utile par exemple afin de mieux comprendre la demande autour des différentes stations et peut servir à l'évaluation de scénarios. Par exemple, elle pourrait permettre d'estimer l'impact de l'ajout d'une nouvelle station de métro sur la demande des divers modes.

5.1 Typologie de stations selon les caractéristiques

Cette première partie s'intéresse à la création d'une typologie de stations de métro basée sur les caractéristiques du voisinage des stations à l'aide d'un regroupement hiérarchique. Les différents indicateurs de voisinage des stations de métro sont décrits au Tableau 3-7 et des statistiques descriptives sont présentées au Tableau 3-4. Les indicateurs sont tout d'abord normalisés avant d'effectuer le partitionnement de données.

Tout d'abord, les variables de différents ordres de grandeur sont normalisées afin qu'elles soient considérées également lors du partitionnement de données. Chaque variable X est centrée et réduite en utilisant la moyenne μ et l'écart-type σ, tel que décrit par l'équation suivante :

$$x = \frac{X - \mu}{\sigma}$$

Ensuite, une segmentation des stations de métro est effectuée par une méthode de regroupement hiérarchique. Le regroupement hiérarchique est un ensemble de méthodes de partitionnement de données qui consiste à créer une série imbriquée de segmentations des individus. Les individus sont comparés en utilisant une mesure de dissimilarité. Pour cette analyse, une méthode ascendante hiérarchique est utilisée, soit une méthode qui fusionne des groupes de façon itérative (James, Hastie, James et Witten, 2013). La dissimilarité est initialement évaluée par une distance euclidienne, puis la dissimilarité interclasse est évaluée par la méthode de Ward (Murtagh et Legendre, 2014). Le partitionnement par la méthode des k-moyennes est également testé, mais la

méthode hiérarchique est retenue dû à ses résultats plus intuitifs et à la possibilité de visualiser la classification par un dendrogramme. Le nombre de groupes est identifié de façon intuitive, en s'assurant que les stations près du centre-ville soient regroupées et que les stations comportant plusieurs écoles soient aussi regroupées. Les groupes ont été ordonnés par distance moyenne en métro à partir du centre-ville, c'est-à-dire que le groupe 1 est celui le plus près du centre-ville et le groupe 5 est le plus éloigné.

Le dendrogramme résultant de la classification hiérarchique est présenté à la Figure 5-1. Sur cette figure, chaque couleur identifie un groupe différent et l'axe horizontal indique la dissimilarité entre les groupes. La ligne pointillée indique la dissimilarité critique utilisée pour former les groupes. Il peut être plus facile de visualiser ces groupes sur une carte, c'est pourquoi ceux-ci ont été tracés à la Figure 5-2.

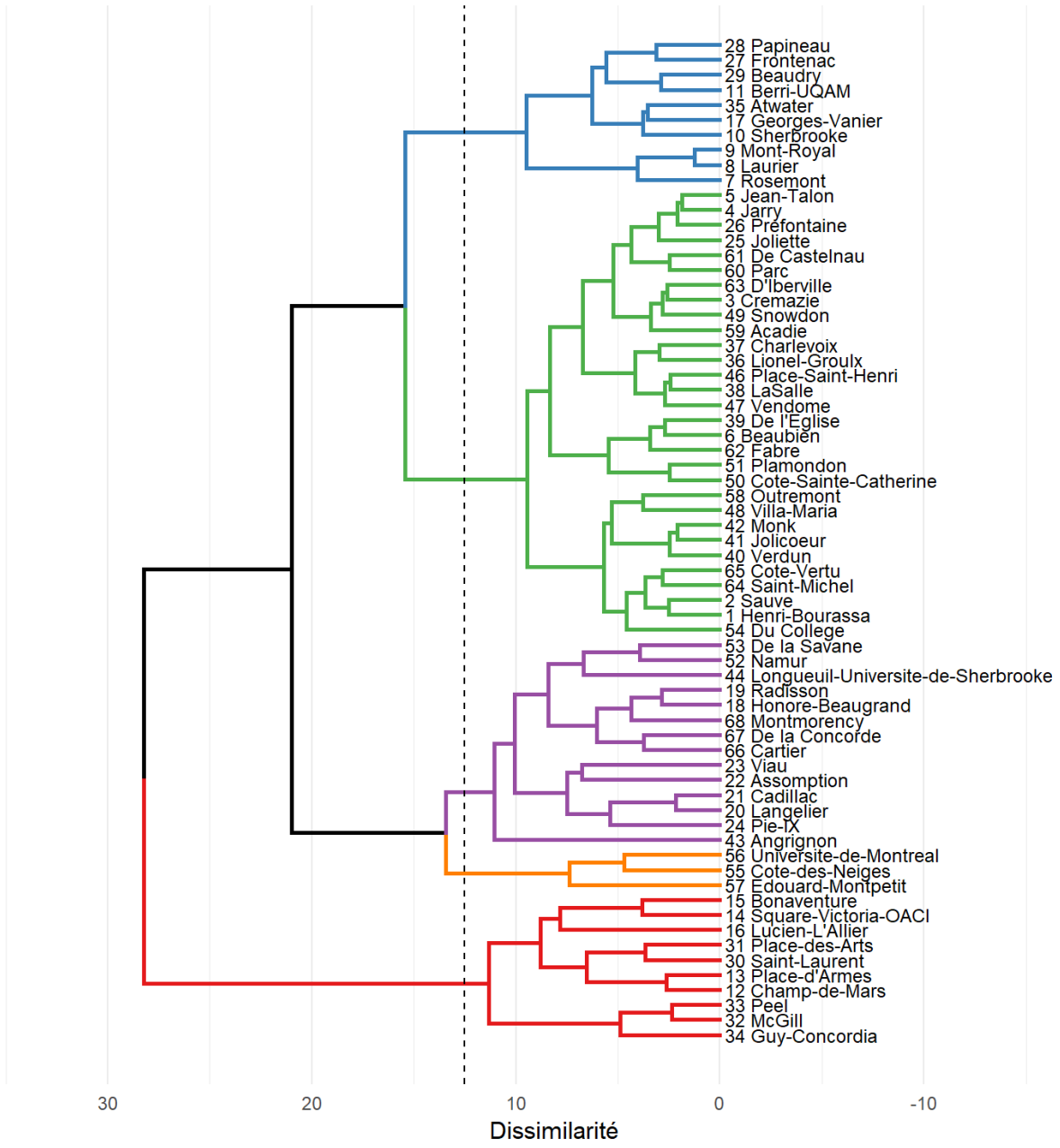

Figure 5-1 : Dendrogramme résultant de la classification selon les caractéristiques

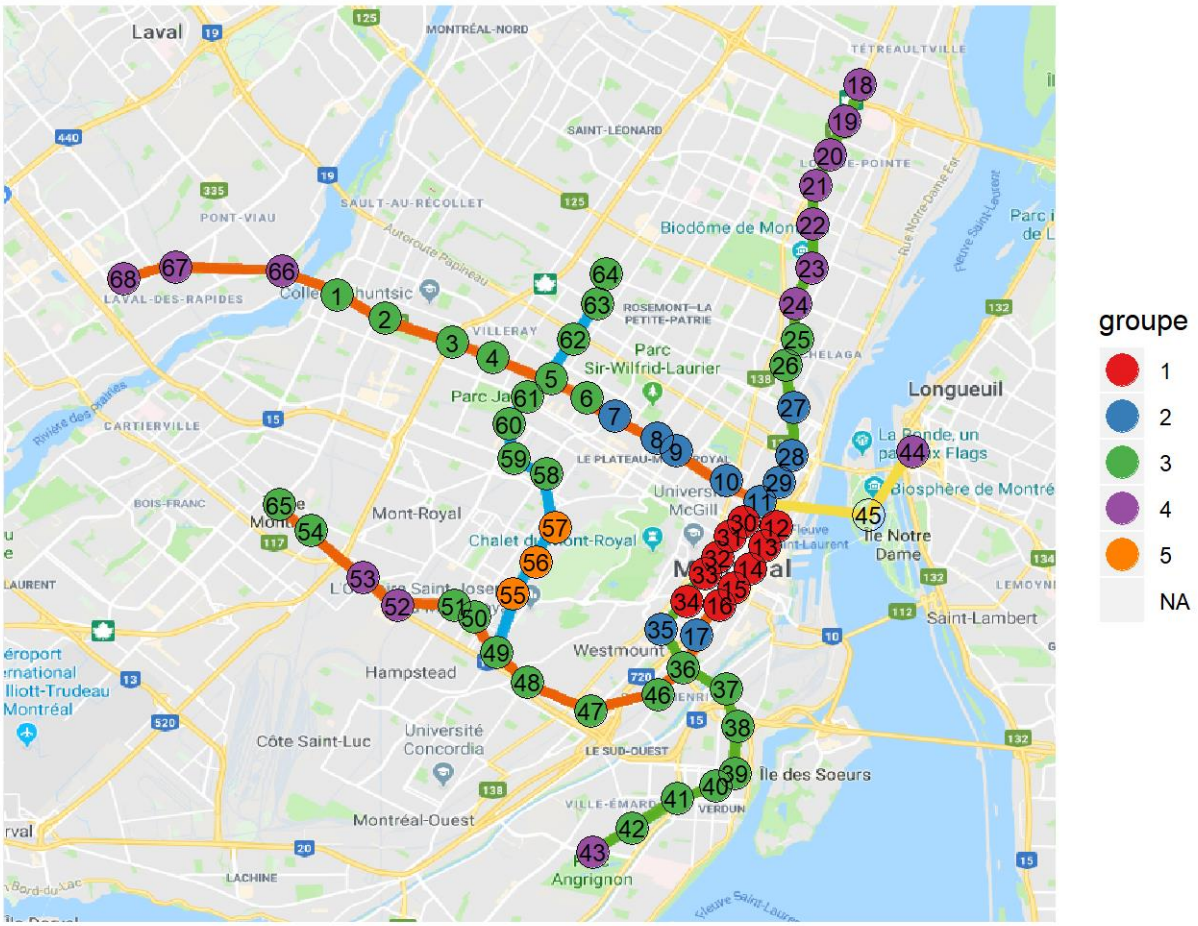

Figure 5-2: Répartition des stations par groupe

Il est possible d'observer des regroupements spatiaux de stations similaires. La distance au centre-ville semble influencer le groupe d'appartenance de la station. Tout d'abord, le groupe 1 est caractérisé par des stations autour desquelles la densité d'emploi est très importante. Le groupe 1 contient uniquement des stations du centre-ville. Ce groupe compte en moyenne 117 000 déplacements pour le travail par jour ouvrable dans le voisinage de chaque station, tel qu'indiqué au Tableau 5-1. De plus, ce groupe se distingue par une superficie occupée par des bureaux aussi élevée que 15,5% et un nombre élevé d'activités près des stations.

Le groupe 2 est caractérisé par des stations en zone majoritairement résidentielle, mais avec une quantité tout de même élevée d'emplois. Ce groupe contient des stations situées très près du centre-ville où se destinent environ 21 900 travailleurs par jour ouvrable, soit l'équivalent du nombre d'emplois. Il s'agit du groupe qui présente la plus forte population dans le voisinage des stations.

Dans ce groupe, la proportion de résidents ayant 65 ans et plus est la plus faible, soit de 11,3 % seulement.

Le groupe 3 est caractérisé par des stations en zone résidentielle comportant peu d'emplois. Ces stations sont plus éloignées du centre-ville que le groupe 2, mais tout de même situées au centre de l'île. La distance au centre-ville en métro à partir des stations est en moyenne de 13,8 minutes. Il s'agit du groupe qui présente la superficie résidentielle la plus élevée, soit de 41,5%. Le nombre de résidents moyen est légèrement plus faible que celui du groupe 2, soit de 19 474 résidents. Le groupe 3 est celui qui comprend le plus de stations, soit 30 stations.

Le groupe 4 est caractérisé par des zones résidentielles éloignées du centre-ville. Les stations situées à l'extérieur de l'Île de Montréal y sont incluses. Il s'agit du groupe comportant la population la plus faible avec aussi peu que 8 299 résidents par voisinage de station et le groupe comportant le nombre d'emplois le plus faible. L'âge moyen est le plus élevé, soit de 43,2 ans, et la proportion de résidents de 65 ans et plus est également la plus élevée. Le taux d'accès à une automobile est le plus élevé, soit de 0,61 voitures par personne résidant dans le ménage. Cela pourrait être expliqué par une faible marchabilité et une offre de vélopartage plus limitée. Par ailleurs, le groupe 4 comporte la proportion la plus élevée de parcs dans le voisinage des stations, couvrant en moyenne 6,87% de la superficie. Il peut paraître étrange à première vue que plusieurs stations sur la branche Est de la ligne verte soient incluses dans ce groupe malgré leur proximité au centre-ville, mais certaines caractéristiques de ces stations sont très similaires aux autres stations du groupe 4, comme par exemple une faible densité de population, un âge moyen élevé, une proportion d'aire résidentielle plus faible que le groupe 3 et une proportion élevée du territoire occupée par des parcs.

Finalement, le groupe 5 est caractérisé par une forte quantité d'institutions scolaires. Il s'agit du groupe le plus petit, qui ne contient que 3 stations de métro. La proportion de la surface occupée par des institutions est aussi élevée que 46,5% et toutes les stations comprennent au moins une université dans leur voisinage. Il s'agit du groupe qui présente la proportion la plus élevée de résidents de 14 ans ou moins, soit de 16,3 %. Il s'agit également de la station avec le revenu médian le plus faible, ce qui peut être expliqué par la quantité élevée d'étudiants. La proportion de la superficie occupée par des bureaux ou des commerces est très faible.

Tableau 5-1 : Centres des groupes

	1	2	3	4	5
population	11889	21906	19474	8299	15524
homme	54,8	54,3	48,76	48,6	44,95
age_moy	38,27	38,21	37,98	43,15	38,49
age0_14	6,02	8,4	14,68	12,48	16,27
age15_24	22,33	13,66	12,31	9,34	16,34
age25_39	32,89	37,62	28,85	24,88	25,30
age40_64	25,12	29,05	31,44	32,74	25,01
age65_plus	13,64	11,28	12,72	20,56	17,07
revenu_med	27010	26914	27465	26447	19538
prop_inf20k	41,79	38,31	36,52	29,64	26,64
depl_travail	116784	21873	6778	5149	16189
marchabilite	35,62	24,56	14,05	8,84	11,30
acces_auto	0,41	0,39	0,51	0,61	0,54
nb_ancrages_bixi	4897	2738	688	172	868
aire_residentielle	17,68	36,05	41,47	26,94	29,88
aire_commerciale	10,36	4,81	4,49	12,70	1,05
aire_bureau	15,45	4,20	1,32	1,57	0,92
aire_institutionnelle	12,38	8,16	8,18	12,63	46,47
aire_parc	2,63	4,79	4,87	6,87	4,35
aire_utilite_publique	39,42	37,63	33,7	27,67	16,84
nb_ecole_publique	1,60	3,80	4,43	1,93	1,67
cegep	0,2	0,3	0,07	0,14	0,0
universite	0,3	0,3	0,0	0,07	1,0
hopital	0,4	0,6	0,13	0,21	0,33
temps_cv_metro	1,6	6,7	13,77	17,93	18
nb_activites	118,3	4,1	2,03	14,43	0,0

5.2 Classification selon les patrons journaliers

Cette deuxième partie s'intéresse à la classification des stations de métro en se basant sur les patrons journaliers de la demande des différents modes dans le voisinage des stations et ce, pour les jours ouvrables. Le regroupement hiérarchique est utilisé afin de partitionner les données.

Deux classifications sont considérées, soit une première classification tenant compte du volume de la demande horaire ainsi que de sa répartition journalière, alors que la seconde tient uniquement compte de la répartition journalière de la demande.

5.2.1 Classification selon le volume et la répartition journalière

Pour cette première classification, le volume de la demande et sa répartition journalière sont considérés. Pour effectuer la classification, les étapes suivantes sont complétées : le filtrage des données transactionnelles, le calcul des indicateurs d'intensité, la création des vecteurs et le partitionnement de données.

Premièrement, les données transactionnelles horaires des différents modes sont filtrées et combinées. Il est décidé d'exclure les données du bus de cette analyse, étant donné qu'elles comportent très peu de variabilité entre les stations dû au fait que les lignes de bus entières sont considérées. En effet, l'ensemble des lignes de bus passant par une station de métro sont considérées, sans tenir compte de la direction, ce qui rend difficile la comparaison entre les stations. Par ailleurs, la station Parc Jean-Drapeau est exclue de cette classification étant donné que celle-ci comporte des patrons d'usage plutôt atypiques qui ont peu de similarité avec les autres stations. Les données sont également filtrées pour ne conserver que les jours ouvrables.

Deuxièmement, l'indicateur d'intensité est calculé pour chaque combinaison de station, de mode et d'heure. Cet indicateur est basé sur le nombre de départs normalisé selon le mode de transport, afin d'éliminer les différences entre les modes dues à l'ordre de grandeur d'usage. L'équation suivante est utilisée pour calculer cet indicateur.

$$i_{t,m,h} = \frac{d_{t,m,h}}{r_m}$$

où

$$r_m = \frac{1}{n}\sum_{t=1}^{n}\sum_{h=1}^{24} d_{t,m,h}$$

i : Intensité des déplacements

d : Nombre de déplacements

r : Nombre de départs de référence (moyenne journalière des stations)

t : Numéro de la station

m : Mode de transport

h : Heure

n : Nombre de stations

Troisièmement, les vecteurs sont formés pour chaque combinaison de mode et de station, en regroupant les indicateurs calculés précédemment. Les stations de métro sans stations de vélopartage à proximité ont un usage nul pour ce mode. Pour éviter que l'algorithme de classification n'isole les stations sans vélopartage dans un même groupe, l'usage moyen du vélopartage des autres stations est imputé. Ainsi, la classification des stations sans vélopartage repose davantage sur l'usage du métro et du taxi.

Les vecteurs formés sont présentés au Tableau 5-2.

Tableau 5-2 : Extrait des vecteurs utilisés pour la classification

	Vélopartage				Métro				Taxi			
station	6h	7h	8h	9h	6h	7h	8h	9h	6h	7h	8h	9h
10	0,033	0,118	0,286	0,182	0,021	0,065	0,109	0,064	0,035	0,053	0,113	0,116
11	0,024	0,083	0,176	0,113	0,025	0,059	0,094	0,080	0,034	0,049	0,103	0,122
12	0,010	0,035	0,085	0,058	0,003	0,012	0,019	0,014	0,023	0,046	0,101	0,122
13	0,007	0,027	0,064	0,052	0,005	0,014	0,021	0,023	0,029	0,060	0,138	0,168
14	0,023	0,075	0,137	0,086	0,007	0,028	0,034	0,025	0,030	0,060	0,148	0,171

Quatrièmement, la classification des stations de métro est effectuée à partir des vecteurs formés. L'intensité horaire est cumulée pour les heures subséquentes d'une même combinaison de mode et

de station, afin de réduire la distance entre deux patrons horaires similaires n'ayant qu'un léger décalage horaire. Le partitionnement est effectué à l'aide de la classification hiérarchique avec une distance euclidienne. Le nombre de groupes sélectionné est de 5 groupes, de façon à avoir le même nombre de groupes que la typologie de station basée sur les caractéristiques.

Le dendrogramme résultant de la classification hiérarchique est présenté à la Figure 5-3. Pour faciliter l'interprétation des résultats, la localisation des stations est présentée à la Figure 5-4. Les intensités moyennes des groupes sont présentées à la Figure 5-5, permettant d'identifier comment se distingue chaque groupe.

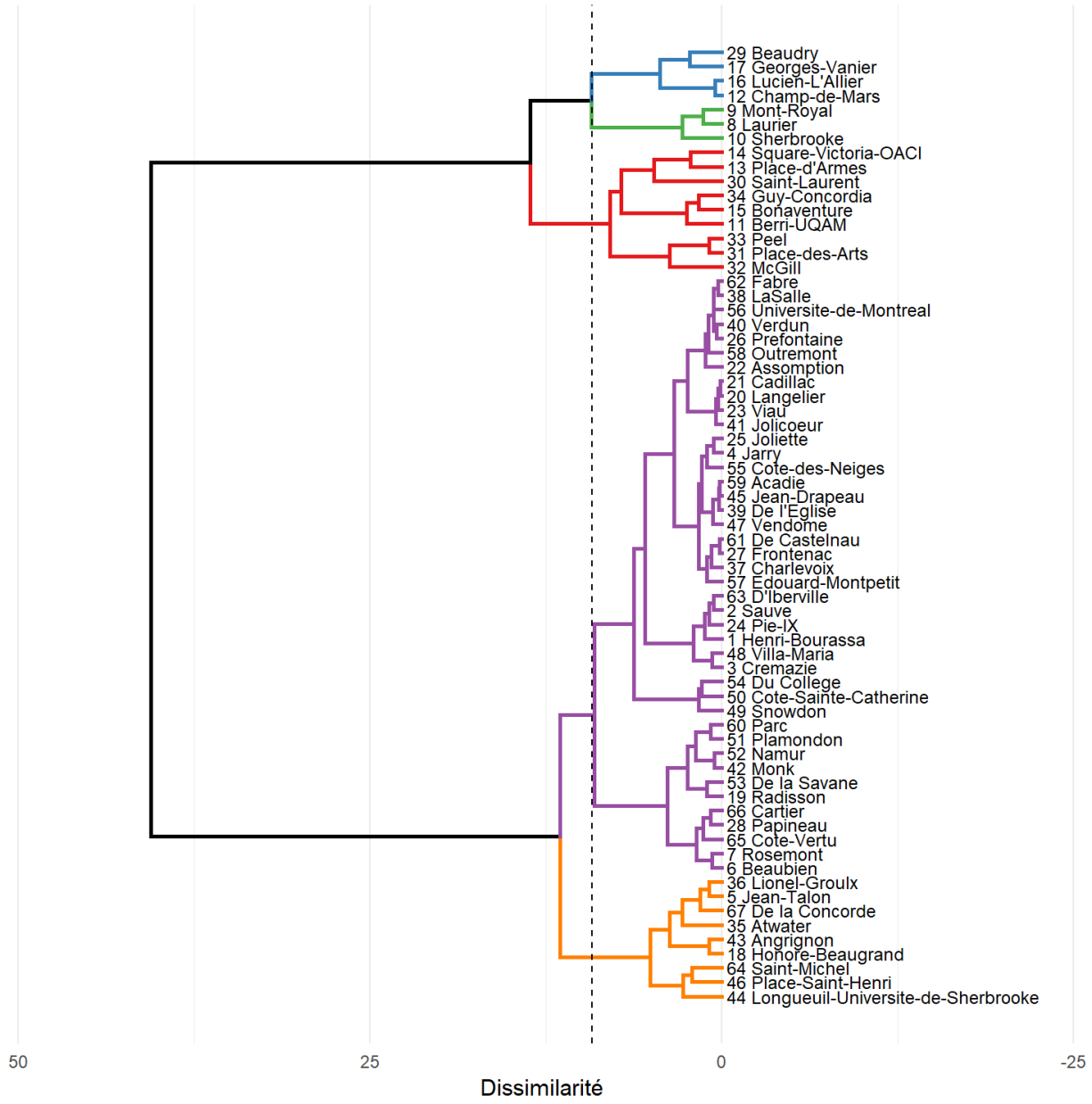

Figure 5-3 : Dendrogramme résultant de la classification selon les patrons journaliers

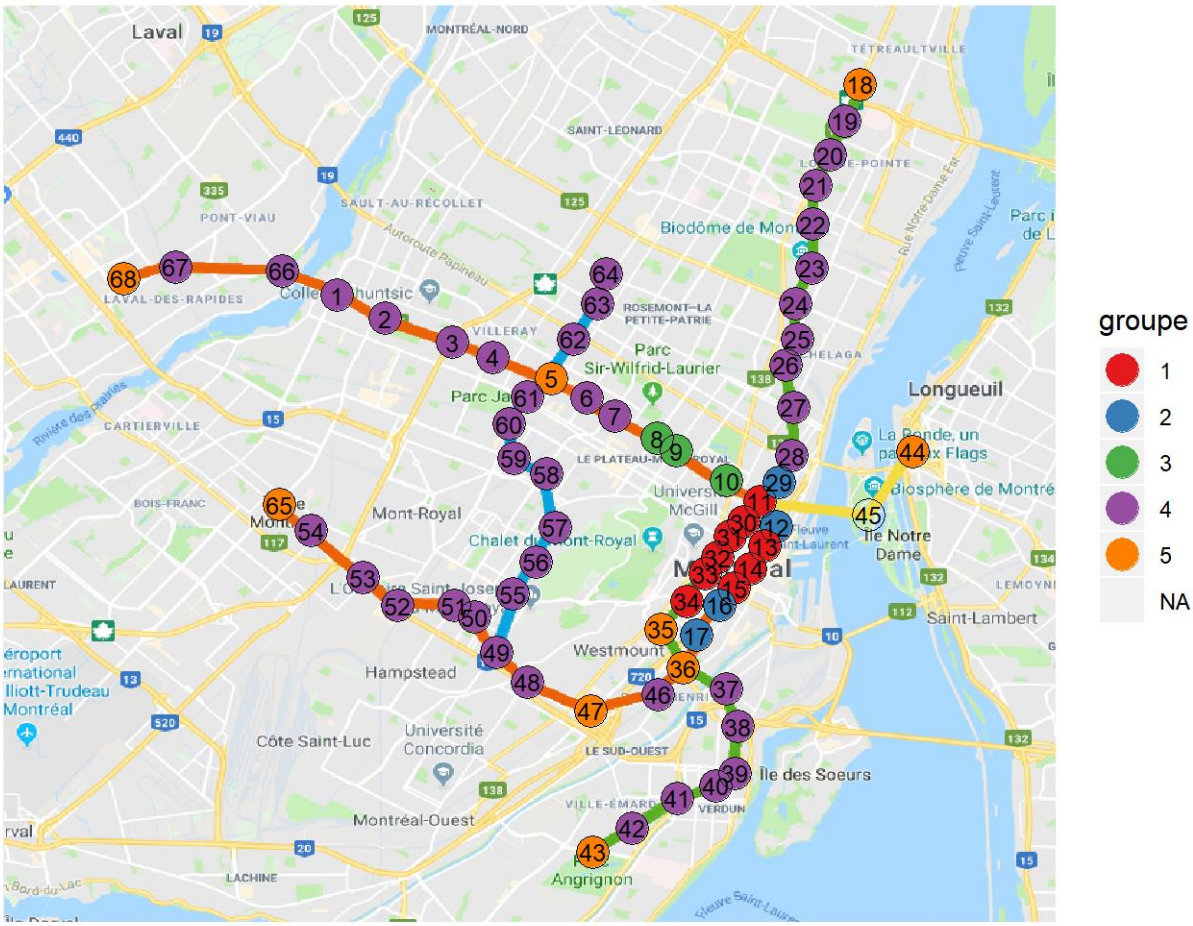

Figure 5-4 : Répartition spatiale des stations par groupe

Les stations similaires semblent groupées spatialement, tel que c'était le cas pour la typologie de station basée sur leurs caractéristiques. Tout d'abord, le groupe 1 contient uniquement des stations du centre-ville. Il s'agit du groupe qui présente la plus forte intensité de départs en pointe du soir pour les trois modes étudiés. En effet, cette zone comporte une grande quantité d'emplois, ce qui explique le grand nombre de départs en fin de journée pour le retour au domicile. Pour le vélopartage et le métro, l'intensité est maximale durant la 17ème heure de la journée. Pour le taxi, cette période de pointe est beaucoup moins prononcée et les déplacements sont plus uniformément répartis durant la journée. Pour les stations de ce groupe, l'usage du taxi est très important le soir et la nuit.

Le groupe 2 contient des stations situées près du centre-ville. La demande pour le métro est plutôt faible et celle-ci est plus importante en pointe du soir que du matin. L'usage du taxi est très élevé dans cette zone.

Le groupe 3 est caractérisé par des stations situées près du centre-ville, plus spécifiquement à l'Est de celui-ci. Ces stations se distinguent par une forte utilisation du vélopartage, tant le matin que le soir. L'intensité d'usage du métro et du taxi est légèrement plus élevée le matin que le soir. L'usage du taxi est important en fin de soirée et atteint un maximum à la 3ème heure de la journée alors que le métro est fermé.

Le groupe 4 contient des stations plus éloignées du centre-ville, majoritairement dans des secteurs résidentiels. Ce groupe présente une intensité d'usage plus élevée en pointe du matin que celle du soir. Or, cette demande demeure plutôt faible.

Finalement, le groupe 5 contient des stations hors centre-ville avec une forte pointe matinale pour le métro. Il regroupe plusieurs stations à l'extrémité des lignes de métro qui collectent les passagers des bus en pointe matinale pour les rediriger vers le métro. L'usage du métro de ce groupe est maximal à la 7ème heure de la journée.

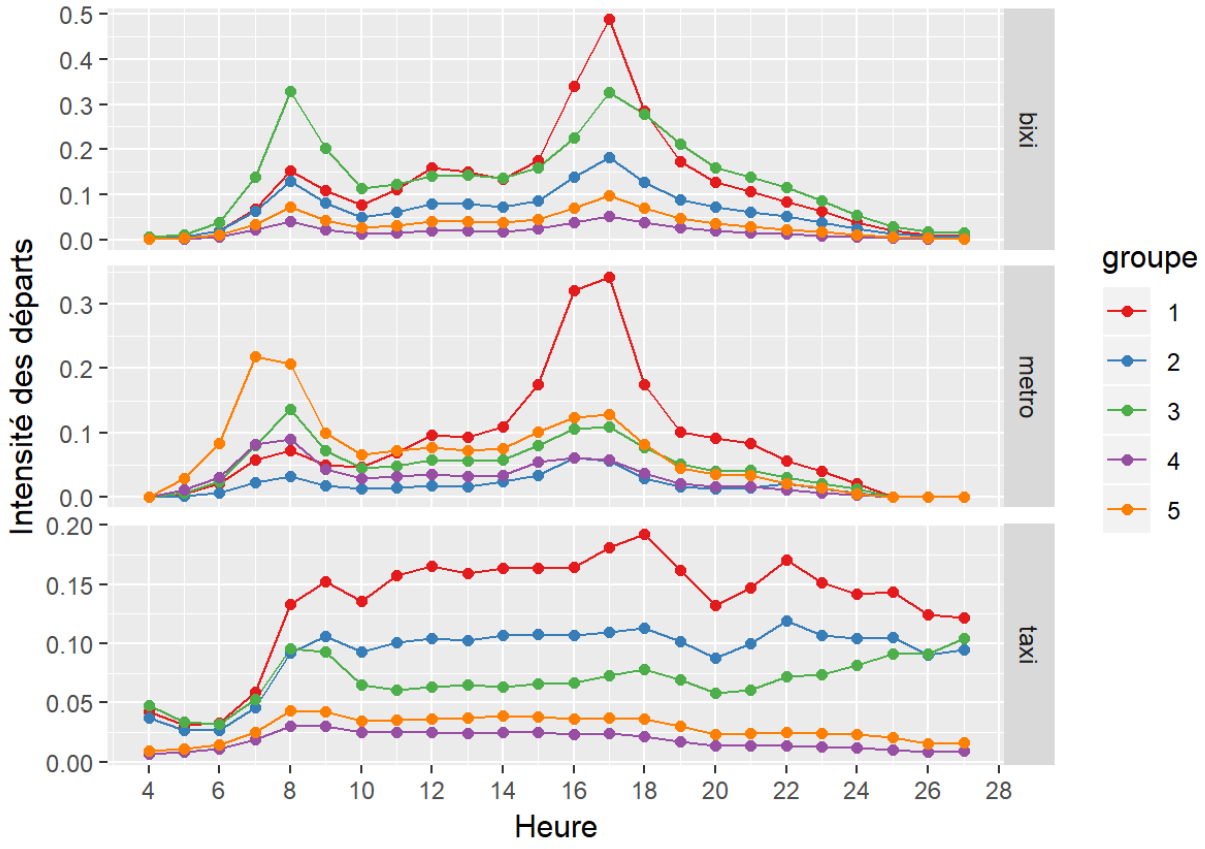

Figure 5-5 : Centres des groupes

Cette classification permet de regrouper des stations similaires selon le volume de la demande et sa répartition durant la journée. Cependant, le volume total journalier semble prendre beaucoup d'importance dans la classification des stations par rapport à la répartition journalière, dû à des grandes différences d'ordres de grandeur dans le volume total de déplacements originant des stations. Voici pourquoi la section suivante présente une classification qui se base uniquement sur la répartition journalière de la demande.

5.2.2 Classification selon la répartition journalière

Cette classification est uniquement basée sur la répartition journalière de la demande et non sur le volume de déplacements. Pour effectuer cette classification, la même procédure que pour la classification précédente est suivie, excepté pour la formule définissant l'indicateur de l'intensité des départs. Le nombre de départs est normalisé selon le mode et la station grâce à l'équation suivante.

$$i_{t,m,h} = \frac{e_{t,m,h}}{r_{t,m}}$$

où

$$r_{t,m} = \sum_{h=1}^{24} e_{t,m,h}$$

i : Intensité des déplacements

d : Nombre de déplacements

r : Nombre de départs de référence (moyenne journalière)

t : Numéro de la station

m : Mode de transport

h : Heure

Une fois les vecteurs créés et la répartition cumulative déterminée, le partitionnement est effectué à l'aide de la classification hiérarchique. Cinq groupes sont sélectionnés, pour tenter de les faire correspondre à ceux identifiés dans la typologie de stations basée sur les caractéristiques des stations.

Le dendrogramme résultant de la classification est présenté à la Figure 5-6 et la localisation des stations est présentée à la Figure 5-7. Les intensités moyennes par groupe sont présentées à la Figure 5-8.

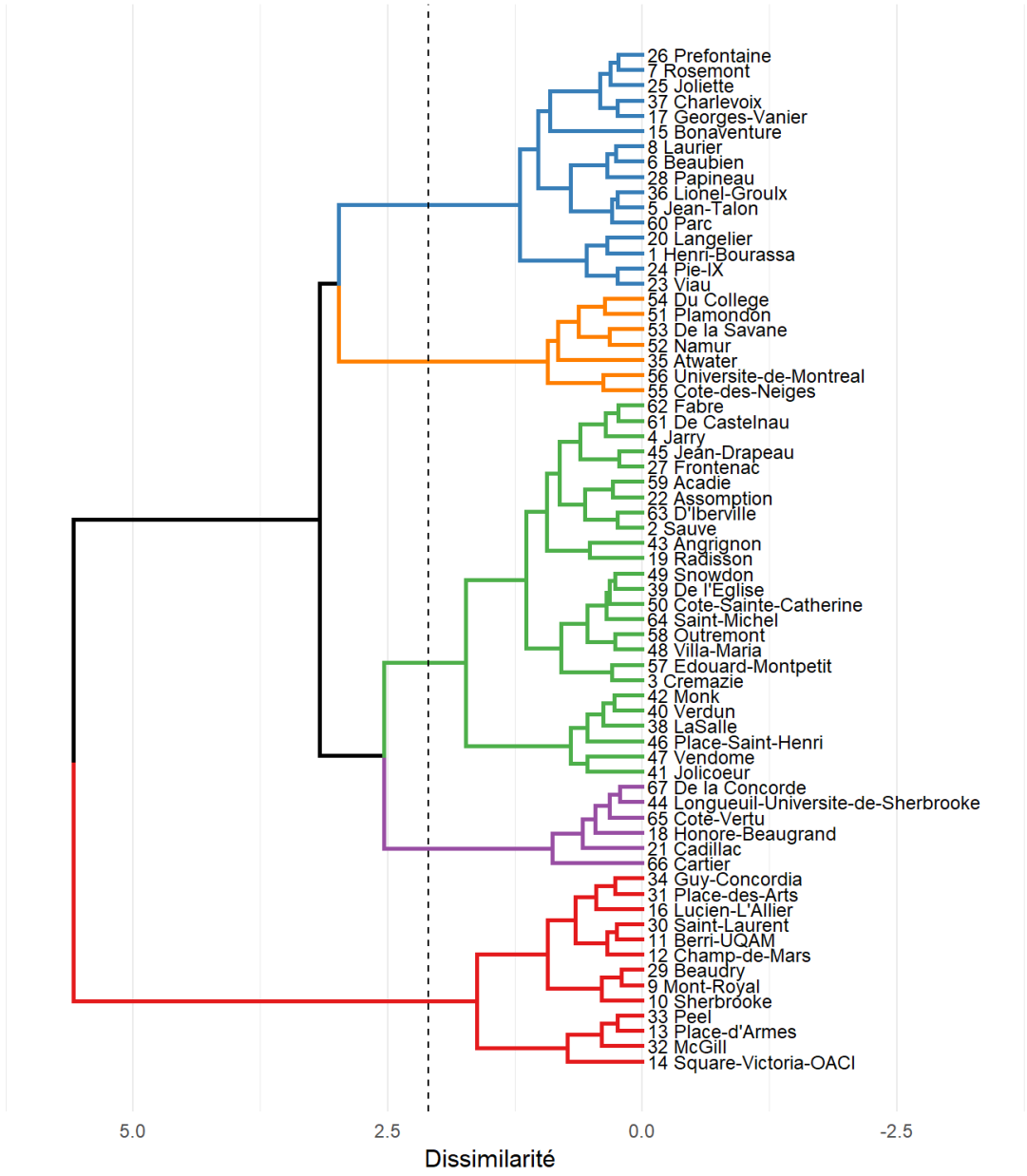

Figure 5-6 : Dendrogramme résultant de la classification selon les patrons journaliers

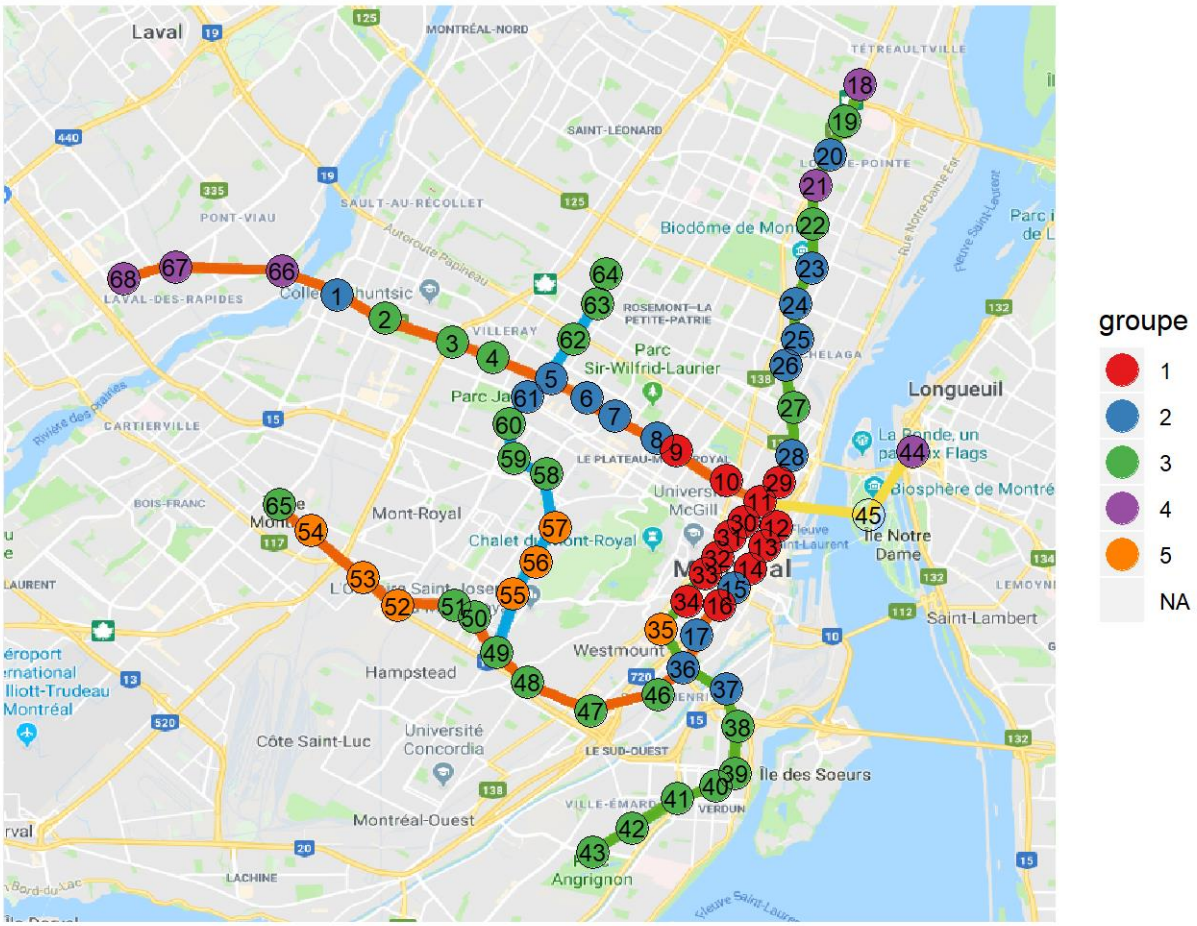

Figure 5-7 : Répartition spatiale des stations par groupe

Les stations semblent groupées de façon spatiale, généralement en fonction de leur proximité au centre-ville. Pour commencer, le groupe 1 contient des stations au centre-ville ou à proximité de celui-ci, où la densité d'emplois est très importante. La demande en pointe matinale est plus faible le matin que le soir pour ce groupe. L'usage du taxi est très important en fin de soirée.

Le groupe 2 contient des stations situées autour du centre-ville en zone davantage résidentielle. La demande du métro et du taxi est plus importante en matinée qu'en soirée. Pour ces modes, la demande est maximale pour la 8ème heure. Similairement au groupe 1, l'usage du taxi est important en fin de soirée.

Le groupe 3 est caractérisé par des stations plus éloignées du centre-ville majoritairement en zone résidentielle. Pour ces stations, l'usage en période matinale est plus élevé que celui de l'après-midi.

Il s'agit du groupe où l'intensité d'usage du vélopartage est la plus élevée en pointe matinale. La demande des différents modes est très faible en période du soir et la nuit.

Le groupe 4 contient les stations les plus éloignées du centre-ville, dont plusieurs stations aux extrémités des lignes de métro. Ce groupe présente un usage très important du métro en pointe du matin, expliqué par plusieurs usagers qui transfèrent de l'autobus au métro.

Finalement, le groupe 5 contient des stations hors centre-ville situées près de grandes institutions scolaires, soit de cégeps ou d'universités. En effet, les stations 52 à 54 sont situées près des cégep Marie-Victorin, du cégep Saint-Laurent et du cégep Vanier. Les stations 55 à 57 sont situées près de l'Université de Montréal et du Collège Brébeuf. Finalement, la station 35 est située près du Collège Dawson. Ces stations présentent un usage du vélopartage et du métro similaire au groupe 1, soit un usage très important en période d'après-midi. Cependant, le groupe 5 se distingue du groupe 1 par un usage du taxi très faible le soir et la nuit.

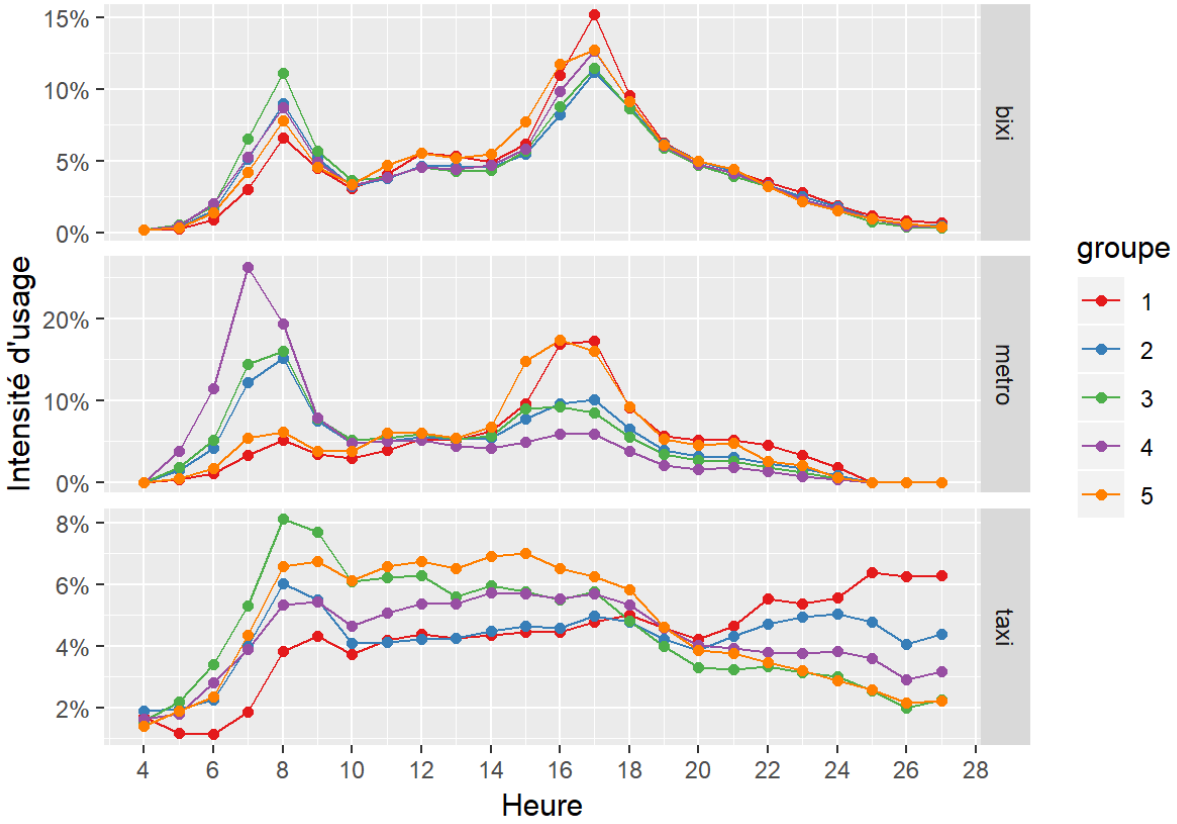

Figure 5-8 : Centres des groupes

Cette seconde classification basée uniquement sur la répartition journalière de la demande semble être plus représentative du type d'usage des différentes stations et semble mieux correspondre à la typologie créée précédemment. Cette classification sera employée afin de relier la typologie des stations au type d'usage.

5.3 Association des patrons journaliers à la typologie

Cette section s'intéresse à la comparaison de la typologie de station basée sur les caractéristiques du voisinage avec les patrons journaliers d'usage identifiés à la section précédente. L'objectif est de déterminer s'il est possible de prédire le type d'usage des stations à partir des caractéristiques de celles-ci. Les deux classifications sont comparées en associant les groupes les plus similaires.

Le Tableau 5-3 compare les deux classifications des stations par une matrice de confusion. La diagonale de la matrice présente les paires de groupes qui concordent le mieux. Il faut rappeler que les groupes ont été ordonnés par distance croissante au centre-ville en métro.

Tableau 5-3 : Matrice de confusion

		Classification selon les patrons journaliers					
		1	2	3	4	5	Total
Typologie basée sur les caractéristiques	1	9	1	0	0	0	10
	2	4	4	1	0	1	10
	3	0	8	21	0	1	30
	4	0	3	3	6	2	14
	5	0	0	0	0	3	3
	Total	13	16	25	6	7	67

Le nombre de stations dans chaque groupe est différent pour chaque classification, c'est pourquoi il y a forcément des stations qui n'ont pas été classées dans les groupes concordants. Un total de 43 stations sont classées de manière identique par les deux classifications, ce qui correspond à un taux de concordance de 64%. La majorité des stations mal classées se trouvent dans deux classes voisines, ce qui est le cas pour 79% des stations dont les classements diffèrent. Ces groupes voisins se ressemblent généralement, étant donné que les 5 groupes ont été ordonnés par distance moyenne en métro à partir du centre-ville. La Figure 5-9 permet de localiser les stations dont les classifications diffèrent.

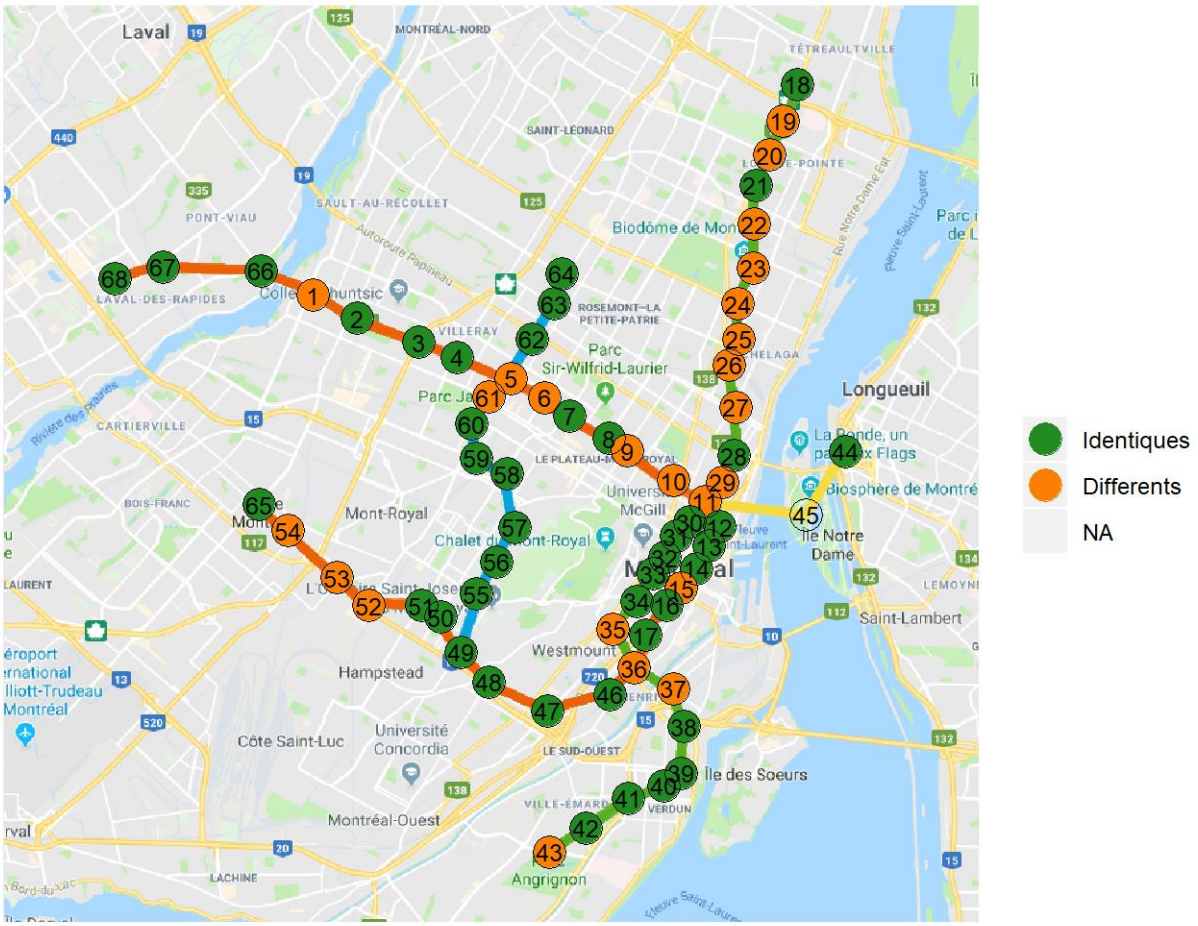

Figure 5-9 : Spatialisation des différences de classification

Les stations du centre-ville semblent généralement bien classées. Plusieurs stations différemment classées se retrouvent à l'Est de la ligne verte. Plusieurs sont classées dans le groupe 4 par la typologie basée sur les caractéristiques, alors qu'elles sont classées dans le groupe 3 par la classification selon les patrons journaliers. Par ailleurs des stations comportant des cégeps importants à proximité (stations 35, 52, 53 et 54) n'ont pas été classées dans le même groupe que les stations 55 à 57 près de l'Université de Montréal pour la typologie basée sur les caractéristiques. Ensuite, la station 15 (Bonaventure) a été classée dans le groupe 2 par la classification basée sur les patrons journaliers malgré le fait qu'elle soit située au cœur du centre-ville. Cela peut être expliqué par un nombre important de transferts effectués à la station, où les usagers valident à nouveau leur carte à puce. En effet, à cette station en matinée, un grand nombre d'usagers arrivent en bus ou en train et transfèrent au métro, ce qui implique une validation supplémentaire lors de l'entrée dans le réseau du métro.

Pour conclure, une telle démarche pourrait servir à estimer l'impact d'une modification au réseau existant de métro. En effet, cela peut permettre d'estimer la demande potentielle des différents modes après l'ajout d'une nouvelle station de métro, en se basant sur les caractéristiques de son voisinage. Même si le taux de succès n'est que de 64%, cette démarche permet tout de même de donner un aperçu de la demande des différents modes, sans recourir à des méthodes complexes de modélisation. Une autre utilité possible est la simplification de certains modèles en regroupant les stations similaires.

CHAPITRE 6 IMPACT D'ÉVÈNEMENTS ATYPIQUES SUR L'USAGE

Cette section a pour objectif d'estimer la fluctuation de la demande horaire causée par certains évènements atypiques et ce, pour les jours ouvrables. Les évènements étudiés sont des évènements météorologiques, les interruptions de service du métro et différentes activités (concerts, festivals, matchs sportifs, etc.). D'autres types d'évènements peuvent avoir un impact sur l'usage tel que la variation du prix de l'essence ou des modifications à l'offre de transport. Or, cette analyse se concentre sur des types d'évènements qui influencent l'usage à court terme. L'unité temporelle considérée est l'heure, soit une période suffisamment courte pour capturer l'impact des évènements de courte durée, comme par exemple des courtes interruptions de service du métro.

Quatre modes sont comparés, soit le vélopartage, le taxi, le métro et l'autobus. Des modèles explicatifs sont calibrés individuellement pour chaque mode, afin d'évaluer l'impact des différents facteurs sur la demande. Une meilleure compréhension des facteurs influençant la demande en transport permet aux opérateurs d'envisager une adaptation plus dynamique de l'offre de transport selon la demande.

Tout d'abord, un modèle est calibré sur l'ensemble des stations, afin d'identifier les variables évènementielles ayant un impact significatif sur l'usage d'un mode. Ensuite, des modèles différents sont calibrés pour chaque type de station défini par la typologie de stations de métro. Puis des modèles sont calibrés pour chaque station, de façon à observer des corrélations spatiales.

6.1 Modèles sur l'ensemble des stations

L'objectif est de calibrer un modèle de type GAM (modèle additif généralisé) pour chaque mode de transport étudié en utilisant les données de l'ensemble des stations de métro, afin de comparer l'influence d'évènements atypiques sur l'usage des divers modes. La variable d'intérêt est donc le nombre de déplacements horaires initiés.

Les différentes étapes méthodologiques sont présentées ici. Tout d'abord, les données sont filtrées, le nombre de déplacements est normalisé, puis les données sont fusionnées. Ensuite, la variable d'intérêt est transformée, les variables sont sélectionnées et finalement les périodes horaires sont formées. Les prochaines sections décrivent chacune de ces étapes.

6.1.1 Filtrage des données

La première étape consiste à filtrer les données afin de ne conserver que les observations pertinentes à la calibration des modèles. Tout d'abord, seuls les jours ouvrables sont conservés, car il est attendu que ceux-ci soient différemment influencés par les évènements spéciaux que les jours non-ouvrables.

Les données sont également filtrées afin de ne conserver que les observations en journée et en soirée entre 6h et 23h inclusivement. Il s'agit des heures complètes comprises dans la période d'ouverture du métro en semaine. Il s'agit donc d'une période commune où les quatre modes étudiés sont disponibles. Éliminer les heures nocturnes permet également d'éviter d'avoir trop d'observations horaires avec un nombre de déplacements nul, ce qui pourrait être problématique pour la calibration d'un modèle de forme linéaire. En effet, un nombre trop élevé d'observations nulles rendrait l'impact des variables exogènes difficilement identifiable.

Quelques stations à faible usage sont exclues des observations pour le vélopartage et le taxi, étant donné que les fluctuations d'usage sont difficiles à identifier. Ainsi, les observations du vélopartage des 15 stations ayant le moins de déplacements sont exclues. Similairement, les observations du taxi des 12 stations présentant le moins de déplacements sont exclues.

Pour le vélopartage, seule la période de disponibilité de service est conservée. Les observations des deux premières semaines de service au printemps et des deux dernières semaines à l'automne sont exclues étant donné qu'il s'agit de périodes de transition durant lesquelles l'usage n'est pas représentatif du reste de la saison.

6.1.2 Normalisation de la demande

La seconde étape consiste à normaliser le nombre de déplacements, dans l'objectif de combiner les données des différentes stations bien que les nombres de déplacements soient d'ordres de grandeur différents. Par ailleurs, certains modes tel que le vélopartage voient leur nombre de déplacements augmenter d'année en année, c'est pourquoi le nombre de départs est aussi normalisé par rapport à l'année. Une fois normalisé par rapport à la station et à l'année, le nombre de déplacements est désigné comme une intensité de déplacements. Il est à noter que cette normalisation est effectuée indépendamment pour chaque mode. L'équation suivante résume le calcul de cette intensité.

Équation 6-1 : Calcul de l'intensité des déplacements pour un mode

$$intensite_{h,a,s} = \frac{déplacements_{h,a,s}}{m_{a,s}}$$

où $m_{a,s} = \frac{1}{n_a} \sum_{h=1}^{n_a} déplacements_{h,a,s}$

h : période horaire

a : année

s : station de métro

$m_{a,s}$: nombre moyen de déplacements pour une année-station

n_a: nb périodes horaires pour l'année a

Ainsi, une intensité de 1,05 signifie que le nombre de déplacements est 5% supérieur au nombre de déplacements moyen pour la station et l'année concernées.

6.1.3 Fusion de données

La troisième étape consiste à fusionner les données de diverses sources. La fusion permet de regrouper dans une table commune les données des déplacements, la météorologie horaire, la météorologie journalière, les activités, les interruptions de service et certaines variables temporelles. Cette fusion est effectuée en se basant sur les champs d'horodatage et de station, excepté pour la météorologie qui est supposée identique sur l'ensemble des stations.

En ce qui concerne les données d'activités, celles-ci ne comprennent que les activités majeures à proximité du réseau de métro, pour lesquelles la STM a prévu une augmentation significative de la demande. La majorité des activités sont associées à une seule station de métro, soit celle la plus près de l'évènement. Deux variables binaires sont ajoutées à la table pour chaque catégorie d'activité. Ces variables binaires indiquent la présence d'une activité, mais pour deux périodes différentes, soit la période durant l'activité et la période à la fin de l'activité. La période durant l'activité s'étend de son heure de début jusqu'à une heure après son heure de fin. De son côté, la période à la fin de l'activité débute une heure avant la fin de l'activité et se termine une heure après sa fin. Cela permet de capturer 2 types d'activités, soit un type où les participant peuvent quitter au moment désiré tel qu'une exposition, et un autre type où tous les participants quittent simultanément à la fin, tel qu'il est généralement le cas pour une partie de soccer. D'autres périodes

horaires autour de l'activité ont été testées, mais les deux périodes présentées ici sont celles qui se sont avérées les plus pertinentes à la minimisation de l'erreur des modèles.

Les interruptions de service du métro sont ajoutées à la table par une variable totalisant le nombre de minutes en interruption pour chaque combinaison d'heure et de station. Une valeur de 0 signifie donc qu'il n'y a pas eu d'arrêts de service, alors qu'une valeur de 60 signifie que l'heure entière a été en arrêt de service à cette station.

6.1.4 Transformation de la variable d'intérêt

La variable d'intérêt est transformée par une fonction logarithmique pour que sa distribution tende davantage vers une distribution normale. L'intensité est augmentée d'une unité avant d'appliquer la transformation logarithmique, afin d'éviter que le logarithme d'un nombre de déplacements nul soit une valeur indéfinie. La figure suivante compare la distribution des intensités de déplacement avec et sans transformation logarithmique.

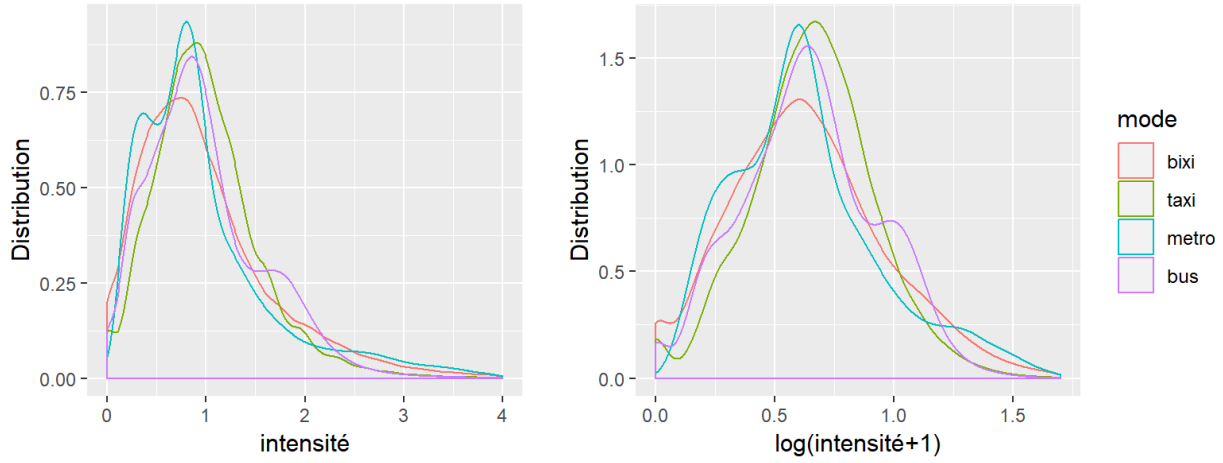

Figure 6-1 : Distribution des intensités de déplacement selon la transformation

Il est possible d'observer que la distribution avec la transformation logarithmique s'apparente davantage à la distribution normale, ce qui est souhaitable pour un modèle linéaire. Le graphe quantile-quantile (*qq-plot*) a également été tracé pour chaque mode afin de confirmer le tout (les tests statistiques n'étant pas recommandés pour de grands échantillons).

Pour chaque mode, la transformation logarithmique entraine une diminution de l'erreur du modèle. La transformation par la racine carrée est aussi considérée, mais elle n'est pas retenue étant donné

que les coefficients résultants sont difficilement interprétables et que l'erreur quadratique est similaire.

6.1.5 Sélection des variables

Pour la cinquième étape, les variables pertinentes aux modèles sont sélectionnées. Pour ce faire, des graphiques de la variable d'intérêt en fonctions des variables explicatives sont tracés, ce qui permet d'identifier s'il y a présence de corrélation et si celle-ci semble linéaire. Dans le cas d'une relation non-linéaire, la variable peut être transformée par une fonction par paliers ou par une spline de lissage.

Un corrélogramme est tracé afin d'identifier les variables colinéaires. Une paire de variables ayant une corrélation supérieure à 0.4 est considérée comme colinéaire. Dans cette situation, une seule variable parmi celles colinéaires est ajoutée au modèle afin d'assurer que l'effet de chaque variable est dissociable. La sélection de la variable la plus pertinente est effectuée selon le critère de l'AIC (critère d'information d'Akaike). Ce critère à minimiser tient compte de la vraisemblance et pénalise les modèles avec un plus grand nombre de paramètres.

6.1.6 Formation des périodes horaires

La sixième étape a comme objectif de faciliter l'interprétation des résultats du modèle en diminuant le nombre de coefficients liés à la temporalité. Pour ce faire, les heures consécutives similaires sont regroupées au sein d'une même période horaire. La méthode est présentée ici en utilisant les données du vélopartage et cette procédure est identique pour les autres modes.

Tout d'abord, il est nécessaire de déterminer si les différents jours de la semaine peuvent être agrégés. L'intensité horaire moyenne selon le jour est donc tracé à la Figure 6-2 pour le vélopartage. Les différents jours semblent avoir des intensités horaires similaires. Le test de Kolmogorov–Smirnov à 2 échantillons est complété sur chaque paire de jours. Le test ne rejette pas l'hypothèse que les échantillons des différents jours proviennent de la même distribution, et ce à un niveau de confiance de 95%. Les résultats détaillés du test sont présentés à l'Annexe D. Ainsi, il est décidé de grouper les différents jours pour le vélopartage étant donné leur distribution horaire similaire.

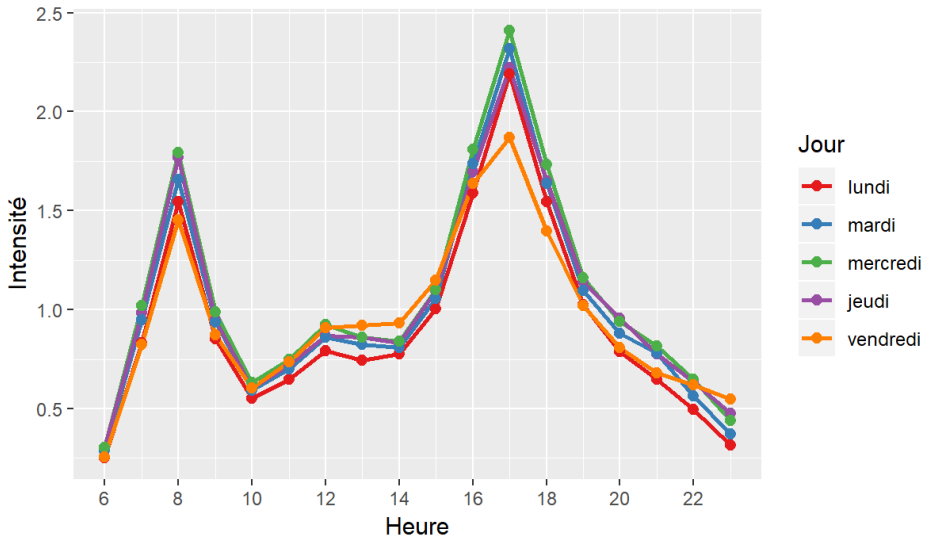

Figure 6-2 : Intensité horaire moyenne selon le jour, pour le vélopartage

Le nombre de périodes horaires est réduit en regroupant les heures similaires consécutives, tout en s'assurant que l'erreur du modèle soit peu affectée. L'indicateur proposé pour comparer les heures est la distribution des intensités de déplacement, ce qui est plus rigoureux que la médiane ou la moyenne. Les distributions statistiques pour chaque heure sont présentées à la Figure 6-3.

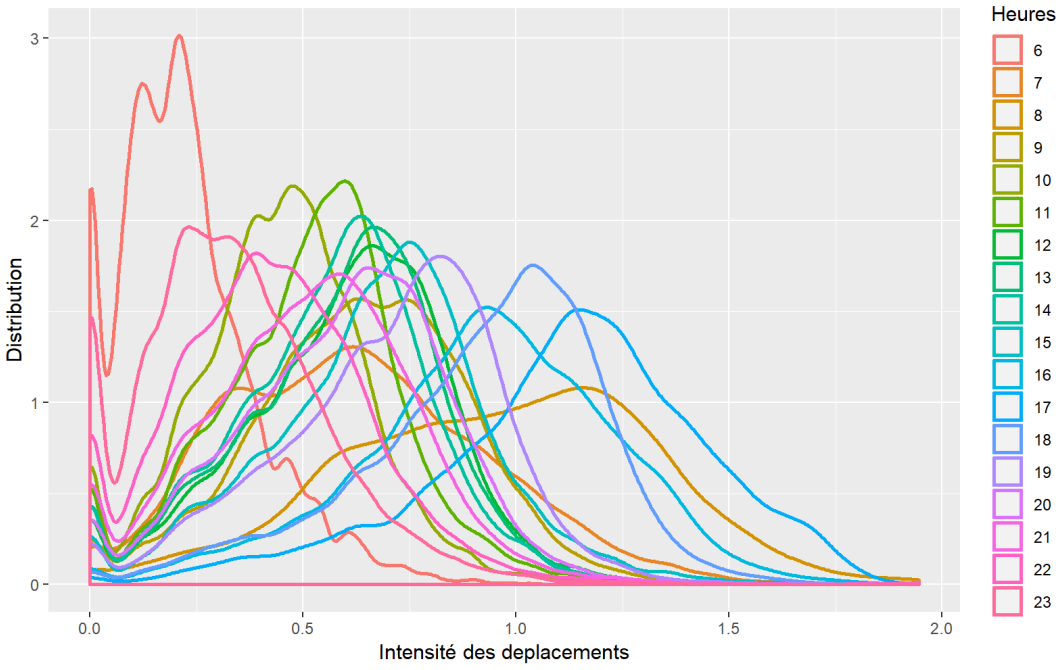

Figure 6-3 : Distribution des intensités de déplacements par heure, pour le vélopartage

Il est possible d'identifier plusieurs distributions similaires, comme par exemple celles entre 12h et 15h environ. Celles-ci pourraient possiblement être groupées afin de simplifier le modèle. Le test de Kolmogorov-Smirnov à deux échantillons permet de déterminer si deux distributions statistiques diffèrent. La statistique du test est la distance verticale maximale qui sépare les deux fonctions de répartition. L'hypothèse nulle est que les deux échantillons proviennent de la même distribution. Les heures peuvent être comparées deux à deux afin de déterminer si leur regroupement est envisageable. La Figure 6-4 présente, en guise d'exemple, la comparaison des heures 7 et 8 avec le test de Kolmogorov-Smirnov. Sur cette figure, la distance verticale maximale entre les deux fonctions de répartition est identifiée en pointillés.

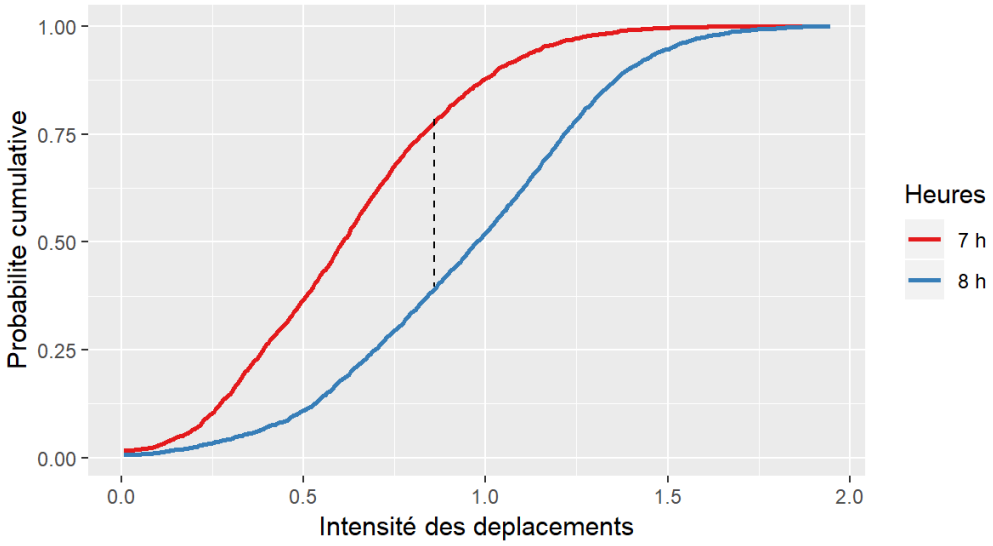

Figure 6-4 : Distance maximale obtenue par le test KS pour une paire d'heures

Dans ce cas, il est possible de constater que les deux distributions diffèrent grandement, ce qui est confirmé par une statistique du test élevée de 0.39. Cette statistique est calculée pour les autres paires d'heures et les résultats sont présentés à la Figure 6-5.

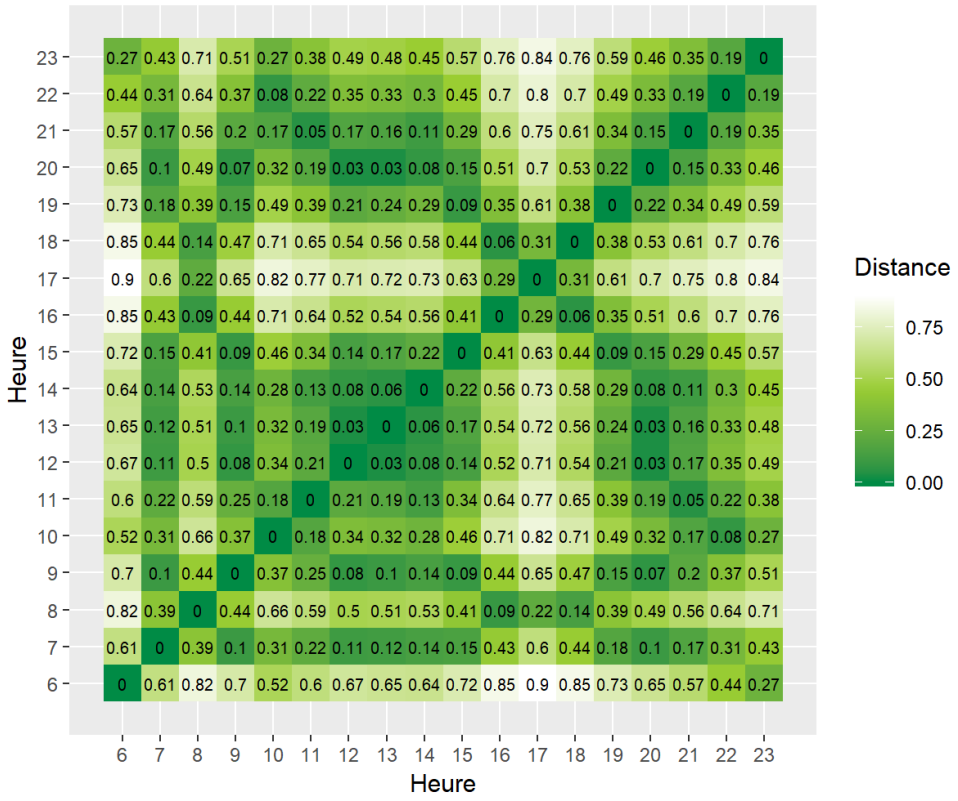

Figure 6-5 : Matrice de distances maximales pour chaque paire d'heures, pour le vélopartage

Des distances plus élevées sont observables entre les heures en pointe et hors-pointe. La période continue avec les distances les plus faibles est celle au milieu de la journée, entre 12h et 15h environ. Or, un seuil doit être établi afin de fixer la distance maximale acceptable pour regrouper deux heures dans la même période. Pour ce faire, il est possible d'estimer l'erreur moyenne quadratique d'entrainement du modèle pour différents seuils de distance. Le modèle utilisé est celui défini par l'Équation 6-2, mais dont les variables binaires identifiant la période horaire varient.

La Figure 6-6 présente le nombre de périodes formées selon la distance seuil fixée et la Figure 6-7 présente l'erreur selon la distance seuil fixée.

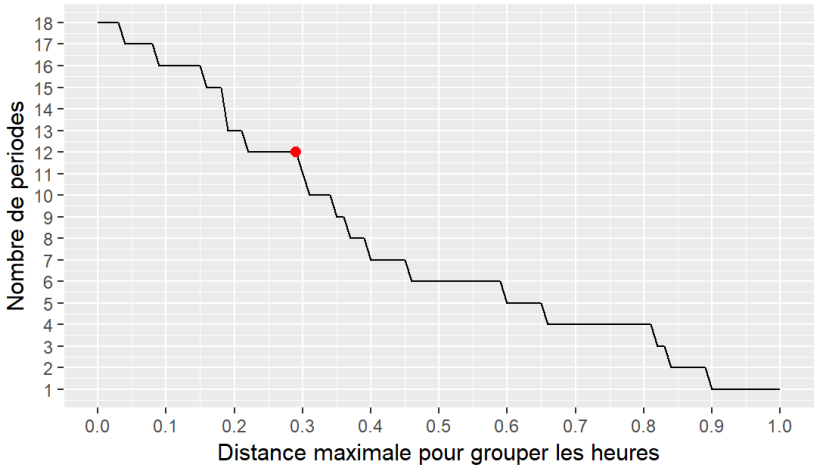

Figure 6-6 : Nombre de périodes formées selon la distance seuil choisie, pour le vélopartage

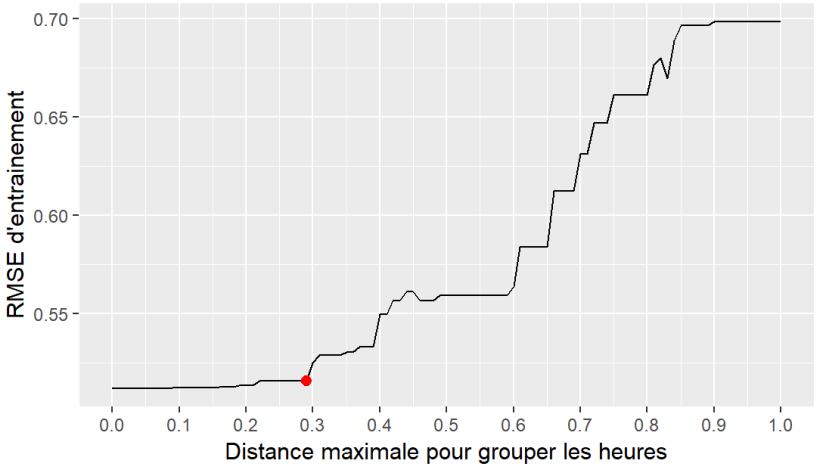

Figure 6-7 : Erreur du modèle selon la distance seuil choisie, pour le vélopartage

Le nombre de périodes diminue graduellement avec l'augmentation du seuil de distance, étant donné que davantage de regroupements sont effectués. De son côté, l'erreur quadratique d'entrainement augmente progressivement étant donné que le nombre de variables explicatives du modèle diminue. Une distance maximale de 0,29 est fixée, soit celle identifiée sur les graphiques par un point rouge. Ce seuil de distance correspond à 12 périodes. Jusqu'à ce seuil, l'augmentation de l'erreur est négligeable, soit de 0,511 à 0,516. Pour des seuils de distance supérieurs, l'erreur augmente de façon plus importante, ce qui diminue les performances du modèle de façon plus importante.

La Figure 6-8 présente l'erreur quadratique du modèle selon le nombre de périodes formées. Plus le nombre de périodes est élevé, plus faible est l'erreur quadratique du modèle.

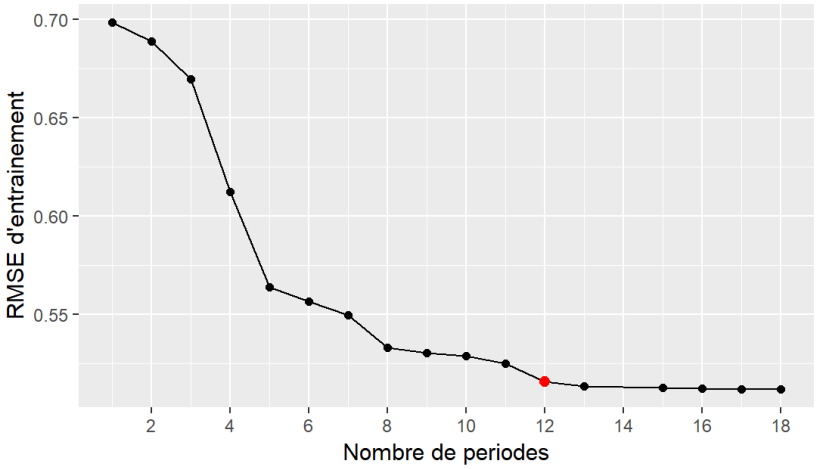

Figure 6-8 : Erreur du modèle selon le nombre de périodes choisi, pour le vélopartage

En réduisant le nombre de périodes de 18 à 12, cela facilite l'interprétation des coefficients du modèle sans toutefois augmenter l'erreur de façon importante. La Figure 6-9 présente les paires d'heures respectant la distance seuil (colorées en gris). Les heures regroupées dans une même période sont identifiées par la même couleur.

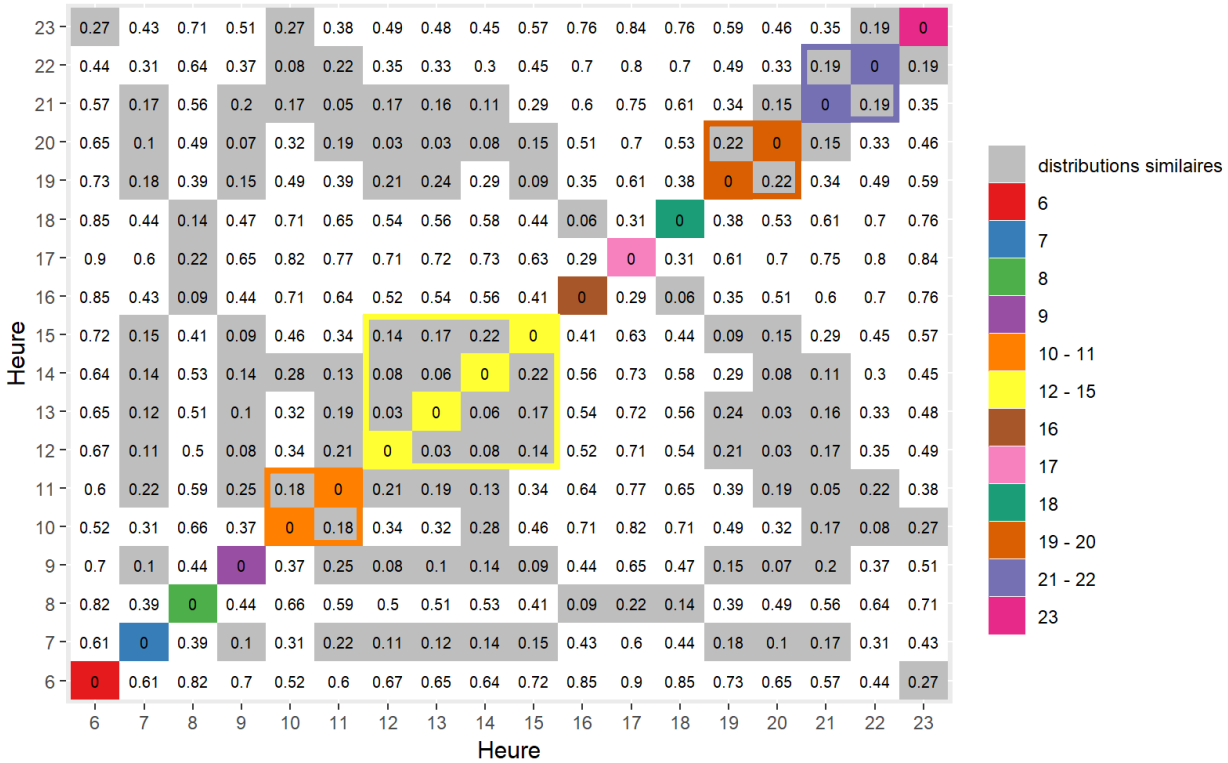

Figure 6-9 : Périodes horaires formées selon le seuil maximal de distance, pour le vélopartage

Les regroupements d'heures sont formés de paires d'heures consécutives respectant le seuil, c'est pourquoi chaque période délimitée par un carré ne contient que des distances inférieures au seuil, soit des paires d'heures colorées en gris. Parmi les 12 périodes horaires, quatre regroupent plusieurs heures. La période la plus longue est d'une durée de 4 heures et s'étend de 12h à 15h.

La distribution statistique pour chaque période est présentée à la Figure 6-10.

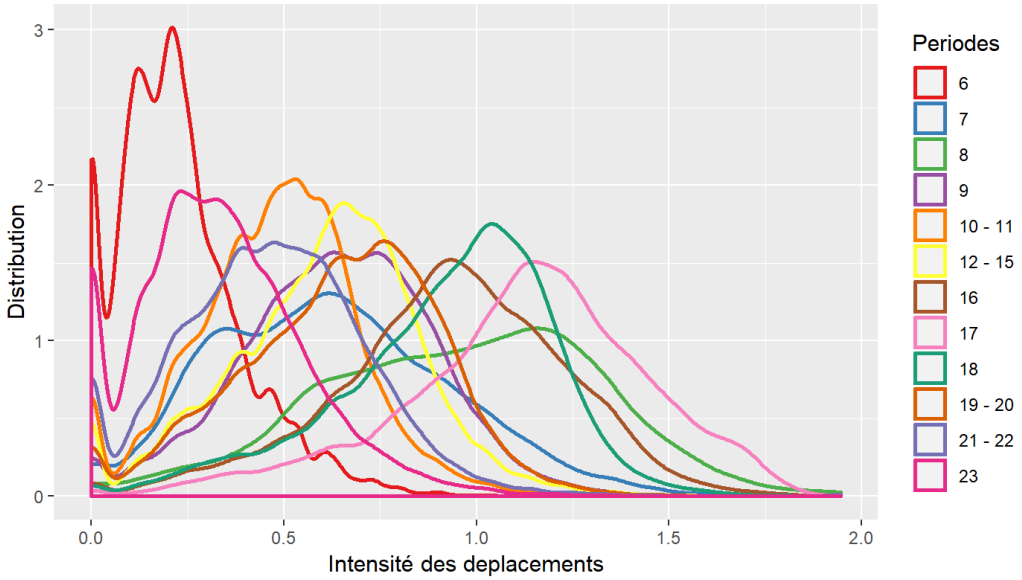

Figure 6-10 : Distribution des déplacements par période, pour le vélopartage

Les périodes de pointe du vélopartage sont facilement identifiables par des distributions dont l'intensité est plus élevée que les autres. Certaines distributions sont très similaires mais n'ont pas été regroupées étant donné qu'elles ne sont pas consécutives.

Cette même méthode est utilisée afin de former des périodes horaires pour les autres modes de transport. Le tableau suivant présente les résultats de la formation des périodes horaires pour le vélopartage, le métro et le bus. Les résultats détaillés incluant le choix du seuil sont présentés à l'annexe A.

Tableau 6-1 : Périodes horaires formées selon le mode

<table>
<tr><th>Heures</th><th>Vélopartage</th><th>Métro</th><th>Bus</th></tr>
<tr><td>6</td><td>6</td><td>6</td><td>6</td></tr>
<tr><td>7</td><td>7</td><td rowspan="2">7 - 8</td><td rowspan="2">7 - 8</td></tr>
<tr><td>8</td><td>8</td></tr>
<tr><td>9</td><td>9</td><td>9</td><td rowspan="6">9 - 14</td></tr>
<tr><td>10</td><td rowspan="2">10 - 11</td><td rowspan="5">10 - 14</td></tr>
<tr><td>11</td></tr>
<tr><td>12</td><td rowspan="3">12 - 15</td></tr>
<tr><td>13</td></tr>
<tr><td>14</td></tr>
<tr><td>15</td><td rowspan="3">15 - 17</td><td rowspan="3">15 - 17</td></tr>
<tr><td>16</td><td>16</td></tr>
<tr><td>17</td><td>17</td></tr>
<tr><td>18</td><td>18</td><td>18</td><td>18</td></tr>
<tr><td>19</td><td rowspan="2">19 - 20</td><td rowspan="3">19 - 21</td><td rowspan="3">19 - 21</td></tr>
<tr><td>20</td></tr>
<tr><td>21</td><td rowspan="2">21 - 22</td></tr>
<tr><td>22</td><td rowspan="2">22 - 23</td><td rowspan="2">22 - 23</td></tr>
<tr><td>23</td><td>23</td></tr>
</table>

Les périodes horaires du bus et du métro sont très similaires. Le vélopartage compte 12 périodes horaires, alors que le métro en contient 8 et le bus n'en contient que 7.

Du côté du taxi, l'intensité moyenne horaire varie de façon importante selon le jour de la semaine tel que démontré sur la Figure 6-11, c'est pourquoi chaque jour est traité indépendamment.

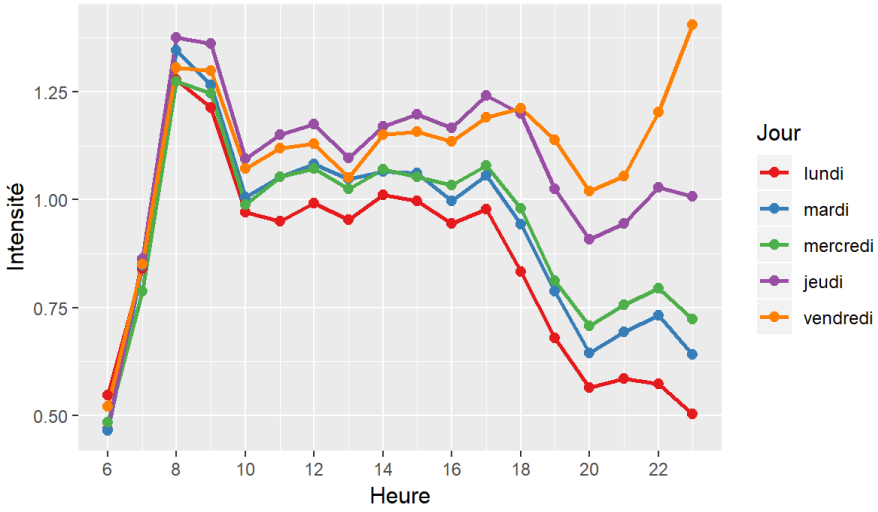

Figure 6-11 : Intensité moyenne horaire selon le jour, pour le taxi

La Figure 6-11 montre que les intensités se ressemblent en début de journée, puis diffèrent en milieu et en fin de journée. Plus la semaine progresse, plus l'intensité en fin de journée augmente. Au milieu de la journée, l'intensité est relativement constante, laissant croire que ces heures peuvent être regroupées en une seule et même période. Le Tableau 6-2 présente les regroupements d'heures continues en périodes horaires pour chaque jour de la semaine.

Tableau 6-2 : Périodes horaires formées pour le taxi

<table>
<tr><th>Heures</th><th>Lundi</th><th>Mardi</th><th>Mercredi</th><th>Jeudi</th><th>Vendredi</th></tr>
<tr><td>6</td><td>6</td><td>6</td><td>6</td><td>6</td><td>6</td></tr>
<tr><td>7</td><td>7</td><td>7</td><td>7</td><td>7</td><td>7</td></tr>
<tr><td>8</td><td rowspan="2">8 - 9</td><td rowspan="2">8 - 9</td><td rowspan="2">8 - 9</td><td rowspan="2">8 - 9</td><td rowspan="2">8 - 9</td></tr>
<tr><td>9</td></tr>
<tr><td>10</td><td rowspan="9">10 - 18</td><td rowspan="9">10 - 18</td><td rowspan="9">10 - 18</td><td rowspan="8">10 - 16</td><td rowspan="8">10 - 17</td></tr>
<tr><td>11</td></tr>
<tr><td>12</td></tr>
<tr><td>13</td></tr>
<tr><td>14</td></tr>
<tr><td>15</td></tr>
<tr><td>16</td></tr>
<tr><td>17</td><td rowspan="2">17 - 18</td></tr>
<tr><td>18</td><td rowspan="2">18 - 19</td></tr>
<tr><td>19</td><td rowspan="4">19 - 22</td><td rowspan="5">19 - 23</td><td rowspan="5">19 - 23</td><td rowspan="5">19 - 23</td></tr>
<tr><td>20</td><td rowspan="2">20 - 21</td></tr>
<tr><td>21</td></tr>
<tr><td>22</td><td>22</td></tr>
<tr><td>23</td><td>23</td><td>23</td></tr>
</table>

Cette méthode permet de réduire le nombre de périodes de 90 à 30, ce qui diminue le nombre de variables à évaluer pour le modèle. En début de journée, les périodes horaires sont similaires pour tous les jours de la semaine. La période de 8h à 9h semble être la pointe du matin tel qu'observé sur la Figure 6-11. En fin de journée, les périodes des premiers jours de la semaine sont similaires, alors que le vendredi se distingue par davantage de périodes horaires. Le nombre élevé de périodes horaires, le vendredi soir, peut être expliqué par une diminution d'achalandage en soirée, suivi d'une augmentation importante de l'achalandage, tel qu'observé à la Figure 6-11.

6.1.7 Résultats des modèles

Les équations suivantes présentent les modèles de type GAM (Modèle additif généralisé) calibrés pour chaque mode. Le Tableau 6-3 définit les différentes variables.

Équation 6-2 : Équation du modèle pour le vélopartage

$$ln(intensité_{vélo}) = \beta_0 + \beta_1\ pluie_{3h+A} + \beta_2\ vent + f_1(température) + f_2(panne) + \sum_{i=1}^{8} \alpha_i\ activité_i + \sum_{i=1}^{11} \gamma_i\ période_i + \epsilon$$

Équation 6-3 : Équation du modèle pour le taxi

$$ln(intensité_{taxi}) = \beta_0 + \beta_1\ pluie_{1h+A} + f_1(température) + f_2(panne) + \sum_{i=1}^{8} \alpha_i\ activité_i + \sum_{i=1}^{29} \gamma_i\ période_i + \epsilon$$

Équation 6-4 : Équation du modèle pour le métro

$$ln(intensité_{métro}) = \beta_0 + \beta_1\ pluie_{jour} + f_1(température) + f_2(panne) + \sum_{i=1}^{10} \alpha_i\ activité_i + \sum_{i=1}^{7} \gamma_i\ période_i + \epsilon$$

Équation 6-5 : Équation du modèle pour le bus

$$ln(intensité_{bus}) = \beta_0 + \beta_1\ pluie_{jour} + f_1(température) + f_2(panne) + \sum_{i=1}^{6} \alpha_i\ activité_i + \sum_{i=1}^{6} \gamma_i\ période_i + \epsilon$$

Tableau 6-3 : Définition des variables

Variable	Format	Description
$intensité_m$	Numérique continu	Intensité horaire des déplacements du mode m
$pluie_{3h+A}$	Numérique discret	Heures de pluie (3 dernières heures et heure actuelle)
$pluie_{1h+A}$	Numérique discret	Heures de pluie (heure précédente et heure actuelle)
$pluie_{jour}$	Binaire	Présence de pluie durant la journée
$vent$	Numérique continu	Vitesse du vent (km/h)
$température$	Numérique continu	Température (°C)
$panne$	Numérique continu	Durée totale d'interruption de service du métro à la station durant l'heure observée (min)
$activité_i$	Binaire	Présence d'une activité de catégorie i à la station
$période_i$	Binaire	Identifie la période i

Chacune des équations contient des variables significatives portant sur la météorologie, les interruptions de service du métro, les activités et la temporalité. Cela signifie que chacun de ces éléments influence l'achalandage des quatre modes étudiés.

Du côté de la météorologie, la pluie influence chacun des modes, mais pas tout à fait de la même façon. Pour chaque mode, différentes possibilités de variables explicatives ont été testées, soit des variables couvrant différentes périodes horaires dont certaines sont continues ou binaires. La variable sélectionnée pour chaque modèle est celle qui permet de minimiser l'AIC. Pour le vélopartage, la variable sélectionnée est le nombre d'heures de pluie comprises dans les trois heures précédant le début du déplacement et l'heure présente. La sélection de cette variable indique que les usagers du vélopartage semblent baser leur choix modal non seulement sur la météo de l'heure présente, mais aussi sur celle passée. Du côté du taxi, son usage est influencé par l'heure présente et l'heure précédant le départ. Pour le métro et le bus, leur achalandage semble être influencé par la présence de pluie dans la journée entière. Par ailleurs, la vitesse du vent semble uniquement influencer l'usage du vélopartage. La température présente une relation non-linéaire avec l'intensité pour chacun des modes, c'est pourquoi une spline de lissage a été employée.

D'autres variables météorologiques ont aussi été considérées mais n'ont pas été retenues pour le modèle. Par exemple, la présence de neige avait une forte corrélation de -0,42 avec la température. Or, la présence de variables colinéaires aurait rendu difficilement dissociable l'effet de chaque variable. Seule la température a donc été conservée, étant donné qu'elle contribuait davantage au modèle selon l'indicateur AIC. Similairement, l'humidité n'a pas été retenue étant donné qu'elle a une forte corrélation de 0,4 avec la présence de pluie. Par ailleurs, d'autres variables ont été considérées mais ne contribuaient pas significativement au modèle, comme la pression atmosphérique, la présence d'orages et la différence de température par rapport à la normale.

En ce qui concerne les interruptions de service du métro, celles-ci semblent avoir un impact sur l'achalandage de tous les modes dans le voisinage de la station touchée par l'interruption. Cette relation non-linéaire est décrite par une spline de lissage.

Pour les activités, leur présence est indiquée pour chaque catégorie par des variables binaires. Or, certaines catégories d'activités ne semblent pas avoir d'impact significatif sur l'achalandage pour certains modes, étant donné que les modèles du vélopartage du taxi par exemple ne comptent que huit types d'activité sur les dix.

Finalement, les périodes horaires formées précédemment sont incluses dans les modèles à l'aide de variables binaires. La période de référence est la période la plus longue de la journée pour chaque mode, soit de 12h à 15h pour le vélopartage, de 10h à 14h pour le métro, de 9h à 14h pour le bus et finalement le lundi de 10h à 18h pour le taxi.

La figure suivante présente une matrice de corrélation entre les différentes variables sélectionnées pour le vélopartage. Les variables binaires identifiant la période horaire sont exclues pour simplifier la figure et il a été vérifié qu'elles ne présentent aucune colinéarité avec les autres variables.

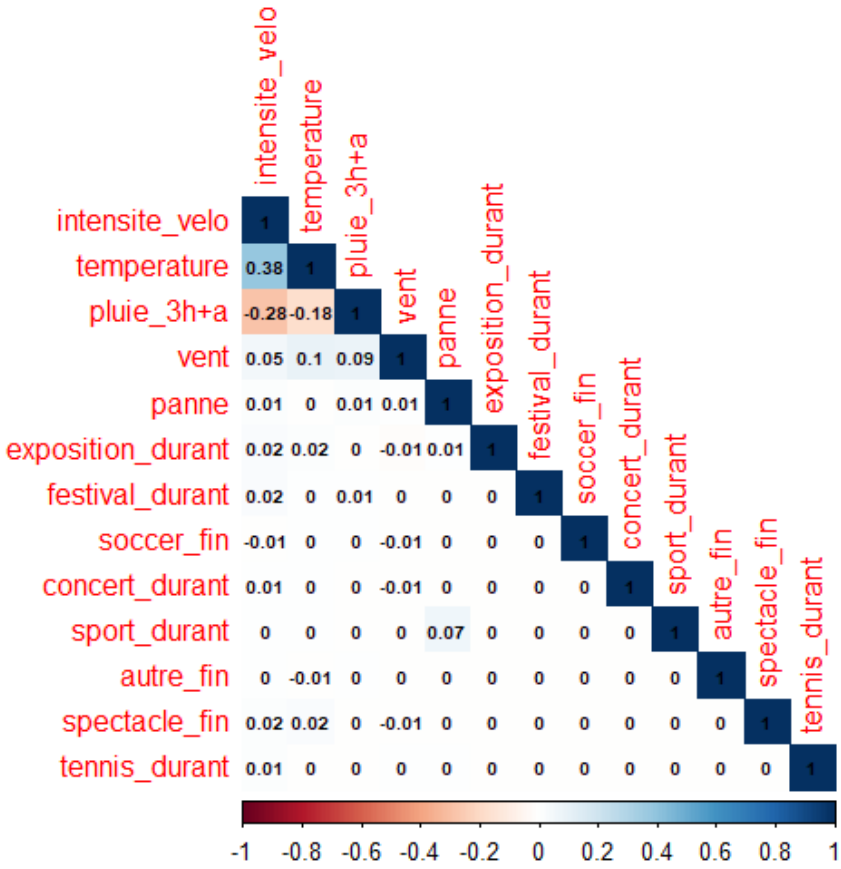

Figure 6-12 : Matrice de corrélation entre les variables sélectionnées pour le vélopartage

Les corrélations les plus fortes sont avec la variable *intensité*$_{vélo}$, soit la variable d'intérêt. La température et la pluie comportent une corrélation négative de 0,18, ce qui n'est pas suffisamment élevé pour les considérer comme colinéaires. Aucune combinaison de variables ne semble colinéaire. La même vérification est effectuée pour les variables sélectionnées pour les modèles des autres modes.

Les coefficients du modèle sont calculés par la méthode des moindres carrés ordinaires. Les valeurs des coefficients sont présentées à l'annexe B pour chaque mode. Ceux-ci sont difficilement interprétables étant donné que la variable dépendante des modèles a été transformée par un logarithme. Pour faciliter leur interprétation, il est possible d'obtenir la variation relative de l'intensité des déplacements liée à chaque variable du modèle en utilisant l'équation suivante.

$$VR_{intensité} = \exp(coef) - 1$$

Le Tableau 6-4 présente la variation relative de l'intensité due à chaque variable. Étant donné que les variables *température* et *panne* ont des relations non linéaires avec la variable d'intérêt, celles-ci sont présentées sur la Figure 6-13 et sur la Figure 6-14. Les variables temporelles liées aux périodes horaires ne peuvent être comparées entre modes étant donné que les périodes varient, c'est pourquoi elles sont uniquement présentées à l'annexe B.

Tableau 6-4 : Variation relative de l'intensité selon chaque variable

Variable	Vélopartage	Taxi	Métro	Bus
$pluie_{3h+A}$	-6,977 %	-	-	-
$pluie_{1h+A}$	-	3,468 %	-	-
$pluie_{jour}$	-	-	-1,144 %	-1,161 %
$vent$	-0,082 %	-	-	-
$température$	voir Figure 6-13			
$panne$	voir Figure 6-14			
$activité_{hockey}$	-	61,0 % →	75,4 % →	-15,8 % ↔
$activité_{exposition}$	9,4 % ↔	20,0 % →	13,6 % ↔	1,7 % ↔
$activité_{festival}$	19,5 % ↔	17,2 % ↔	57,3 % ↔	2,7 % ↔
$activité_{soccer}$	107,9 % →	93,7 % →	120,9 % →	3,6 % ↔
$activité_{concert}$	4,2 % ↔	27,4 % ↔	57,7 % →	-14,3 % ↔
$activité_{sport}$	10,4 % ↔	15,7 % ↔	29,9 % ↔	-
$activité_{autre}$	13,0 % →	-	37,7 % ↔	-
$activité_{football}$	-	-	37,1 % ↔	-
$activité_{spectacle}$	30,5 % →	13,5 % →	58,3 % →	-8,7 % ↔
$activité_{tennis}$	10,6 % ↔	31,9 % ↔	42,0 % →	-
$période_i$	voir l'annexe B			

↔ : Période durant l'activité
→ : Période à la fin de l'activité

Pour le vélopartage, le coefficient lié à la variable $pluie_{3h+A}$ est de -0,07232. La variation relative de l'intensité pour cette variable s'interprète par une baisse de 6,977% de l'intensité des déplacements en vélopartage pour chaque heure de pluie (parmi les 3 heures précédant le déplacement et l'heure actuelle). Cela signifie que s'il pleut durant ces 4 heures, l'usage du vélopartage diminue d'environ 27,9%, soit quatre fois la valeur pour une heure. Cette diminution peut être expliquée par un inconfort pour les cyclistes lorsque la chaussée est mouillée ou que l'humidité est élevée (tel qu'il est généralement le cas après de la pluie). Il est probable que plusieurs cyclistes choisissent une alternative modale ou choisissent de ne pas se déplacer.

Du côté du taxi, la présence de pluie a plutôt un impact positif sur son usage, de l'ordre de 3,5% par heure de pluie (parmi l'heure précédente et l'heure actuelle). Cela pourrait être expliqué par un changement modal des modes actifs vers le taxi.

Pour le métro et le bus, la présence de pluie durant la journée diminue légèrement leur usage de 1,1% et de 1,2% respectivement. L'impact de la pluie sur ces deux modes de transport en commun semble donc être moins important que pour le vélopartage et le taxi.

La vitesse du vent a également un impact négatif sur l'usage du vélopartage, de l'ordre de 0,082% par km/h de vent. Cela signifie qu'une augmentation du vent de 40 km/h entraine une diminution de l'usage d'environ 3,3%. De leur côté, les autres modes ne sont pas être influencés significativement par la vitesse du vent.

La température a une relation non-linéaire avec la variable d'intérêt, c'est pourquoi la relation est décrite par une des splines de lissage par la méthode des plaques minces. Le degré de liberté optimal de la spline est fixé en minimisant l'indice AIC. Pour chaque mode, une spline de lissage est tracée et présentée à la Figure 6-13.

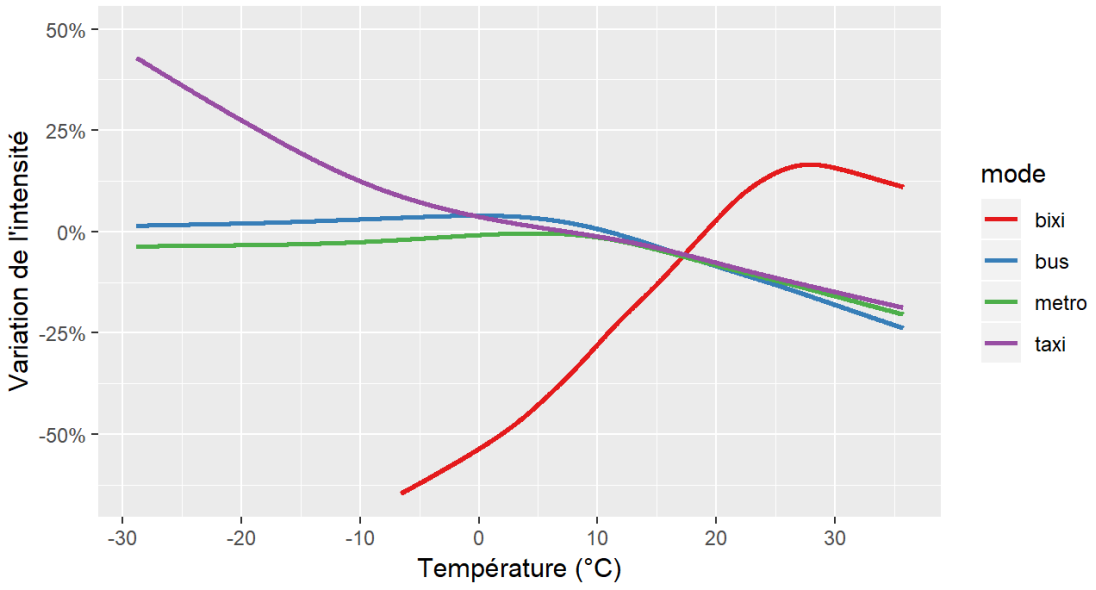

Figure 6-13 : Variation relative de l'intensité selon la température

Tout d'abord, l'intensité des déplacements en vélopartage augmente avec la température, jusqu'à atteindre une intensité maximale à 28°C, après quoi une légère diminution de l'intensité est observée. Cela est expliqué par un inconfort pour les cyclistes lors de températures faibles ou très élevées. De leur côté, l'intensité du bus et du métro présentent des relations similaires. L'intensité est relativement constante sous 5°C, alors qu'elle diminue progressivement pour des températures plus élevées. Cette diminution pourrait être attribuable à une variable non contrôlée par le modèle, soit les vacances d'été durant lesquelles moins de déplacements utilitaires sont effectués. En ce qui concerne le taxi, son intensité maximale est observée à très basse température. L'intensité diminue de façon quasi-linéaire avec l'augmentation de la température. Cette figure laisse croire que le vélopartage serait en compétition avec les autres modes analysés, étant donné que son intensité d'utilisation augmente avec la température alors que c'est le phénomène inverse pour les autres modes. L'importance de ce transfert modal pourrait être estimée à l'aide d'analyses au niveau de l'usager.

Pour continuer, la durée des interruptions de service a également une relation non-linéaire avec l'intensité de la demande, c'est pourquoi des splines de lissage sont utilisées pour décrire cette relation. La Figure 6-14 présente la variation d'intensité selon la durée d'interruption de service. Il est important de rappeler que cette variable désigne la durée totale en interruption de service pour l'heure observée et non le temps écoulé depuis le début d'une panne.

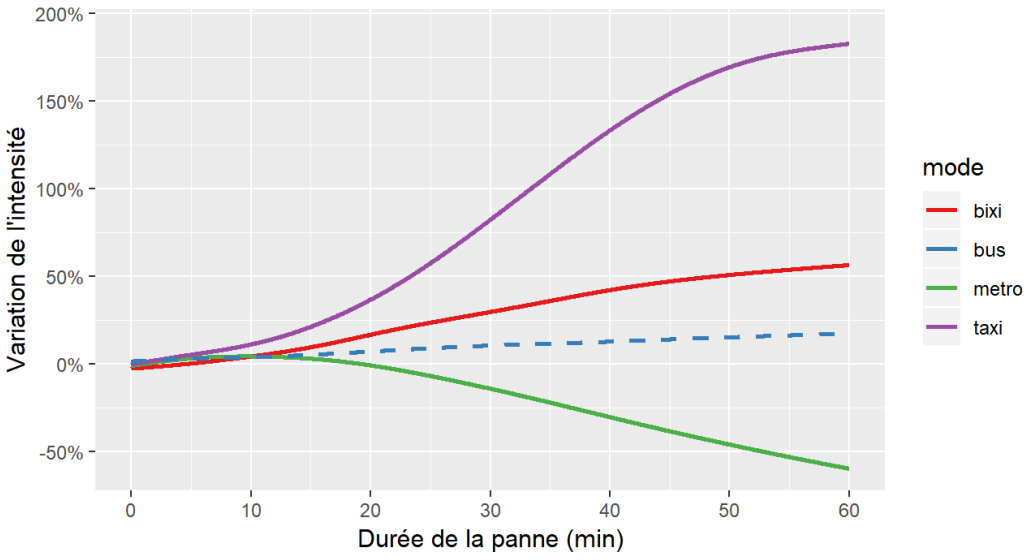

Figure 6-14 : Variation relative de l'intensité selon la durée d'interruption de service

Pour une durée d'interruption faible, les variations d'intensité sont négligeables. Pour le métro, l'intensité d'usage diminue avec l'augmentation de la durée de la panne, jusqu'à atteindre une diminution de l'intensité de 59% lorsque l'heure entière est en interruption de service. Étant donné que la validation de cartes à puce se fait à l'entrée du réseau de métro et non à l'embarquement dans le véhicule, l'intensité d'usage observée compte les usagers qui sont entrés dans le réseau de métro mais qui après avoir constaté la panne en sont ressortis sans entrer dans le train. Cela signifie que l'intensité des déplacements réellement complétés en métro est inférieure à celle observée ici. Pour des durées de pannes plus élevées, la panne est généralement médiatisée ce qui contribue à diminuer les validations de cartes à puce, étant donné que les usagers peuvent choisir des alternatives modales avant même d'être arrivé à l'entrée du métro.

Lors d'une panne de métro, plusieurs usagers considèrent des modes alternatifs afin de compléter leur déplacement plus rapidement. L'intensité d'usage des autres modes augmente donc, tel qu'observé sur la Figure 6-14. Cette augmentation n'est pas linéaire. Celle-ci est plutôt faible pour des courtes durées de panne, puis augmente rapidement pour des durées entre 20 à 45 minutes, et finalement s'adoucit pour des durées approchant les 60 minutes. Pour des durées de pannes faibles, l'augmentation de l'intensité est peu importante puisque les usagers espèrent la reprise rapide du service. Lorsque la panne perdure, plusieurs utilisateurs décident de se tourner vers d'autres modes de transport.

Pour le taxi, une panne de métro s'étendant sur l'heure entière entraîne une augmentation d'achalandage d'environ 182% dans le voisinage de la station de métro en interruption. Du côté du vélopartage, cette augmentation est plutôt de l'ordre de 55%. L'intensité maximale du vélopartage semble tendre vers une constante, ce qui pourrait être expliqué par la capacité limitée des stations en termes de nombre de vélos disponibles.

Pour le bus, son usage lors d'interruptions de service est sous-évalué, dû à l'omission des données du service provisoire par bus. En effet, la Société de transport de Montréal (STM) ajoute du service temporaire par bus lors d'interruptions de service de longue durée, passant par les stations de métro en interruption. Or, les validations de cartes à puce à bord de ces bus temporaires n'ont pas pu être identifiées au trajet emprunté par les bus et ont donc été omises. De plus, une pratique répandue des chauffeurs lors de pannes de métro consiste à accélérer l'embarquement des usagers en évitant la perception. Le chauffeur place sa main sur le lecteur de cartes à puce et fait signe aux usagers d'entrer sans payer. Cette même pratique est faite sur les lignes régulières de bus ainsi que sur le service provisoire par bus. Il devient donc impossible de connaitre le nombre d'embarquements à partir des données de cartes à puce. L'augmentation d'intensité de 20% observée sur la Figure 6-14 est donc largement sous-évaluée.

Par ailleurs, des variables pour les activités sont présentes pour les catégories d'activités dont l'effet est significatif. Les variations d'intensité présentées au Tableau 6-4 sont majoritairement positives, ce qui est attendu étant donné que les activités sont des lieux d'attraction temporaires qui génèrent des déplacements localement. Or, certains types d'activité ont un impact plus important que d'autres.

Dans le Tableau 6-4, les flèches à côté des valeurs identifient la période pour laquelle l'activité a été considérée, ayant été choisie en minimisant l'indicateur AIC du modèle. En général, les activités de plus longue durée considèrent majoritairement la période durant l'activité, étant donné que plusieurs participants peuvent quitter l'activité avant sa fin. En effet, les activités de catégories exposition, festival, sport, autre et tennis ont des durées médianes de 6 heures et plus (voir les durées médianes par catégorie au Tableau 3-10) et leur période considérée est majoritairement celle durant l'activité. De leur côté, les activités de courte durée considèrent majoritairement la période de fin d'activité, mais de façon moins marquée. Cela peut être expliqué par la similarité entre les deux périodes considérées lorsque l'activité est de courte durée. En effet, pour une activité d'une

durée de 2 heures par exemple, la période durant l'activité s'étend du début jusqu'à une heure après la fin, alors que la période de la fin s'étend d'une heure avant la fin de l'activité jusqu'à une heure après la fin, ce qui ne représente qu'une seule heure de différence entre les deux périodes.

La figure suivante présente la variation d'intensité selon la catégorie d'activité et le mode. Il s'agit des mêmes valeurs que celles présentées au Tableau 6-4, mais présentées de façon à faciliter leur comparaison.

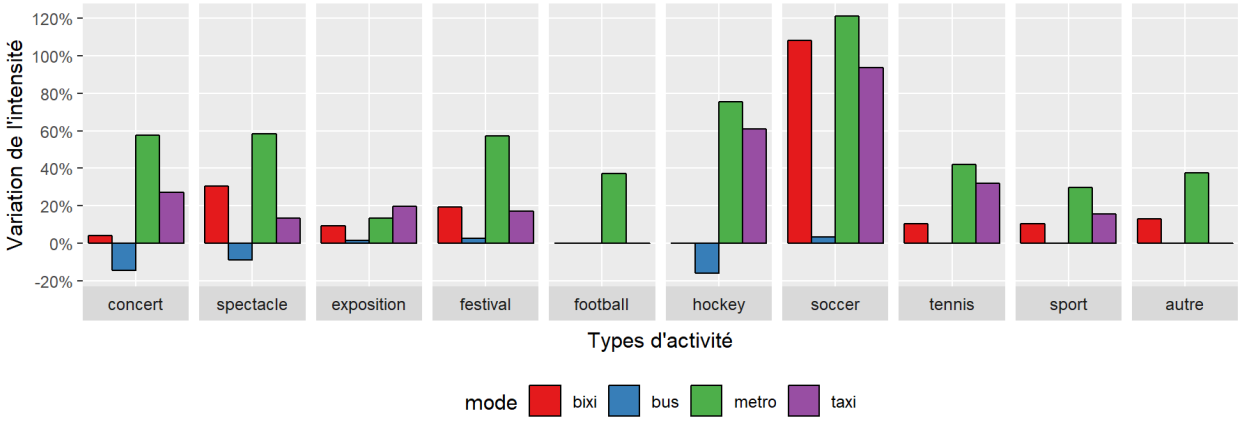

Figure 6-15 : Variation de l'intensité des départs selon l'activité et le mode

Pour le vélopartage, la catégorie d'activités qui entraine l'augmentation d'intensité la plus importante est la catégorie soccer, avec une augmentation de près de 107,9%. Elle est suivie par la catégorie spectacle qui présente une augmentation de 30,5%. Les activités de catégorie hockey ne semblent pas affecter significativement le vélopartage, ce qui peut être expliqué par le court chevauchement de cette activité avec la saison de vélopartage à l'automne, où la température devient fraiche ce qui devient moins agréable pour les cyclistes. Il est important de mentionner que la disponibilité de vélos aux stations de vélopartage n'est pas considérée. Il n'est donc pas possible d'observer la demande latente, soit les déplacements qui auraient été faits si plus de vélos avaient été disponibles.

Du côté du taxi, les deux catégories d'activité avec les hausses d'achalandage les plus grandes sont le soccer et le hockey, avec des augmentations respectives de 93,7% et 61,0%. Les catégories d'activités entraînent toutes une augmentation de l'intensité d'usage, excepté pour les activités de catégorie football et autre.

Pour le métro, le soccer est la catégorie d'activités qui entraîne la plus forte hausse d'achalandage de 120,9%. Elle est suivie par les activités de catégorie hockey avec une augmentation d'environ 75,4%. Le métro est le seul mode pour lequel l'ensemble des 10 catégories d'activités influencent significativement la demande.

Les résultats du modèle pour le bus sous-évaluent l'intensité réelle des déplacements lors d'activités, étant donné que les données considérées n'incluent pas les services de bus pour évènements spéciaux ajoutés au service régulier, mais seulement les bus réguliers. Or, les déplacements effectués avec ces services de bus n'ont pas été utilisées pour cette étude. Ce modèle évalue donc l'effet des activités sur les bus réguliers plutôt que sur l'ensemble des services de bus. Par ailleurs, il est important de rappeler que les embarquements sur l'ensemble des lignes de bus passant par la station de métro sont comptabilisés, ce qui atténue l'impact local des activités. Bref, les résultats obtenus pour le bus sont difficilement interprétables dû à plusieurs limitations.

La Figure 6-16 présente les résidus du modèle pour le vélopartage par rapport aux valeurs prédites. Une spline de lissage a été ajoutée en rouge afin de mieux voir la tendance. Les points semblent distribués uniformément autour d'un résidu nul et ne semblent suivre aucun patron, ce qui semble indiquer que la transformation logarithmique est adéquate. La Figure 6-17 présente le graph quantile-quantile (*qq-plot*) des résidus. Celui-ci s'apparente à une droite, confirmant que les résidus sont normalement distribués.

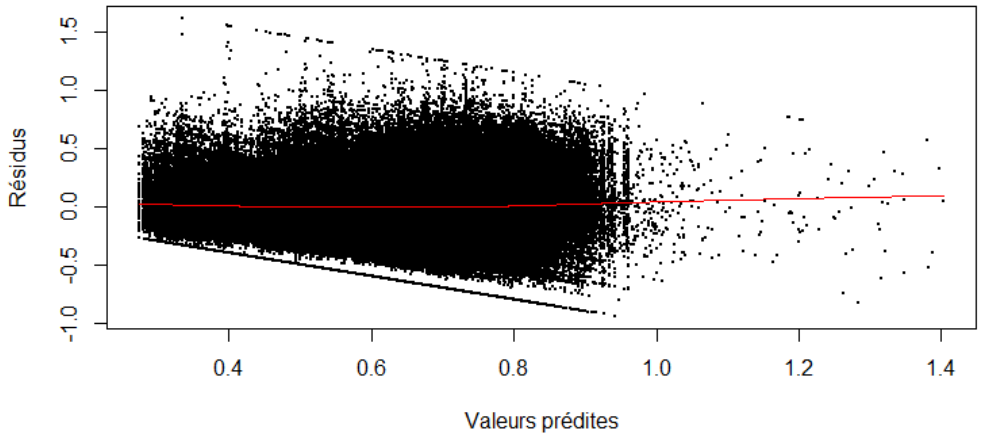

Figure 6-16 : Résidus selon les valeurs prédites, pour le modèle du vélopartage

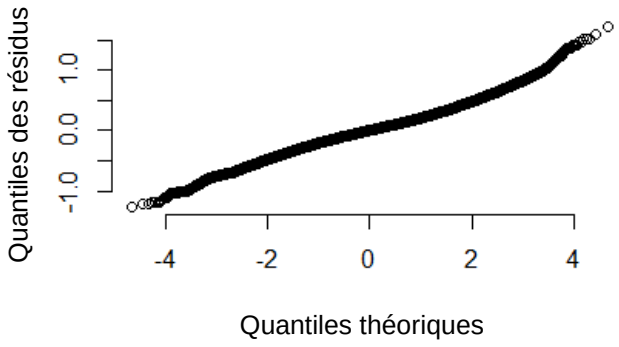

Figure 6-17 : Graphe quantile-quantile des résidus

La figure suivante compare les valeurs prédites aux valeurs réelles pour le modèle du vélopartage. Les valeurs prédites se rapprochent de la droite indiquant la prévision parfaite. La concentration des points autour de la droite est très élevée et semble diminuer en s'y éloignant. Les points les plus éloignés de la droite ont un levier important. Il pourrait être pertinent d'analyser ceux-ci pour s'assurer qu'ils ne correspondent pas à des valeurs aberrantes.

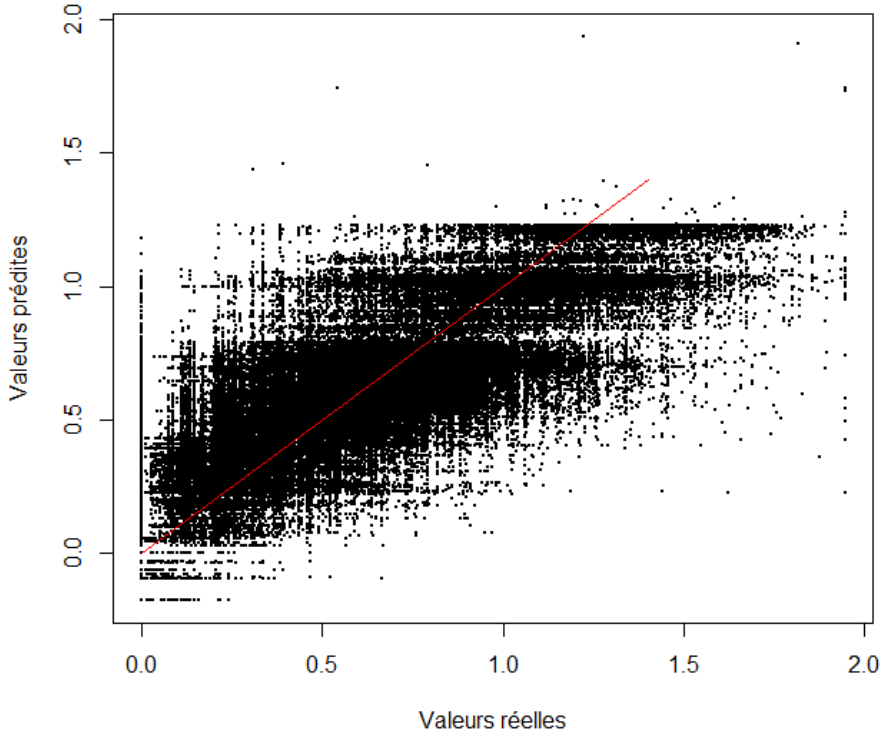

Figure 6-18 : Valeurs prédites par rapport aux valeurs réelles, pour le modèle du vélopartage

Le tableau suivant présente le critère d'information d'Akaike pour les quatre modèles. Cet indicateur à minimiser est utile afin de juger de la qualité de l'ajustement d'un modèle. Il n'est

cependant pas possible de comparer les indicateurs des différents modèles entre eux, étant donné que ces modèles ont des variables d'intérêt différentes.

Tableau 6-5 : Critère d'information d'Akaike (AIC) des différents modèles

Variable	AIC
Vélopartage	- 36 952
Taxi	- 7 576
Métro	- 115 473
Bus	- 1 060 992

6.2 Modèles par type de station

L'objectif de cette section est de mieux comprendre l'influence de certaines variables météo sur l'usage des cinq types de station identifiés dans la typologie à la section 5.1. Les variables explicatives météorologiques sont celles ayant été sélectionnées à la section précédente, soit pour le modèle sur l'ensemble des stations. Les variables explicatives concernant les activités et les interruptions de service ne peuvent être intégrées aux modèles par type de station, car elles sont spécifiques à une station et non à un type de station.

Pour chaque type de station, le nombre moyen de déplacements horaires est calculé avant de déterminer l'intensité des déplacements avec l'Équation 6-1. Cela permet d'obtenir une série temporelle continue pour chaque type de station, ce qui permet l'utilisation de modèles pour séries temporelles. Deux types de modèles sont comparés, soit le modèle additif généralisé (GAM) et le modèle ARIMA, soit un modèle pour séries temporelles.

Les données du vélopartage sont utilisées comme exemple, mais cette méthode pourrait aussi être appliquée aux autres modes.

6.2.1 Modèles GAM

Les coefficients résultants des modèles additifs généralisés sont présentés à l'Annexe C. Or, les constats importants sont repris ici. Tout d'abord, les variations relatives de l'intensité d'usage pour la pluie et le vent sont présentées au Tableau 6-6.

Tableau 6-6 : Variation relative de l'intensité selon la pluie et le vent, pour le vélopartage

Type de station	$pluie_{3h+A}$	$vent$
Type 1	-6,97 %	-0,055 %
Type 2	-7,02 %	-0,080 %
Type 3	-7,30 %	-0,092 %
Type 4	-6,12 %	-0,099 %
Type 5	-7,64 %	-0,093 %

Toutes les variations d'intensité sont négatives, ce qui indique une diminution de l'usage du vélopartage en présence de pluie pour tous les types de station. Or, les stations près du centre-ville

semblent moins affectées par la pluie que celles qui s'en éloigne, excepté pour le type 4 qui semble peu affecté par la pluie. De façon similaire, les stations près du centre-ville semblent moins affectées par la vitesse du vent.

La Figure 6-19 présente les splines de lissage qui décrivent la relation entre la température et l'intensité d'usage pour chaque type de station.

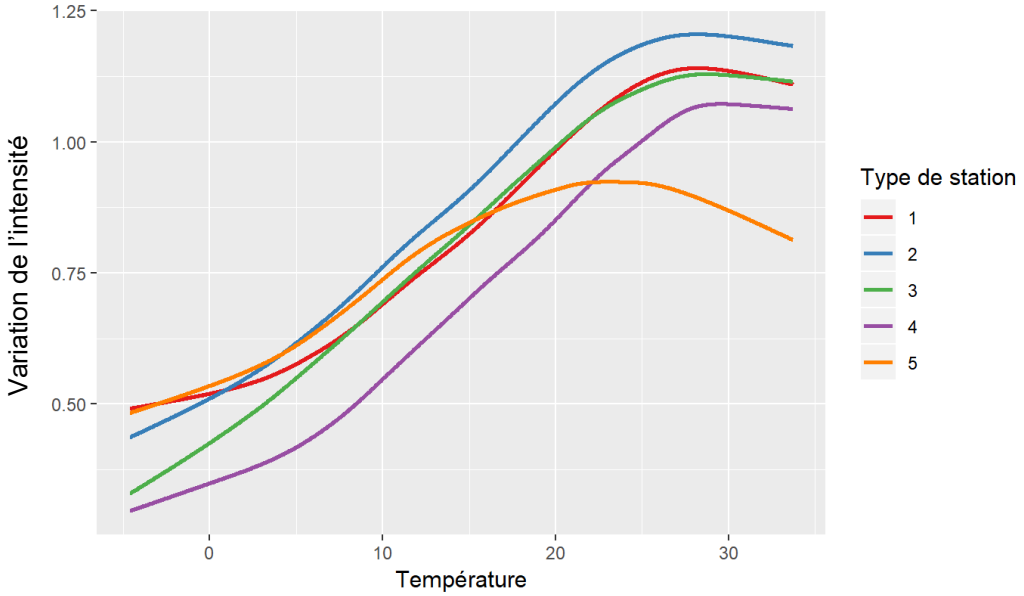

Figure 6-19 : Intensité des déplacements selon la température par type de station, pour le vélopartage

Lorsque la température est très faible, l'intensité la plus élevée est pour les stations de type 1, 2 et 3, soit des stations autour desquelles la demande utilitaire est importante. Pour des températures plus élevées, l'intensité des stations de type 5 est plus faible que pour les autres stations, ce qui peut être expliqué par la période de vacances durant l'été pour les étudiants, étant donné que ce type de station présente beaucoup d'institutions scolaires à proximité.

6.2.2 Modèles ARIMA

Le modèle ARIMA est un modèle autorégressif qui peut prendre en compte une saisonnalité (dans ce cas le cycle journalier) et des variables externes (dans ce cas des variables météorologiques). L'interprétation des résultats est très similaire au modèle linéaire, étant donné qu'il s'agit aussi d'un modèle additif. La fonction « auto.arima » de la librairie « forecast » de R est utilisée pour calibrer le modèle en minimisant l'AIC (Hyndman et Khandakar, 2007).

Le modèle ARIMA se différencie principalement du modèle GAM par sa prise en compte de coefficients autorégressifs. Ces coefficients autorégressifs peuvent être avantageux afin d'offrir une plus grande flexibilité et ainsi de mieux capturer les variations d'intensité d'usage selon le temps. Cependant, une plus grande flexibilité peut potentiellement causer du surapprentissage si l'échantillon est trop faible.

Le modèle ARIMA présente certaines limitations, tel que la prise en compte d'une seule saisonnalité. Le cycle hebdomadaire ne peut donc pas être pris en charge dans le même modèle que le cycle journalier. Par ailleurs, le modèle ARIMA ne permet pas la calibration de splines de lissage pour certaines variables tel qu'il a été fait avec le modèle GAM. La variable température sera donc supposée linéaire pour la calibration de ce modèle, malgré le fait qu'il s'agisse d'une simplification qui pourrait entrainer une perte de précision.

Les coefficients résultant de la calibration du modèle ARIMA pour chaque type de station sont présentés à l'annexe C. Tous les modèles calibrés par la fonction « auto.arima » présentent des coefficients autorégressifs, ce qui semble indiquer une corrélation entre les heures consécutives. Les coefficients « ar1 » sont tous positifs, ce qui indique que la corrélation avec l'heure précédente est positive. Chaque type de modèle présente également des coefficients autorégressifs saisonniers, indiquant la présence d'une corrélation avec l'intensité d'usage à la même heure le jour précédent.

6.2.3 Comparaison des modèles

Le Tableau 6-7 compare les résultats du modèle GAM à ceux du modèle ARIMA pour la variable $pluie_{3h+A}$ pour le vélopartage. Tous les coefficients sont négatifs. Ceux du modèle ARIMA sont moins importants, mais sont du même ordre de grandeur. La même conclusion peut être tirée pour les 2 modèles, soit que la pluie a une moins grande incidence sur les stations près du centre-ville.

Tableau 6-7 : Variation relative de l'intensité selon la variable $pluie_{3h+A}$, pour le vélopartage

Type de station	Modèle GAM	Modèle ARIMA
Type 1	-6,97 %	-4,70 %
Type 2	-7,02 %	-4,91 %
Type 3	-7,30 %	-5,70 %
Type 4	-6,12 %	-5,51 %
Type 5	-7,64 %	-5,80 %

Pour continuer, la Figure 6-20 compare les valeurs prédites par les deux modèles aux valeurs réelles. Une semaine de printemps 2016 a été choisie à titre d'exemple, soit une semaine qui comporte deux évènements de pluie. Il est possible d'observer que l'intensité prédite par les 2 modèles suit relativement bien l'intensité réelle. Des différences plus grandes sont observées pour les heures de pointe et également pour le vendredi soir où la demande est fort probablement influencée par différentes activités sociales. Les deux évènements de pluie introduisent une diminution de l'intensité d'usage que les modèles expliquent bien.

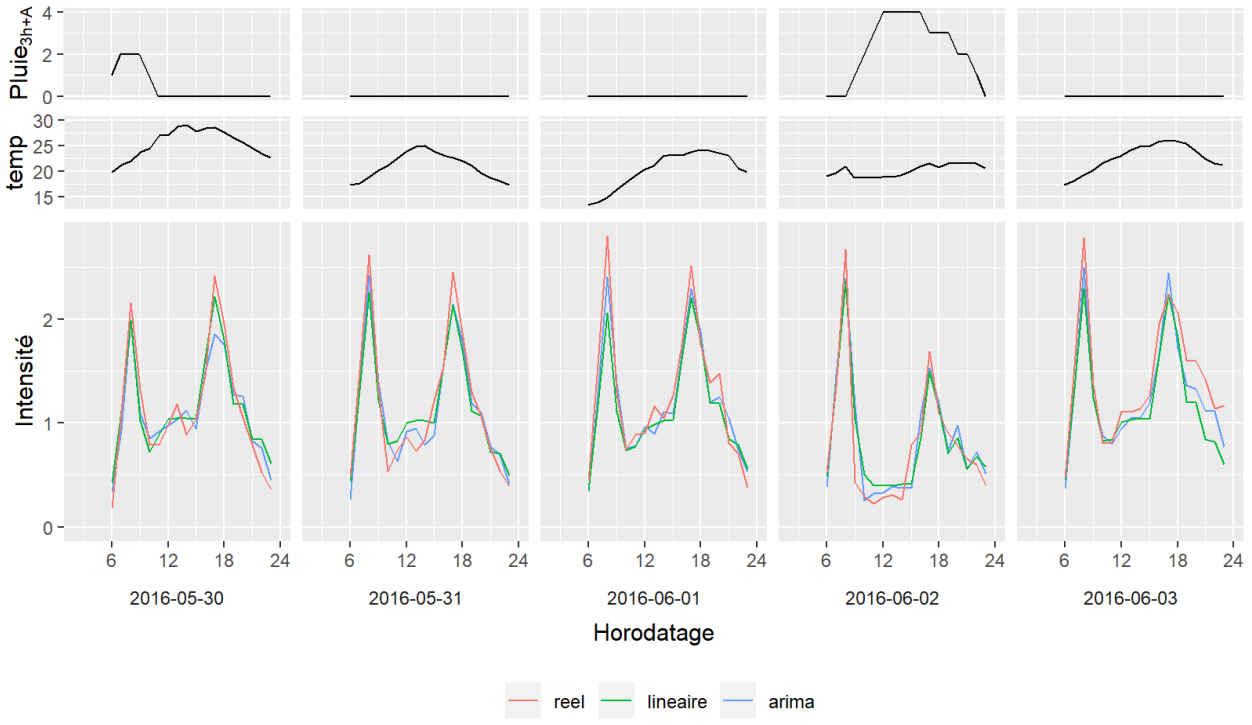

Figure 6-20 : Exemple de prévision de l'intensité pour les stations de type 3, pour le vélopartage

Le Tableau 6-8 compare le critère d'information d'Akaike (AIC) pour les deux types de modèle. Pour tous les type de station, le modèle ARIMA semble mieux performer que le modèle GAM étant donné que le critère AIC est plus faible.

Tableau 6-8 : AIC des modèles selon le type de station

Type de station	Modèle GAM	Modèle ARIMA
Type 1	-10405	-14797
Type 2	-12100	-15844
Type 3	-10805	-14102
Type 4	-1363	-2150
Type 5	-2720	-4591

Pour continuer, la performance des deux modèles est évaluée par validation croisée. Les deux premières années de données sont utilisées afin d'entrainer les modèles et la dernière année est utilisée comme test. Le Tableau 6-9 compare la racine de l'erreur moyenne quadratique (RMSE) pour les deux types de modèle. La variable prédite a été transformée dans l'unité originale afin de calculer l'erreur sur l'intensité d'usage, plutôt que sur le logarithme de l'intensité.

Tableau 6-9 : Erreur RMSE par type de modèle, pour le vélopartage

Type de station	Erreur d'entrainement		Erreur test	
	Modèle GAM	Modèle ARIMA	Modèle GAM	Modèle ARIMA
Type 1	0,240	0,202	0,262	0,304
Type 2	0,195	0,158	0,209	0,224
Type 3	0,222	0,173	0,238	0,383
Type 4	0,524	0,498	0,510	0,596
Type 5	0,450	0,389	0,461	0,463

Pour la majorité des modèles, l'erreur sur les données de test est plus élevée que l'erreur d'entrainement, ce qui est attendu. Les erreurs d'entrainement sont plus faibles pour les modèles

ARIMA que pour les modèles GAM, ce qui est attendu étant donné que le modèle ARIMA est plus flexible. Cependant, les erreurs test sont plus élevées pour les modèles ARIMA, ce qui est probablement causé par du surapprentissage. Le modèle GAM semble donc plus approprié pour décrire ces données.

6.3 Modèles par station

Différents modèles sont calibrés pour chaque station, afin de comparer l'influence des différents évènements sur l'usage de chaque station. La sélection des variables est la même que pour le modèle portant sur l'ensemble des stations. Cependant, les interruptions de service sont exclues étant donné que l'échantillon de données pour plusieurs stations est trop faible pour une calibration adéquate. Les variables binaires indiquant la présence d'activités sont uniquement incluses pour les stations qui présentent au minimum une activité de la catégorie concernée. L'analyse utilise les données du vélopartage à titre d'exemple, mais la méthode peut également être appliquée à d'autres modes.

6.3.1 Modèles descriptifs

Tout d'abord, des modèles GAM sont calibrés pour chaque station. Les résultats pour la variable $pluie_{3h+A}$ sont présentés à la Figure 6-21.

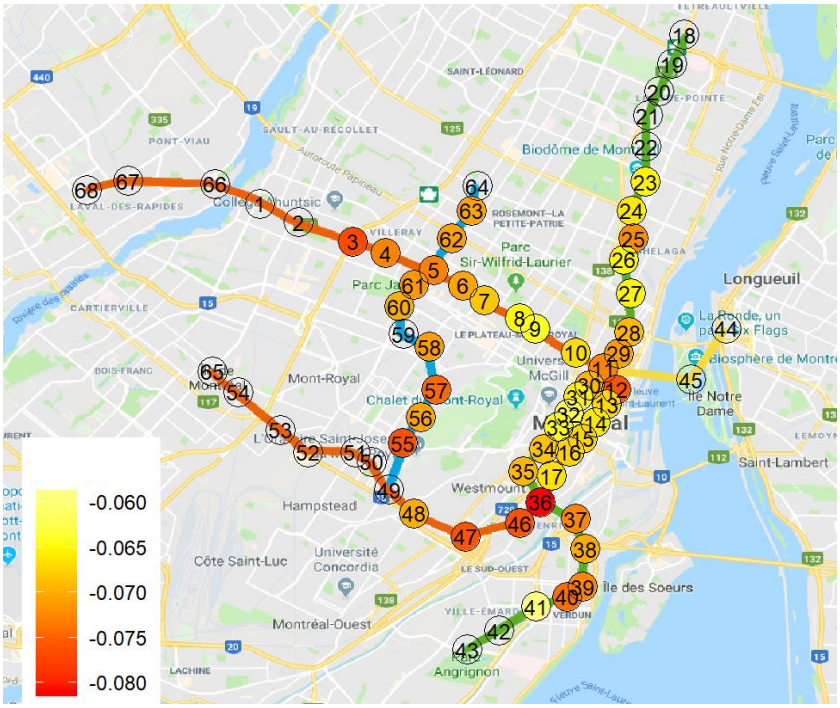

Figure 6-21 : Variation relative de l'intensité pour le vélopartage selon la variable $pluie_{3h+A}$

Les stations près du centre-ville voient leur intensité d'usage diminuer d'environ 6% (pour chaque heure de pluie comprise dans les 3 heures précédant le départ et l'heure du départ), alors que les stations en périphérie observent une diminution plus importante d'environ 7,5%. Une explication possible est que les déplacements près du centre-ville sont généralement plus courts, ce qui rend la pluie moins dérangeante pour les cyclistes.

La variation d'intensité pour la station 12 (Champ-de-Mars) de 7,7% est plus faible que pour les autres stations à proximité. Cela peut être attribué à un usage récréatif important de cette station, dû à plusieurs attractions touristiques à proximité. En effet, un usage récréatif est généralement plus sensible aux conditions météorologiques que l'usage utilitaire comme c'est le cas pour la majorité des stations près du centre-ville (Brandenburg et al., 2007; Richardson, 2000; Thomas et al., 2008).

Un modèle ARIMA est entrainé sur les mêmes données pour chaque station, afin de comparer les résultats avec le modèle GAM. La Figure 6-22 compare les valeurs prédites par les deux modèles aux valeurs réelles, avec une semaine du printemps 2016 à titre d'exemple.

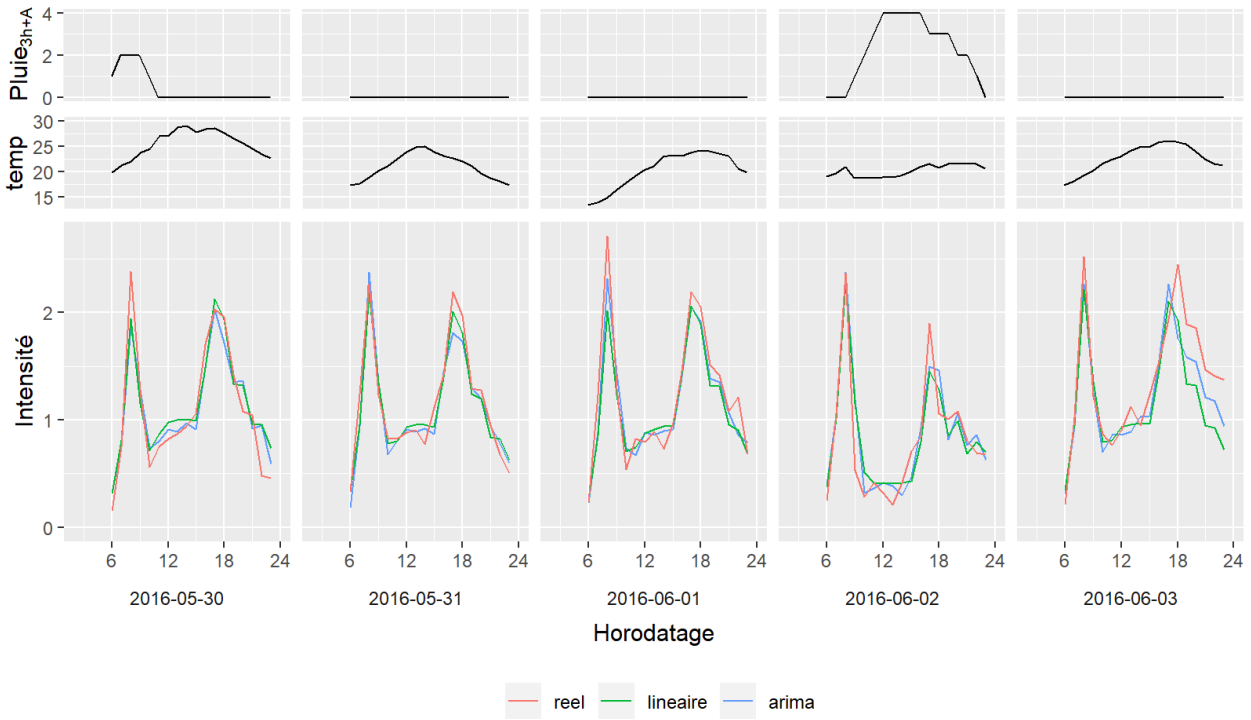

Figure 6-22 : Exemple de prévision de l'intensité pour le vélopartage à la stations Laurier

Il est possible d'observer que les modèles prédisent généralement bien l'intensité d'usage réelle. Or, certaines fluctuations de l'intensité d'usage ne sont pas capturées par les modèles, comme par exemple la pointe du matin plus importante qu'à l'habitude le 1er juin 2016 ou l'intensité d'usage élevée du vendredi soir. Le modèle ARIMA semble un peu mieux ajusté aux données réelles, mais les deux modèles demeurent très similaires.

6.3.2 Modèles prédictifs

Des modèles prédictifs peuvent être utilisés pour estimer l'intensité d'usage en se basant sur les données passées et les données d'évènements. Quatre horizons de prévision sont évalués, soit pour une durée d'un mois, d'une semaine, d'un jour et d'une heure.

La validation-croisée est utilisée afin de comparer les performances des modèles GAM et ARIMA sur ces différents horizons. Une fenêtre glissante d'une durée de deux mois définit l'étendu des données d'entrainement. Itérativement, des modèles sont entrainés sur deux mois de données horaires pour prédire l'intensité pour la période subséquente. Cette fenêtre de deux mois est suffisamment large pour apprendre l'effet de la majorité des évènements, tout en étant suffisamment courte pour limiter la durée de calcul des modèles.

Le Tableau 6-10 présente l'erreur obtenue par validation-croisée pour l'intensité des déplacements en vélopartage pour la station Mont-Royal (station 9) à titre d'exemple. Le type d'erreur calculé est la racine de l'erreur moyenne quadratique (RMSE).

Tableau 6-10 : Erreur (RMSE) par validation croisée pour le vélopartage à la station Mont-Royal

Horizon de prévision	Modèle GAM	Modèle ARIMA
1 mois	0,2378	0,2444
1 semaine	0,2163	0,2361
1 jour	0,2250	0,2358
1 heure	0,2139	0,1928

Pour des prévisions à plus long terme, le modèle GAM semble mieux performer que le modèle ARIMA. La flexibilité du modèle ARIMA semble mener à du surentrainement étant donné l'erreur plus élevée. Cependant, pour des prévisions à court terme, le modèle ARIMA performe mieux que le modèle GAM. Le modèle ARIMA inclut des termes autorégressifs qui permettent d'ajuster la prévision en fonction des valeurs récentes d'intensité d'usage. Cela peut être utile pour compenser pour certains évènements qui ne sont pas pris en compte par les variables explicatives, tel que certaines activités, de la congestion routière ou l'indisponibilité de vélos aux stations de vélopartage. Par conséquent, le modèle GAM performe mieux sur des horizons longs, alors que le modèle ARIMA performe mieux sur des horizons plus courts.

CHAPITRE 7 CONCLUSION

Cette section présente tout d'abord une synthèse de la recherche, puis énonce les contributions apportées, les limitations et les perspectives de l'étude.

7.1 Synthèse de la recherche

Cette étude permet de mettre en évidence les corrélations d'usage entre différents modes de transport en utilisant des données transactionnelles pour quatre différents modes de transport. Les résultats analytiques obtenus sont propres au territoire de Montréal, par son climat, son contexte géographique et la culture de sa population. Toutefois, les différentes méthodes proposées par l'études pourraient être appliquées à des données provenant de d'autres villes. Il est attendu que les résultats varient non seulement selon le territoire étudié, mais également selon le temps.

La première analyse porte sur les partons typiques d'achalandage pour chaque combinaison de mode et de station. Les patrons considérés sont les patrons annuels, hebdomadaires et horaires. Des vecteurs sont créés pour chaque combinaison de mode et de station. Les nombres de déplacements sont normalisés selon le mode et la station, afin de permettre leur comparaison malgré qu'ils soient d'ordres de grandeur très différents. Les vecteurs similaires sont regroupés en utilisant la méthode de partitionnement de données des k-moyennes.

Les patrons annuels du métro et du bus sont très similaires, soit un usage plutôt constant durant l'année, excepté pour la période de vacances estivales et celle des fêtes lors desquelles l'usage est plus faible. Pour le vélopartage, son achalandage augmente progressivement au printemps avec l'augmentation de la température, puis diminue à l'automne. Du côté du taxi, son usage se distingue par une diminution progressive au cours de l'année. Les vecteurs du vélopartage et du taxi sont tous deux uniques, c'est pourquoi les groupes dans lesquels ils sont classés comprennent presque uniquement des vecteurs d'un seul mode. Finalement, il semble y avoir une relation de compétitivité entre le vélopartage et les autres modes, étant donné que la période d'usage maximale du vélopartage (durant l'été) correspond à une période de plus faible usage pour les autres modes.

Les patrons hebdomadaires pour le métro et le bus sont généralement attribués aux mêmes groupes, soit à des groupes ayant un usage plus élevé la semaine que la fin de semaine. En ce qui concerne le vélopartage, son usage est aussi élevé la fin de semaine que la semaine au milieu de l'été. Par ailleurs, le taxi présente un usage maximal le vendredi. Les stations situées dans la partie Ouest du

centre-ville et près des universités ont des patrons qui se distinguent par un achalandage très faible la fin de semaine.

Les patrons journaliers pour les jours de semaine ont typiquement des périodes de pointe le matin et le soir, alors que l'achalandage est plus constant pour les jours de fin de semaine. Cependant, le taxi a un usage différent dont les périodes de pointe sont peu marquées et dont l'usage en soirée est plus élevé que les autres modes. Pour le métro, les stations près des pôles d'activités sociales ont des périodes de pointe plus importantes le soir, alors que pour les autres stations la période de pointe du matin est la plus importante. Le vélopartage a une pointe du soir plus importante que celle du matin et plus tardive que celle des autres modes. Du côté du taxi, son usage est important le soir, particulièrement pour les stations à l'Est du centre-ville. Par ailleurs, il semble y avoir une relation de complémentarité entre le métro et le bus, étant donné que l'usage du bus augmente de façon importante lors de la période de fermeture du métro (durant la nuit).

La seconde analyse a comme objectif de créer une typologie de stations de métro basée sur les caractéristiques du voisinage des stations, afin de mieux comprendre leur usage. Les stations sont classées dans 5 groupes selon un algorithme hiérarchique de partitionnement de données. Le premier groupe comprend les stations situées au centre-ville comportant un nombre important d'emplois à proximité, alors que le deuxième groupe comprend des stations situées très près du centre-ville et qui ont une densité d'emplois légèrement plus faible. Le troisième groupe comprend des stations ayant une densité de logement élevée et le quatrième groupe comprend des zones résidentielles de plus faible densité. Finalement, le groupe 5 comprend trois stations près des différents campus de l'Université de Montréal.

Les stations sont également classées selon leur patron type d'usage journalier et ce, pour les jours ouvrables. Cette classification est comparée à la typologie de stations basée sur les caractéristiques des stations, afin de déterminer s'il est possible de prédire le type d'usage des stations à partir des caractéristiques de celles-ci. Les groupes similaires pour chacune des deux méthodes de classification sont associés. La typologie basée sur les caractéristiques permet donc d'identifier le bon type d'usage pour 64% des stations.

La troisième analyse s'intéresse à la fluctuation de l'achalandage due à divers évènements tels que des évènements météorologiques, des interruptions de service du métro et des activités. Tout d'abord, un modèle de type GAM est développé pour chaque mode, afin d'identifier les variables

ayant un impact significatif sur l'usage. Les résultats suggèrent que la présence de pluie diminue l'usage du vélopartage, du métro et du bus, alors qu'elle augmente l'usage du taxi. L'usage du vélopartage semble aussi être affecté par la présence de pluie dans les trois heures précédant le déplacement. L'achalandage du vélopartage augmente avec la température, alors qu'il diminue pour le taxi. La relation entre la température et l'intensité d'usage est non linéaire, c'est pourquoi elle est prise en compte dans le modèle à l'aide d'une spline de lissage. Pour continuer, l'achalandage augmente de manière significative durant des interruptions de service du métro pour le vélopartage, le taxi et le bus. Pour le taxi, cette augmentation atteint 182% après une heure complète en interruption de service. Par ailleurs, les activités influencent l'usage des quatre modes, mais le mode qui semble le plus affecté par celles-ci est le métro.

Ensuite, des modèles sont calibrés pour chaque type de station ayant été défini dans la typologie de station de métro. Deux types de modèles descriptifs sont comparés, soit le modèle GAM et le modèle ARIMA. Les coefficients obtenus sont du même signe et du même ordre de grandeur. La performance de ces modèles est estimée par validation croisée. Les erreurs test sont plus faibles pour le modèle GAM, ce qui semble indiquer que ce type de modèle est plus approprié.

Finalement, des modèles sont calibrés pour chaque station de métro. Les résultats des modèles descriptifs suggèrent que la pluie a un impact plus important sur les stations plus éloignées du centre-ville. De plus, les stations proposant des activités touristiques dans leur voisinage semblent davantage affectées par la météo. Pour continuer, l'intensité d'usage est prédite sur différents horizons à l'aide de modèles prédictifs de type GAM et ARIMA. La performance de ceux-ci est comparée par validation croisée. Le modèle ARIMA semble mieux performer pour des horizons de prévision courts, alors que le modèle GAM semble mieux performer pour des horizons plus longs. La moins bonne performance du modèle ARIMA sur des horizons de prévision plus longs est probablement attribuable à une flexibilité du modèle trop élevée menant à du surapprentissage.

7.2 Contributions

Cette étude permet des apports méthodologiques en plus des contributions au niveau des résultats déjà mentionnées dans la section précédente. Tout d'abord, la méthodologie proposée permet l'analyse simultanée de quatre différents modes de transport en utilisant des données transactionnelles, ce qui a très peu été exploré dans la littérature. La définition de zones de

voisinage des stations de métro permet de simplifier la dimension spatiale de l'étude. Ensuite, la méthode de normalisation proposée permet de comparer des nombres de déplacements d'ordre de grandeurs très différents.

Par ailleurs, les effets de plusieurs types d'évènements atypiques (météo, activités, pannes de métro) sur l'achalandage sont simultanément évalués dans un même modèle, alors que les autres études consultées ne prennent en compte qu'un seul type d'évènement. Deux types de modèles sont comparés, soit un modèle additif généralisé (GAM) et un modèle ARIMA. Dans le modèle GAM, la relation entre la température et l'achalandage est exprimée par une spline de lissage qui offre davantage de flexibilité par rapport à la relation linéaire généralement utilisée par les études consultées.

Cette étude permet de mieux comprendre les habitudes de déplacement de la population montréalaise. Tout d'abord, l'analyse des patrons type d'usage permet de mieux comprendre les corrélations d'usage entre les différents modes et ainsi évaluer si la relation entre deux modes semble complémentaire ou plutôt compétitive. Ensuite, la typologie de stations de métro propose une classification des stations qui peut être utile pour évaluer divers scénarios, tel que l'ajout d'une nouvelle activité. Finalement, l'analyse des évènements atypiques permet d'identifier les évènements ayant un impact significatif sur l'achalandage et d'estimer cet impact.

Cette recherche contribue au développement d'approches de modélisation et de prévision de la demande en transport. La méthode proposée permet de considérer un ajustement plus dynamique de l'offre de transport pour correspondre à la demande et ce, en se basant sur des données d'évènements météorologiques, d'activités et d'interruptions de service du métro. Ainsi, les modèles développés permettent à un opérateur de transport en commun d'ajuster la fréquence des passages sur les diverses lignes selon divers évènements. De son côté, ces modèles pourraient aider un opérateur de taxi à mieux répartir les véhicules sur le territoire et aider un opérateur de vélopartage à mieux répartir les vélos aux différentes stations. Les résultats de cette étude permettent de soutenir une prise de décision plus éclairée tant sur la planification stratégique d'un réseau intégré et multimodal de transport, que sur la priorisation de certains modes lors de l'allocation d'espace urbain et de ressources.

7.3 Limites

Cette étude comporte plusieurs limites qui peuvent être attribuées aux données utilisées ou aux méthodes employées. Ces limites sont discutées dans cette section.

7.3.1 Limites dues aux données

Certaines limites sont liées aux données transactionnelles utilisées. Tout d'abord, il n'y a aucun identifiant unique entre les différentes bases de données afin de suivre un même utilisateur entre les quatre modes de transport étudiés (sauf pour le métro et le bus qui emploient un système de validation commun). Les différents segments des déplacements multimodaux sont donc considérés comme des déplacements indépendants. Pour continuer, le lieu d'embarquement des déplacements en autobus n'est pas connu. Seule la ligne d'autobus empruntée est connue, c'est pourquoi tous les déplacements d'une même ligne sont attribués aux stations dont elle passe par le voisinage. Par ailleurs, le lieu de débarquement pour le métro et le bus sont inconnus, ce qui rend difficile l'analyse des lieux de destination des déplacements.

Ensuite, les données de validation du bus pour les service spéciaux (lors d'activités) et pour les services provisoires (lors d'arrêts de service du métro) sont incomplètes, c'est pourquoi elles ont dû être omises. En effet, les trajets des services de bus spéciaux et provisoires ne sont pas standards et ceux-ci ne peuvent être retracés et attribués aux données de validation de cartes à puce. Voici pourquoi seules les données transactionnelles des trajets réguliers de bus sont utilisées.

Par ailleurs, plusieurs déplacements en bus ne sont pas pris en compte par le système de validation de carte à puce lors d'interruptions de service du métro. En effet, une pratique répandue des chauffeurs lors des pannes consiste à accélérer l'embarquement des usagers en évitant la perception. Ainsi, le nombre de déplacements en bus lors d'interruptions de service du métro est sous-estimé par les données transactionnelles.

D'autres limites sont liées aux données d'évènements utilisées. Par exemple, les données météorologiques utilisées sont supposées identiques pour l'ensemble des stations de métro, bien que la station de métro la plus éloignée est située à environ 21 km de la station météorologique. Pour continuer, les données sur les interruptions de service du métro indiquent uniquement le lieu d'origine de l'interruption et non l'ensemble des stations touchées par la panne. Cette information est difficile à estimer, étant donné que les stations touchées par une panne varient dans le temps,

selon l'accumulation des trains aux stations précédant la panne et selon les services provisoires instaurés.

Une autre limite vient de la méthode de compilation des données d'activités autour des stations de métro. Celles-ci ont été compilées sans indication claire sur la taille des évènements à inclure. De plus, les activités sont généralement attribuées à la station de métro la plus proche seulement, en négligeant les autres stations à proximité qui pourraient être empruntées par les participants de l'activité.

7.3.2 Limites méthodologiques

Plusieurs limites sont liées aux différentes méthodes employées dans cette étude. Par exemple, les zones de voisinage des stations de métro sont non-exclusives, ce qui signifie qu'elles peuvent se chevaucher. Un déplacement peut donc être compté à la fois dans plusieurs voisinages de stations de métro, ce qui rend l'usage des stations voisines corrélées. Pour continuer, la variation de l'offre de transport dans le temps et dans l'espace est négligée.

En ce qui concerne l'analyse des patrons types d'usage, le nombre de groupe choisi pour le partitionnement de données a certainement une influence sur les groupes formés. De plus, les groupes qui contiennent un nombre important d'observations peuvent dissimuler des patrons moins fréquents.

La typologie de stations de métro dépend grandement des variables choisies pour la classification. Un choix différent de variables aurait probablement modifié les groupes. Le nombre de groupe fixé est également un choix discrétionnaire.

L'analyse des évènements atypiques ne considère pas la disponibilité de vélos aux stations de vélopartage. Par ailleurs, les variables ayant un impact significatif sur l'usage des modes sont supposées être les mêmes pour toutes les stations de métro. Finalement, l'impact des évènements pourrait varier d'une saison à l'autre ou selon l'heure de la journée, mais cet impact est supposé constant par la méthode empruntée.

7.4 Perspectives

La méthode proposée pourrait être appliquée pour d'autres ville que Montréal. De plus, cette méthode pourrait s'appliquer à d'autres modes de transport.

Alors que les autobus de la Société de transport de Montréal sont maintenant tous équipés de capteurs GPS, le lieu d'origine des déplacements est maintenant connu. Il pourrait être pertinent d'utiliser ces nouvelles données afin d'avoir un meilleur estimé du nombre de déplacements initiés dans le voisinage de chaque station de métro.

D'autres définitions du voisinage des stations de métro pourraient être testées, comme par exemple des zones de taille différentes ou des zones exclusives. Il pourrait également être pertinent d'étudier les variations dans les habitudes de déplacements des usagers lorsque l'offre de transport fluctue. Aussi, les lieux d'arrivé des déplacements pourraient être considérés plutôt que les lieux de départ.

L'analyse des patrons types d'usage pourrait être raffinée en faisant du sous-partitionnement de données pour les groupes contenant le plus d'observations, afin de vérifier s'ils dissimulent des patrons d'usage uniques. Par ailleurs, certains sous-ensembles de données (tel que les jours fériés ou la période de relâche) pourraient être analysés indépendamment pour évaluer comment les habitudes de déplacement diffèrent. Aussi, d'autres méthodes de partitionnement pourraient être explorées, tel que le partitionnement hiérarchique.

En outre, il serait pertinent d'étudier l'impact d'autres variables explicatives sur l'usage des modes. Par exemple, les variations du prix de l'essence et de l'offre de transport ont certainement un impact sur l'achalandage. D'autres types de modèles pourraient être testés, tel que des modèles multi-niveaux. Ce type de modèle pourrait permettre de combiner les différentes stations dans un même modèle tout en observant les variations inter-stations. D'autres modèles pour séries temporelles pourraient être testés, dans l'objectif de trouver un modèle plus performant.

BIBLIOGRAPHIE

Andersen, J. L. et Landex, A. (2009). *GIS-based Approaches to Catchment Area Analyses of Mass Transit*. Technical University of Denmark (p. 14). Communication présentée au ESRI International User Conference.

BIXI. (2018). Comment utiliser BIXI ? *BIXI Montreal*. Repéré à https://www.bixi.com/fr/comment-utiliser-bixi

BIXI. (2019). Qui sommes-nous? *BIXI Montreal*. Repéré à https://www.bixi.com/fr/qui-sommes-nous

Brandenburg, C., Matzarakis, A. et Arnberger, A. (2007). Weather and cycling—a first approach to the effects of weather conditions on cycling. *Meteorological Applications*, *14*(1), 61-67.

Cools, M., Moons, E. et Wets, G. (2010). Assessing the Impact of Weather on Traffic Intensity. *Weather, Climate, and Society*, *2*(1), 60-68.

Dill, J. (2013). *Predicting Transit Ridership at The Stop Level: The Role of Service and Urban Form* (p. 19). Communication présentée au TRB 2013 Annual Meeting.

EXO. (2018). RTM - Horaires jours fériés. Repéré à https://rtm.quebec/fr/service-clientele/horaires-jours-feries

Gallop, C., Zhao, J. et Tse, C. (2012). *A Seasonal Autoregressive Model of Vancouver Bicycle Traffic Using Weather Variables* (p. 17). Communication présentée au TRB 2012 Annual Meeting.

Google. (s. d.). Région métropolitaine de Montréal. *Google Maps*. Repéré à https://www.google.ca/maps

Gouvernement du Canada, S. C. (2017). Série « Perspective géographique », Recensement de 2016. Repéré à https://www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Facts-cma-fra.cfm?LANG=Fra&GK=CMA&GC=462&TOPIC=1

Guo, Z., Wilson, N. H. M. et Rahbee, A. (2007). Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record*, *2034*(1), 3-10. doi:10.3141/2034-01

Hyndman, R. J. et Khandakar, Y. (2007). Automatic Time Series Forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3). Repéré à http://www.jstatsoft.org/v27/i03/

James, G., Hastie, T., James, G. et Witten, D. (2013). *An Introduction to Statistical Learning: With Applications in R*. New York : Springer.

Kamga, C., Yazici, M. A. et Singhal, A. (2013). *Hailing in the Rain: Temporal and Weather-Related Variations in Taxi Ridership 2 and Taxi Demand-Supply Equilibrium*. Washington, D.C. Communication présentée au Transportation Research Board's 92nd Annual Meeting.

Lin, T. Y.-T. (2017). *Transit User Mode Choice Behaviour in Response to TTC Rapid Transit Service Disruption*. University of Toronto, Toronto.

Martel Poliquin, É. (2012). *Mieux comprendre les déterminants du choix modal*. Polytechnique Montréal, Montréal (Québec).

Mattson, J. (2008). *Effects of Rising Gas Prices on Bus Ridership for Small Urban and Rural Transit Systems* (p. 52). North Dakota State University.

Miranda-Moreno, L. F. et Nosal, T. (2011). Temporal Trends and Impact of Weather on Cycling in an Urban Environment. *Transportation Research Record*, *2247*(1), 42-52.

Morency, C., Trépanier, M., Saunier, N., Verreault, H. et Bourdeau, S. (2016). The Challenges of using 5 Parallel Passive Data Streams to report on a Wide Range of Mobility Options. *Transportation Research Procedia*, 17.

Murtagh, F. et Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274-295. doi:10.1007/s00357-014-9161-z

Niemeier, D. A. (1996). Longitudinal Analysis of Bicycle Count Variability: Results and Modeling Implications. *Journal of Transportation Engineering*, *122*(3), 200-206.

Nishiuchi, H., King, J. et Todoroki, T. (2013). Spatial-Temporal Daily Frequent Trip Pattern of Public Transport Passengers Using Smart Card Data. *International Journal of Intelligent Transportation Systems Research*, *11*(1), 1-10. doi:10.1007/s13177-012-0051-7

Nosal, T. et Miranda-Moreno, L. F. (2014). The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts. *Transportation Research Part A: Policy and Practice*, *66*, 213-225. doi:10.1016/j.tra.2014.04.012

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., … White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, *13*(4), 295-306. doi:10.1016/j.tranpol.2005.12.004

Resenbloom, S. (2006). *Understanding Women's and Men's Travel Patterns :The Research Challenge* (vol. 1). Transportation Research Board of the National Academies. doi:10.17226/23274

Richardson, A. (2000). Seasonal and Weather Impacts on Urban Cycling Trips. *TUTI Report*. Repéré à http://www.tuti.com.au/2000veloschweiz.pdf

Sabir, M. (2011). *Weather and Travel Behaviour* (Vrije Universiteit, Amsterdam). Repéré à http://dare.ubvu.vu.nl/bitstream/handle/1871/19500/dissertation.pdf?sequence=1

Singhal, A., Kamga, C. et Yazici, A. (2014). Impact of weather on urban transit ridership. *Transportation Research Part A: Policy and Practice*, *69*, 379-391.

Société de transport de Montréal. (2018). *Rapport annuel 2017* (p. 58). Montréal (Québec).

Stover, V. et McCormack, E. (2012). The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, *15*(1), 95-110.

Thomas, T., Jaarsma, R. et Tutert, B. (2008). Temporal variations of bicycle demand in the Netherlands: The influence of weather on cycling, 17.

Toqué, F. (2019). *Visualization tools for spatio-temporal time-series analysis with context awareness.* Paris, France. Communication présentée au TransitData 2019.

Winters, M., Friesen, M. C., Koehoorn, M. et Teschke, K. (2007). Utilitarian Bicycling: A Multilevel Analysis of Climate and Personal Influences. *American Journal of Preventive Medicine*, *32*(1), 52-58.

ANNEXE A – FORMATION DES PÉRIODES HORAIRES

Cette section présente de l'information complémentaire justifiant les résultats présentés à la section 6.1.6.

Tableau 7-1 : Statistique du test de Kolmogorov–Smirnov pour chaque paire de jours

Jour de la semaine	lundi	mardi	mercredi	jeudi	vendredi
lundi	0	0,0102	0,0137	0,0198	0,0179
mardi	0,0102	0	0,0052	0,0096	0,0163
mercredi	0,0137	0,0052	0	0,0067	0,0155
jeudi	0,0198	0,0096	0,0067	0	0,0178
vendredi	0,0179	0,0163	0,0155	0,0178	0

L'hypothèse nulle n'est pas rejetée puisque les statistiques sont sous la valeur critique de 0,0217.

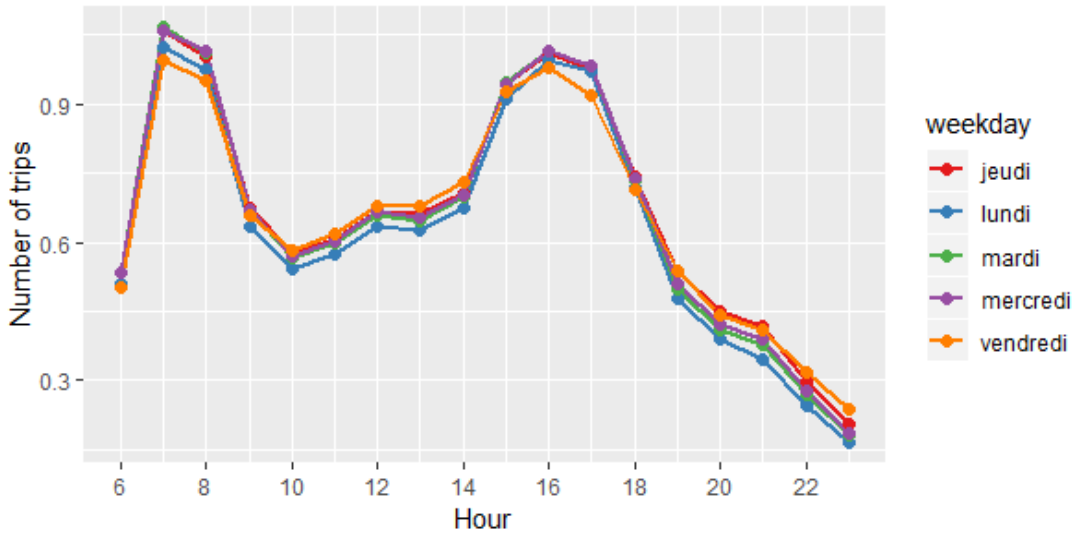

Figure 7-1 : Distribution horaire des déplacements par jour, pour le bus

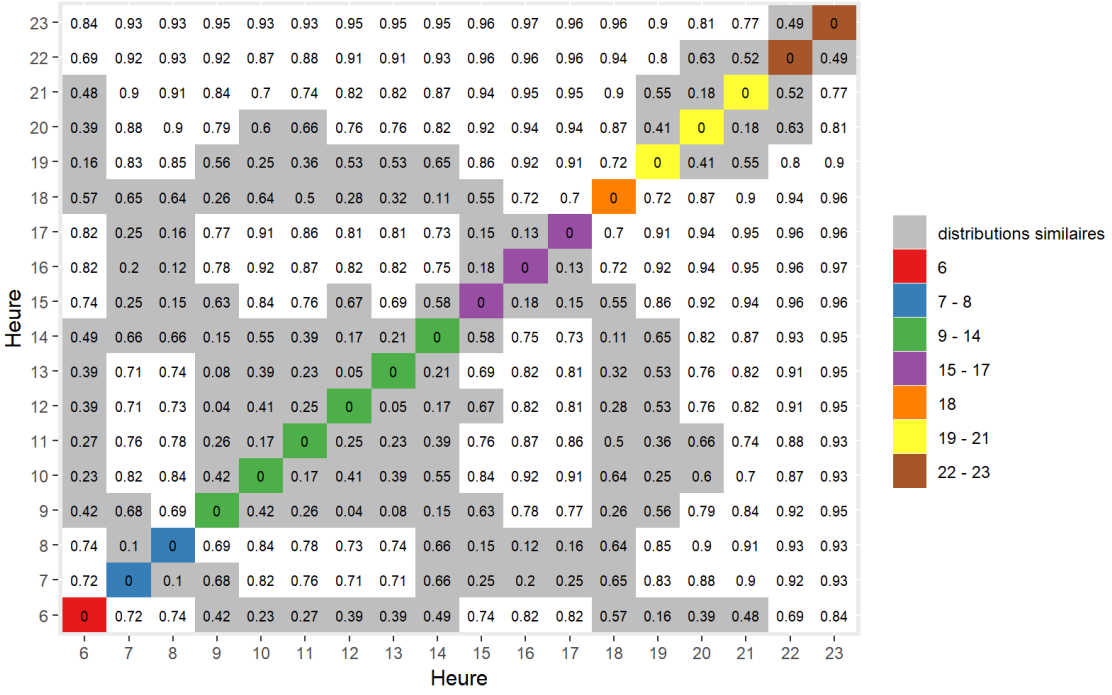

Figure 7-2 : Matrice des distances et périodes horaires formées, pour le bus

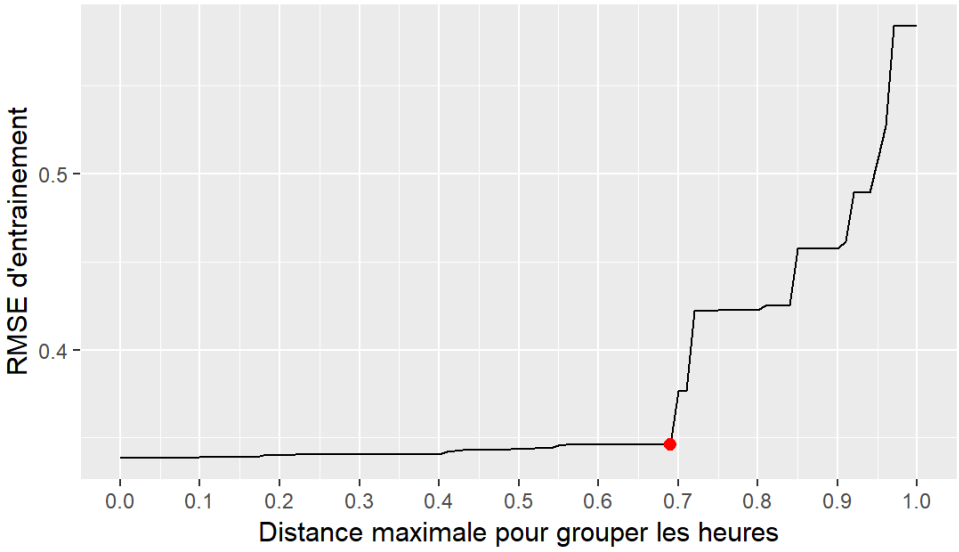

Figure 7-3 : Erreur du modèle selon la distance seuil, pour le bus

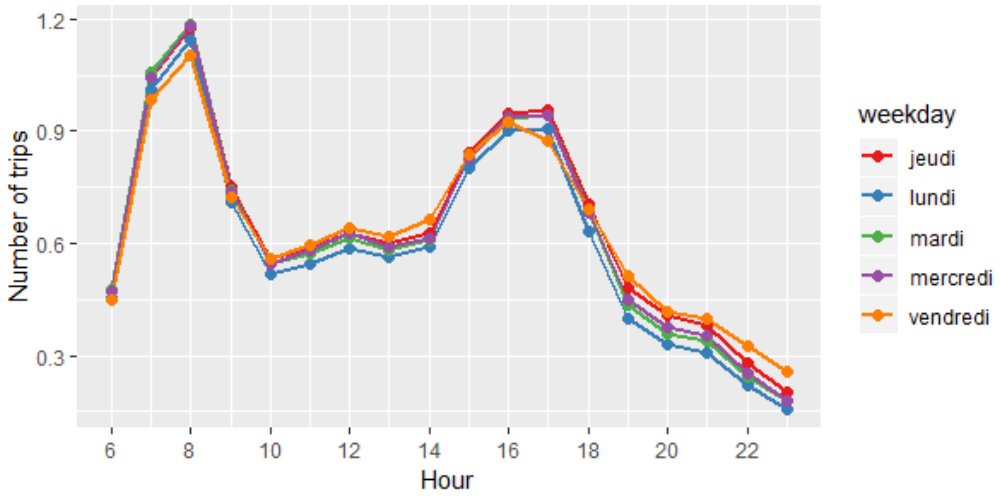

Figure 7-4 : Distribution horaire des déplacements par jour, pour le métro

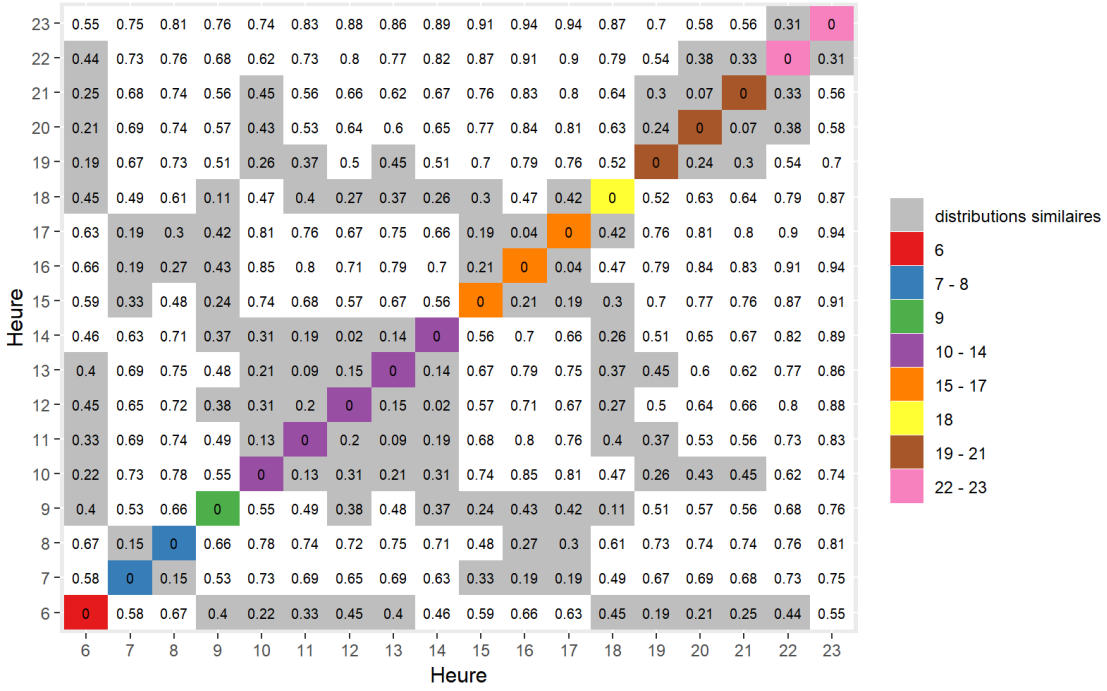

Figure 7-5 : Matrice des distances et périodes horaires formées, pour le métro

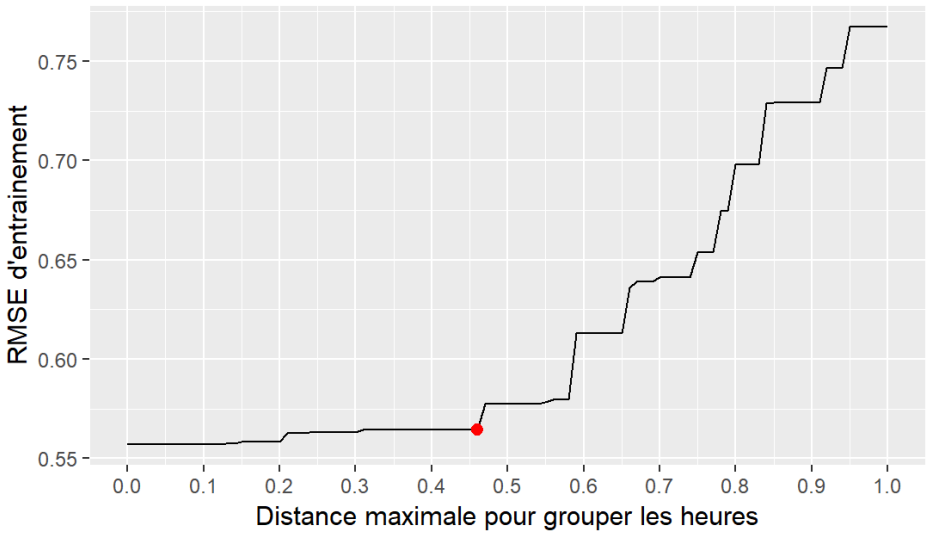

Figure 7-6 : Erreur du modèle selon la distance seuil, pour le métro

ANNEXE B – COEFFICIENTS DES MODÈLES

Cette section présente de l'information complémentaire justifiant les résultats présentés à la section 6.1.7.

Tableau 7-2 : Coefficients du modèle du vélopartage

Variable liée au coefficient	Coefficient	Variation relative de l'intensité
$constante$	0,38119*	0,46403
$pluie_{3h+A}$	-0,07232*	-0,06977
$vent$	-0,00082*	-0,00082
$température$	NA	NA
$panne$	NA	NA
$activité_{exposition_durant}$	0,09003*	0,09421
$activité_{festival_durant}$	0,17824*	0,19511
$activité_{soccer_fin}$	0,73218*	1,07961
$activité_{concert_durant}$	0,04134*	0,04221
$activité_{sport_durant}$	0,09969*	0,10483
$activité_{autre_fin}$	0,12247	0,13029
$activité_{spectacle_fin}$	0,26669	0,30564
$activité_{tennis_durant}$	0,10131*	0,10662
$période_{6h}$	-0,30393*	-0,26209
$période_{7h}$	0,08976*	0,09391
$période_{8h}$	0,40446*	0,49849
$période_{9h}$	0,07644*	0,07944
$période_{10-11h}$	-0,10083*	-0,09591
$période_{16h}$	0,31356*	0,36829
$période_{17h}$	0,50367*	0,65478
$période_{18h}$	0,31063*	0,36428
$période_{19-20h}$	0,04925*	0,05048
$période_{21-22h}$	-0,10712*	-0,10158
$période_{23h}$	-0,23190*	-0,20698

*Significatif à un seuil de 95%

Tableau 7-3 : Coefficients du modèle du taxi (suite)

Variable liée au coefficient	Coefficient	Variation relative de l'intensité
$constante$	0,67398*	0,96203
$pluie_{1h+A}$	0,03409*	0,03468
$température$	NA	NA
$panne$	NA	NA
$activité_{hockey_fin}$	0,47612*	0,60982
$activité_{exposition_fin}$	0,18214*	0,19978
$activité_{festival_durant}$	0,15836*	0,17159
$activité_{soccer_fin}$	0,66099*	0,93671
$activité_{concert_durant}$	0,24212*	0,27395
$activité_{sport_durant}$	0,14605*	0,15726
$activité_{spectacle_fin}$	0,1263*	0,13462
$activité_{tennis_durant}$	0,2776*	0,31996
$période_{lundi_6h}$	-0,28528*	-0,24820
$période_{lundi_7h}$	-0,10763*	-0,10204
$période_{lundi_8-9h}$	0,11493*	0,12180
$période_{lundi_19-22h}$	-0,20255*	-0,18335
$période_{lundi_23h}$	-0,27442*	-0,23999
$période_{mardi_6h}$	-0,32806*	-0,27968
$période_{mardi_7h}$	-0,11108*	-0,10513
$période_{mardi_8-9h}$	0,13566*	0,14529
$période_{mardi_10-18h}$	0,04174*	0,04263
$période_{mardi_19-23h}$	-0,14196*	-0,13235
$période_{mercredi_6h}$	-0,30787*	-0,26499
$période_{mercredi_7h}$	-0,11793*	-0,11125
$période_{mercredi_8-9h}$	0,13297*	0,14221
$période_{mercredi_10-18h}$	0,04774*	0,04890
$période_{mercredi_19-23h}$	-0,10659*	-0,10110
$période_{jeudi_6h}$	-0,31712*	-0,27175
$période_{jeudi_7h}$	-0,08948*	-0,08559
$période_{jeudi_8-9h}$	0,16609*	0,18067
$période_{jeudi_10-16h}$	0,09186*	0,09621
$période_{jeudi_17-18h}$	0,12860*	0,13723

Tableau 7-3 : Coefficients du modèle du taxi (fin)

$période_{jeudi_19-23h}$	0,00720*	0,00722
$période_{vendredi_6h}$	-0,29504*	-0,2555
$période_{vendredi_7h}$	-0,10109*	-0,09615
$période_{vendredi_8-9h}$	0,14513*	0,15619
$période_{vendredi_10-17h}$	0,08209*	0,08555
$période_{vendredi_18-19h}$	0,11445*	0,12125
$période_{vendredi_20-21h}$	0,04127*	0,04214
$période_{vendredi_22h}$	0,10260*	0,10805
$période_{vendredi_23h}$	0,18202*	0,19964

*Significatif à un seuil de 95%

Tableau 7-4 : Coefficients du modèle du métro

Variable liée au coefficient	Coefficient	Variation relative de l'intensité
$constante$	0,59824*	0,81891
$pluie_{jour}$	-0,01144*	-0,01137
$température$	NA	NA
$panne$	NA	NA
$activité_{hockey_fin}$	0,56181*	0,75384
$activité_{exposition_durant}$	0,12768*	0,13619
$activité_{festival_durant}$	0,45286*	0,57280
$activité_{soccer_fin}$	0,79244*	1,20877
$activité_{concert_fin}$	0,45544*	0,57687
$activité_{sport_durant}$	0,26127*	0,29858
$activité_{autre_durant}$	0,32023*	0,37745
$activité_{football_durant}$	0,31552*	0,37097
$activité_{spectacle_fin}$	0,45960*	0,58345
$activité_{tennis_fin}$	0,35038*	0,41960
$période_{6h}$	-0,11259*	-0,10648
$période_{7-8h}$	0,41436*	0,51340
$période_{9h}$	0,08664*	0,09051
$période_{15-17h}$	0,35248*	0,42259
$période_{18h}$	0,11765*	0,12485
$période_{19-21h}$	-0,15198*	-0,14099
$période_{22-23h}$	-0,31639*	-0,27122

*Significatif à un seuil de 95%

Tableau 7-5 : Coefficients du modèle du bus

Variable liée au coefficient	Coefficient	Variation relative de l'intensité
$constante$	0,66140*	0,93749
$pluie_{jour}$	-0,01168*	-0,01161
$température$	NA	NA
$panne$	NA	NA
$activité_{hockey_durant}$	-0,17305*	-0,1589
$activité_{exposition_durant}$	0,01734*	0,01749
$activité_{festival_durant}$	0,02650*	0,02685
$activité_{soccer_durant}$	0,03514*	0,03576
$activité_{concert_durant}$	-0,15418*	-0,14288
$activité_{spectacle_durant}$	-0,09096*	-0,08695
$période_{6h}$	-0,09944*	-0,09465
$période_{7-8h}$	0,36465*	0,44001
$période_{15-17h}$	0,32699*	0,38679
$période_{18h}$	0,09338*	0,09788
$période_{19-21h}$	-0,19929*	-0,18069
$période_{22-23h}$	-0,40074*	-0,33017

*Significatif à un seuil de 95%

ANNEXE C – COEFFICIENTS DES MODÈLES PAR TYPE DE STATION

Cette section présente de l'information complémentaire justifiant les résultats présentés à la section 6.2.

Tableau 7-6 : Coefficients du modèle GAM, pour le vélopartage

Variable	Type 1	Type 2	Type 3	Type 4	Type 5
$pluie_{3h+A}$	-0,0693	-0,0702	-0,0762	-0,0677	-0,0776
$vent$	-0,0006	-0,0008	-0,0010	-0,0012	-0,0009
$temperature$	Spline de lissage				
$période_{6h}$	-0,4689	-0,3191	-0,2448	-0,2791	-0,5536
$période_{7h}$	-0,2431	0,0685	0,2278	-0,0490	-0,3045
$période_{8h}$	0,0473	0,4656	0,5527	0,1420	-0,1041
$période_{9h}$	-0,1144	0,1736	0,1515	-0,1012	-0,3025
$période_{10-11h}$	-0,1895	-0,0710	-0,0770	-0,1634	-0,2198
$période_{16h}$	0,4466	0,2407	0,2501	0,3674	0,4970
$période_{17h}$	0,6964	0,4320	0,4567	0,5412	0,4254
$période_{18h}$	0,3335	0,3082	0,3161	0,3955	0,1780
$période_{19-20h}$	0,0013	0,1053	0,0741	0,0901	-0,1062
$période_{21-22h}$	-0,1626	-0,0258	-0,0895	-0,0198	-0,2433
$période_{23h}$	-0,2782	-0,1407	-0,2129	-0,2205	-0,4507
valeur initiale	0,4401	0,3840	0,3560	0,3217	0,6504

Tableau 7-7 : Coefficients du modèle ARIMA, pour le vélopartage

Variable	Type 1	Type 2	Type 3	Type 4	Type 5
$pluie_{3h+A}$	-0,0481	-0,0498	-0,0586	-0,0567	-0,0599
$vent$	-0,0006	-0,0007	-0,0006	-0,0010	-0,0009
$temperature$	0,0132	0,0133	0,0126	0,0171	0,0124
$période_{6h}$	-0,4250	-0,3005	-0,2344	-0,2710	-0,4961
$période_{7h}$	-0,2051	0,0795	0,2324	-0,0412	-0,2489
$période_{8h}$	0,0911	0,4768	0,5584	0,1515	-0,0467
$période_{9h}$	-0,0515	0,1886	0,1603	-0,0880	-0,2406
$période_{10-11h}$	-0,1703	-0,0635	-0,0696	-0,1504	-0,1775
$période_{16h}$	0,3960	0,2029	0,2107	0,3453	0,4329
$période_{17h}$	0,6538	0,3977	0,4175	0,5262	0,3791
$période_{18h}$	0,2751	0,2566	0,2661	0,3795	0,1426
$période_{19-20h}$	-0,0190	0,0825	0,0520	0,0800	-0,1168
$période_{21-22h}$	-0,1165	0,0029	-0,0657	-0,0154	-0,2167
$période_{23h}$	-0,2140	-0,0999	-0,1804	-0,2054	-0,4048
valeur initiale	0,4299	0,3834		0,3240	
ar1	1,3307	1,2978	0,5973	0,7777	0,6075
ar2	-0,3980	-0,3554			
ma1	-0,6475	-0,6386	-1,0128	-0,5765	-1,3120
ma2	-0,1057	-0,1361	0,0228	-0,0279	0,3191
sar1	0,2925	0,2153		0,1274	
sar1	0,2498	0,1635		0,1173	
sma1			0,1857		0,2086
sma2					0,1833