

UNIVERSITÉ DE MONTRÉAL

ANALYSE COMPARATIVE ENTRE LES ENQUÊTES MÉNAGES ORIGINE-DESTINATION ET
LES SYSTÈMES DE PAIEMENT PAR CARTE À PUCE EN TRANSPORT URBAIN

CARL BLANCHETTE

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INDUSTRIEL)
DÉCEMBRE 2008

©Carl Blanchette, 2008.



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-47651-2
Our file *Notre référence*
ISBN: 978-0-494-47651-2

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ANALYSE COMPARATIVE ENTRE LES ENQUÊTES MÉNAGES
ORIGINE-DESTINATION ET LES SYSTÈMES DE PAIEMENT
PAR CARTE À PUCE EN TRANSPORT URBAIN

présenté par : BLANCHETTE Carl

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

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

M. AGARD Bruno, Doct., président

M. TRÉPANIÉ Martin, ing., Ph.D., membre et directeur de recherche

M. MORENCY Catherine, Ph.D., membre et codirectrice de recherche

M. CHAPLEAU Robert, ing., Ph.D., membre

À ma famille, spécialement mes parents, mes grands-parents et ma sœur qui ont été si généreux et qui ont toujours été présents pour moi. Je dédie aussi ce mémoire à Ariane, qui m'a accompagné et soutenu pendant cette rude épreuve. Finalement, je tiens à remercier mes amis, que j'ai dû négliger, mais qui m'ont toujours épaulé et compris.

« Mais où sont les deux gouttes d'huile que je t'avais confiées? » demanda le Sage.

Le jeune homme, regardant alors la cuiller, constata qu'il les avait renversées.

« Eh bien », dit alors le Sage des Sages, « C'est là le seul conseil que j'aie à te donner : le secret du bonheur est de regarder toutes les merveilles du monde, mais sans jamais oublier les deux gouttes d'huile dans la cuiller. »

L'Alchimiste (1988), de Paulo Coelho, écrivain brésilien (1947 -) et alchimiste à ses heures.

REMERCIEMENTS

Je profite de ces quelques lignes pour remercier chaleureusement mes directeurs de recherche, le professeur Martin Trépanier du département de mathématiques et de génie industriel et la professeure Catherine Morency du département des génies civil, géologique et des mines. Leur appui, autant financier que moral, ainsi que leur conseil m'ont été fort précieux tout au long des expérimentations et de la rédaction de ce mémoire. Mais, au-delà des connaissances et des techniques qu'ils m'ont apprises, ils m'ont inculqué une rigueur de travail et une manière de raisonner qui me resteront pour toujours et dont je leur serai à jamais reconnaissant. Ce fût une expérience inoubliable et je suis choyé de l'avoir réalisé avec des personnes de cette trempe.

Je tiens aussi à remercier la Société de transport de l'Outaouais ainsi que le comité TRANS qui ont gracieusement mis à ma disposition des données riches et complètes sur les déplacements faits dans la grande région de Gatineau. Ces informations ont été inestimables et elles ont joué un rôle prépondérant dans l'élaboration de ce mémoire. Merci.

Je tiens à remercier le département de mathématiques et de génie industriel, le département des génies civil, géologique et des mines ainsi que tous mes professeurs qui ont participé de près ou de loin à ma formation académique. Finalement, je veux remercier les étudiants du département des transports avec qui j'ai eu l'occasion de fraterniser à l'école et lors du 44^e congrès de l'AQTR à Québec, et qui ont bien gentiment répondu à toutes mes questions.

RÉSUMÉ

L'accès à de l'information de qualité pour répondre aux besoins des planificateurs en milieu urbain est préoccupant. Les techniques de collecte de données basées sur des échantillonnages, telle que les enquêtes ménage origine-destination, ont de plus en plus de difficulté à rejoindre la population et à obtenir un échantillon riche, diversifié et précis.

Un complément d'information pouvant en partie résoudre cette problématique pourrait provenir d'une technologie qui s'implante sur plusieurs réseaux de transport en commun à travers le monde et qui gagne en intérêt. Cette technologie, les systèmes de paiement par cartes à puce, permet d'enregistrer les déplacements effectués quotidiennement sur le réseau. Cette source d'information est donc unique puisqu'elle est prise en continu et elle fournit énormément de données sur les déplacements de la population empruntant un réseau spécifique. Cependant, cette source est très pauvre relativement aux informations sur les caractéristiques sociodémographiques de la population.

La fusion de ces deux sources de données s'impose donc comme une solution potentiellement intéressante afin de combler ce besoin en information. La Société de transport de l'Outaouais, un chef de file québécois en innovation technologique, dispose d'un système de cartes à puce déjà implantée et offre une occasion unique d'étudier et de comparer ces données à celles de l'enquête ménage origine-destination qui a eu lieu dans la grande région de la capitale nationale en 2005.

L'objectif principal de ce mémoire est donc de comparer deux bases de données contenant des informations provenant de deux sources distinctes, soit l'enquête ménage origine-destination 2005 et le système de paiement par cartes à puce de la Société de transport de l'Outaouais.

Cette analyse comparative, s'appuyant sur les objets définis par l'approche orientée-objet, se décompose en trois catégories d'analyses, l'analyse globale, segmentée et du réseau. Les analyses globales montrent des écarts considérables entre les données de l'enquête origine-destination et des systèmes de cartes à puce. Mais ces analyses sont limitées par leur niveau de détail puisqu'elles utilisent des données agrégées.

Les analyses segmentées, par contre, permettent d'étudier des données désagrégées selon la journée, l'heure et le titre de transport utilisé lors des déplacements. Ces analyses démontrent que l'écart entre les données n'est pas global, mais bien spécifique à certains titres de transport ou périodes de temps. Ces analyses permettent aussi de mieux définir et comprendre la variabilité quotidienne des déplacements qui est propre aux usagers de la Société de transport de l'Outaouais.

Les analyses du réseau détaillent les déplacements selon les objets arrêts et lignes définis dans l'approche orientée-objet. L'analyse des lignes montre qu'il est possible de comparer les différentes données à un niveau aussi désagrégé que les montées sur les lignes. Cette analyse a aussi permis d'étudier plus spécifiquement la variabilité quotidienne sur une ligne en fonction des titres de transport des usagers et de constater que les étudiants et les adultes ont des habitudes de déplacement similaires.

Aussi, l'étude par paires origine-destination confirme que les données des systèmes de cartes à puce reflètent mieux les déplacements internes. En fait, l'analyse de la distribution des déplacements, de l'enquête et des systèmes de cartes à puce, par paires origine-destination a montré qu'il y a une sous-représentation des déplacements internes de l'enquête ménage origine-destination 2005.

Suite à ces comparaisons, des biais dans les différentes bases de données ont été soulevés et une méthode de correction des facteurs d'expansion de l'enquête origine-destination a été appliquée. Cette méthode a démontré des résultats encourageants et

prometteurs. Elle a aussi démontré qu'il était possible de modifier ou de calculer des facteurs d'expansion à partir des données des systèmes de paiement par cartes à puce.

Une discussion sur les techniques de fusion a aussi soulevé le manque d'expérimentation dans le domaine des transports en commun et elle propose une taxinomie des différentes méthodes de fusion. Des améliorations concernant les différentes techniques de collecte de données, vues dans ce mémoire, ont été proposées afin de faciliter la fusion des données. Finalement, des perspectives de recherche sur la fusion et la distribution des codes d'erreurs dans les bases de données a permis de constater les progrès en la matière et le chemin qu'il reste à faire.

ABSTRACT

Gathering quality information for the needs of transit planning has recently been preoccupying. Most of the actual sources of information based on a sample of the population, like the household travel survey, are encountering a reduction in the response rate; therefore it is harder to obtain a precise and acceptable sample.

Smart card collection fare system may be the answer to this issue. In fact, many transit network around the world are implementing smart card collection fare system or will implement them in a near future. The smart card can provide continuous day-by-day data of the trips made on the network. The data gathered by smart cards are unfortunately lacking information about the traveler and the purpose of the journey.

Data fusion is beginning to stand out as a potentially good solution to fill this need of information. The Société de transport de l'Outaouais, a Canadian leader in innovative solutions for transit network, is already using a smart card collection fare system and have access to data from a 2005 household survey that took place in the area of Ottawa and Gatineau. This is a one of a kind opportunity to study those data and compare them since they were taken at the exact same time.

The primary goal of this work is to compare two different databases coming from two distinctive sources (a household survey and smart cards system). The studies are based on an object-oriented approach and they were divided into three categories, the global analysis, the segmented analysis and the network analysis. The global analysis showed a considerably large gap between the data from the household survey and the smart cards. But, those analyses are limited because they are aggregated.

The data used in the segmented analyses are disaggregated and thus permit to study objects like the days, the hours and the monthly pass holders. Those analyses showed that the discrepancy between the two datasets is local instead of global. Also, a study

of the day-by-day variability of travels is done and it gives us a better understanding of the user behaviors on that particular transit network.

Network analyses were made with the objects routes and bus stops and, like the segmented analyses, they use disaggregated data. A study of the temporal variability on particular routes showed that the students and the adults have similar behaviors.

Analyzing origin-destination matrices confirmed that the estimates from the smart cards are more accurate when it comes to short-distance travels. In fact, analyzing the distribution of the trips, from both databases, by their origin and destination shows lower trip rate from the travel household survey.

After all those comparisons, it has been shown that a bias exists in both datasets and a corrective method has been applied to the household survey estimates. This method showed promising and encouraging results for further research and it demonstrated that it is possible to modify or calculate household estimates from other sources like smart cards database.

A discussion about data fusion techniques confirmed the lack of experiment in the field of knowledge of transit transportation and proposed a terminology of different methods of fusion. Some improvements for the two methods analyzed in this work were proposed to facilitate the integration of data. Finally, thoughts on the future of data fusion and on methods to spread errors codes across the databases showed the progress that has been made and what still need to be done.

TABLE DES MATIÈRES

DÉDICACE.....	iv
REMERCIEMENTS	v
RÉSUMÉ.....	vi
ABSTRACT	ix
TABLE DES MATIÈRES.....	xi
LISTE DES TABLEAUX	xiv
LISTE DES FIGURES	xv
LISTE DES SIGLES ET ABRÉVIATIONS	xix
LISTE DES ANNEXES.....	xx
CHAPITRE 1 INTRODUCTION	1
1.1 Problématique	3
1.2 Buts et objectifs	5
1.3 Structure du mémoire.....	6
CHAPITRE 2 REVUE DE LITTÉRATURE	8
2.1 Les enquêtes transport	8
2.2 L'approche totalement désagrégée.....	13
2.3 Modélisation orientée objets	15
2.4 Carte à puce	19
2.4.1 Enjeux et intérêts	23
2.4.2 Données TC	29
2.5 La fusion des données.....	31

2.5.1	Constats.....	31
2.5.2	Perspectives	32
CHAPITRE 3 MÉTHODOLOGIE		34
CHAPITRE 4 LES SYSTÈMES D'INFORMATIONS		43
4.1	Analyse du contexte de la STO.....	43
4.2	Présentation des données de L'EOD.....	46
4.2.1	Informations sur la population	46
4.3	Présentation des données CAP	53
4.4	Préparation des données.....	55
4.4.1	Les objets étudiés.....	55
4.4.2	Recherche des déplacements CAP dans la base de données EOD	57
4.4.3	Les hypothèses posées et leurs justifications	60
4.4.4	Traitement des données sous Excel 2007.....	61
CHAPITRE 5 ANALYSES ET RÉSULTATS		64
5.1	Analyse globale	64
5.2	Analyse segmentée.....	67
5.2.1	Par journée.....	67
5.2.2	Par type de carte.....	69
5.2.3	Par demi-heure et groupe d'heure	78
5.3	Analyse du réseau.....	89
5.3.1	Par lignes.....	89
5.3.2	Par origine et destination.....	107

CHAPITRE 6 DISCUSSION RELATIVE À L'EOD ET AUX RÉSULTATS OBTENUS	121
6.1 Les facteurs d'expansion.....	121
6.2 Le potentiel de fusion	128
6.3 Les améliorations.....	135
CHAPITRE 7 CONCLUSION	137
RÉFÉRENCES	143
ANNEXES	149

LISTE DES TABLEAUX

Tableau 4.1: Modes de déplacement de la population de Gatineau (Source MTQ)	52
Tableau 4.2: Messages d'erreur pour les enregistrements de la base de données CAP.	54
Tableau 5.1: Analyse statistique selon les journées (CAP et EOD)	68
Tableau 5.2: Analyse des déplacements selon les titres de transport	72
Tableau 5.3: Analyse des déplacements des « adultes réguliers ».....	75
Tableau 5.4: Analyse des déplacements des « étudiants réguliers »	77
Tableau 5.5: Analyse des déplacements CAP et EOD en fonction des groupes d'heures	84
Tableau 5.6: Analyse des déplacements CAP et EOD aux heures de pointe (2,5) par mois	84
Tableau 5.7: Analyse des déplacements CAP et EOD aux heures de pointe (2,5) par journée.....	85
Tableau 5.8: Analyse des lignes 46, 50 et 67 (CAP et EOD)	96
Tableau 5.9: Analyse par titre de transport de la ligne 46 (CAP et EOD)	98
Tableau 5.10: Analyse par titre de transport de la ligne 50 (CAP et EOD)	100
Tableau 5.11: Analyse par titre de transport de la ligne 67 (CAP et EOD)	102
Tableau 5.12: Nombre de montées quotidiennes sur les lignes 50 et 67 en fonction des habitudes de déplacements.....	104
Tableau 5.13: Régions municipales des origines et des destinations.....	111
Tableau 5.14: Matrice OD des déplacements de l'EOD	112
Tableau 5.15: Matrice OD des déplacements des CAP	115
Tableau 5.16: Matrice des proportions par paires OD (EOD/CAP)	119
Tableau 6.1: Matrice OD des déplacements EOD normalisés	124
Tableau A.1: Équivalence des titres de transport (CAP et EOD)	150
Tableau A.2 : Équivalence des lignes du réseau (CAP).....	151
Tableau A.3: Modes de transport enquêtés (EOD).....	153

LISTE DES FIGURES

Figure 1.1: Compatibilité et écarts entre les données provenant de sources différentes (CAP et EOD).....	5
Figure 1.2: Principaux thèmes de la revue de littérature	6
Figure 2.1: Objets et collection d'objets de l'enquête origine-destination (inspiré de Trépanier & Chapleau (2001a)).....	16
Figure 2.2: Modèle orienté objets de la STO (inspiré de Trépanier & Chapleau (2001a))	18
Figure 2.3: Intégration de sources de données distinctes	33
Figure 3.1: Concepts de la méthodologie adoptée.....	34
Figure 3.2: Méthodologie employée pour la création des fichiers comparables d'analyse	36
Figure 3.3: Définition de l'univers d'étude	38
Figure 3.4 : Méthode d'extraction des données.....	39
Figure 3.5 : Résumé des analyses	40
Figure 4.1: Secteurs enquêtés (source : iTRANS Consulting inc. (2006)).....	47
Figure 4.2: Territoire québécois de l'enquête origine-destination 2005 (source: iTRANS Consulting inc. (2006))	48
Figure 4.3: Distribution de la population selon l'âge.....	49
Figure 4.4: Occupation de la population de Gatineau (Transport Québec, 2007)	50
Figure 4.5 Occupation des titulaires d'une carte à puce	51
Figure 4.6: Principales lignes de désir pour tout le territoire enquêté (Source: iTRANS Consulting inc. (2006))	53
Figure 4.7: Tables et champs utilisés dans la base de données EOD.....	56
Figure 4.8: Tables et champs utilisés dans la base de données CAP	57
Figure 4.9: Déplacements CAP de la base de données EOD	59
Figure 4.10: Anomalies des déplacements CAP pour les résidents hors Québec	60

Figure 4.11: Interface de l'utilisateur du fichier de comparaison des données	62
Figure 5.1: Analyse du nombre de personnes utilisant le transport en commun (CAP) .	66
Figure 5.2: Déplacements par titre de transport pour toutes les journées (CAP)	71
Figure 5.3: Comparaison des déplacements selon les titres de transport (CAP et EOD)	73
Figure 5.4: Déplacements des « adultes réguliers » pour toutes les journées (CAP et EOD).....	74
Figure 5.5: Déplacements des « étudiants réguliers » pour toutes les journées (CAP et EOD).....	76
Figure 5.6: Déplacements pour toutes les journées selon les groupes d'heures (CAP) ..	79
Figure 5.7: Écart entre les déplacements CAP (moyenne, max, min) et EOD (moyenne) en fonction des groupes d'heures	81
Figure 5.8: Déplacements pour toutes les journées enquêtées selon le groupe d'heures 2 et 5 (CAP et EOD).....	82
Figure 5.9: Déplacements pour toutes les journées enquêtées selon le groupe d'heures 1, 3, 4, 6 et 7 (CAP et EOD)	83
Figure 5.10: Déplacements pour chaque demi-heure de la journée et pour toutes les journées enquêtées (CAP).....	87
Figure 5.11: Comparaison des moyennes de déplacement des données CAP et EOD par demi-heure	88
Figure 5.12: Analyse globale des lignes (CAP et EOD)	90
Figure 5.13: Analyse des lignes express (CAP et EOD)	92
Figure 5.14: Analyse des lignes régulières (CAP et EOD)	94
Figure 5.15: Analyse des lignes de pointe (CAP et EOD).....	95
Figure 5.16: Analyse des montées sur les lignes 46, 50 et 67 (CAP et EOD)	97
Figure 5.17: Analyse des montées sur la ligne 46 pour les adultes et les étudiants réguliers (CAP et EOD).....	99

Figure 5.18: Analyse des montées sur la ligne 50 pour les adultes et les étudiants réguliers (CAP et EOD).....	101
Figure 5.19: Analyse des montées sur la ligne 67 pour les adultes et les étudiants réguliers (CAP et EOD).....	102
Figure 5.20: Proportion des usagers selon les fréquences de montées sur les lignes 50 et 67 (CAP).....	103
Figure 5.21: Distribution de montées sur les lignes 50 et 67 selon les habitudes de déplacements pour les « adultes réguliers » et les « étudiants réguliers » (CAP).....	105
Figure 5.22: Analyse d'une chaîne de déplacements provenant de l'EOD.....	108
Figure 5.23: Régions ontariennes desservies par la STO (Source: STO)	110
Figure 5.24: Régions québécoises desservies par l'OC Transpo (Source: OC Transpo) .	110
Figure 5.25: Répartition géographique des arrêts.....	113
Figure 5.26: Secteurs municipaux de la région de la capitale nationale (Source : iTRANS Consulting inc., (2006))	114
Figure 5.27: Différence entre l'estimation EOD et CAP des matrices OD.....	117
Figure 6.1: Analyse des mesures de l'erreur par groupe d'heure (CAP et EOD)	125
Figure 6.2: Analyse des mesures de l'erreur par demi-heure (CAP et EOD)	126
Figure 6.3: Analyse des mesures de l'erreur par lignes (CAP et EOD)	127
Figure 6.4: Intégration totale	130
Figure 6.5: Exemple d'intégration partielle	132
Figure 6.6: Exemple d'intégration analytique.....	133
Figure B.1: Champs invisibles lors des requêtes	155
Figure B.2: Première étape de l'importation dans le fichier de comparaison.....	158
Figure B.3: Deuxième étape de l'importation dans le fichier de comparaison	159
Figure B.4: Troisième étape de l'importation dans le fichier de comparaison	159
Figure B.5: Dernière étape de l'importation dans le fichier de comparaison	160

Figure B.6 : Analyse des lignes interzones (CAP et EOD)160

LISTE DES SIGLES ET ABRÉVIATIONS

CAP =	Carte à puce
CATI =	<i>Computer-assisted telephone interview</i>
EOD =	Enquête ménage origine-destination
Matrice OD =	Matrice origine-destination
OD =	Origine-destination
STO =	Société de transport de l'Outaouais
TC =	Transport en commun
TRANS =	Comité conjoint pour la planification des transports dans la région de la Capitale nationale

LISTE DES ANNEXES

ANNEXE A : LES ÉQUIVALENCES DES NOMENCLATURES DE L'EOD ET DES SYSTÈMES DE CAP	149
A.1 : Les titres de transport	149
A.2 : Les lignes du réseau	150
A.3 : Les modes de transport.....	152
ANNEXE B : LES FICHIERS DE COMPARAISONS	154
B.1: L'exportation des données vers le fichier de comparaison	154
B.2 : L'analyse des lignes interzones	160

CHAPITRE 1

INTRODUCTION

Les transports en commun occupent une place importante dans la société d'aujourd'hui. Entre les milliards de personnes qui se déplacent par journée et les problèmes environnementaux et énergétiques de plus en plus préoccupants, le transport en commun est donc amené à s'imposer davantage comme une solution intéressante et sensée. Heureusement, au fil des années, l'offre des réseaux de transports en commun a évolué et des techniques de planification de plus en plus performantes ont vu le jour. Un service amélioré, des heures plus flexibles et une meilleure utilisation des dispositifs informatiques et des équipements font en sorte que les sociétés de transport sont de plus en plus performantes et répondent plus facilement à une demande qui est maintenant bien souvent croissante.

Mais la planification de l'offre n'est pas aisée et elle requiert beaucoup d'informations sur les habitudes de déplacements de la population ainsi que la démographie de la région étudiée. Il existe plusieurs techniques, par exemple les comptes à bord, les recensements et les enquêtes origine-destinations, qui permettent de quantifier et de qualifier l'état de la population. Toutes ces techniques offrent des avantages et elles ont toutes aussi des défauts inhérents à leur technologie et à la précision de leurs données.

Au Québec, une technique fréquemment utilisée pour établir un portrait de la population (ménages, personnes, déplacements) est l'enquête ménage téléphonique origine-destination (EOD). En fait, cette méthode est couramment utilisée et perfectionnée depuis les années 1960-1970. Cette grande expertise dans ce domaine fait du Québec l'un des chefs de file et ceci incite tous ses acteurs à rechercher et développer un savoir-faire encore plus appréciable dans l'utilisation des données récoltées.

Ces enquêtes, traditionnellement faites grâce à des entretiens téléphoniques, permettent de faire le portrait des déplacements selon un jour moyen. Les modes utilisés, les origines et les destinations des déplacements effectués sont autant d'informations nécessaires recueillies grâce à ces enquêtes. Bien que celles-ci tiennent compte de tous les types de déplacements (par voiture, transport en commun, etc.), l'enquête est également utilisée afin de tracer un portrait de la demande sur le réseau de transport en commun. Ce portrait a donc pour but d'offrir un meilleur service, au moment opportun, et, surtout, à la clientèle visée. Les enquêtes sont aussi conçues afin de représenter la réalité actuelle de la société et, par conséquent, elles sont constamment remodelées afin d'obtenir des données concernant des enjeux de l'heure. C'est donc un outil qui est encore couramment utilisé et qui a su faire ses preuves au fil des années.

Parallèlement à ces enquêtes, d'autres techniques font tranquillement leur apparition dans le milieu du transport en commun québécois. L'une d'entre elles, les systèmes de paiement par cartes à puce (CAP), suscite un intérêt grandissant, notamment en Outaouais, où un tel système est implanté depuis 2001, et à Montréal où il est en cours d'implantation. Cette technologie, utilisée pour améliorer et simplifier les systèmes de paiement, permet aussi d'obtenir des données sur les habitudes de déplacement des titulaires.

Le principal avantage d'un tel système réside dans le fait que les données recueillies représentent, à quelques erreurs près, les déplacements des usagers à tout moment de la journée. Bref, ces données peuvent virtuellement être considérées comme l'équivalent d'un historique des produits vendus dans une entreprise manufacturière. Cependant, l'analyse et l'exploitation de ces données afin d'améliorer la qualité du service demeurent encore un sujet partiellement exploré et où il reste beaucoup de recherche à réaliser. Bien que ces données, nombreuses et complexes, aient un potentiel pour la planification encore en développement, elles sont très prometteuses.

Comme toutes les méthodes existantes, les cartes à puce ont des limites ou des manques à combler relativement à l'information pouvant aider à la planification de l'offre sur les réseaux où elles sont utilisées. Mais cette information manquante, principalement d'ordre socio-démographique, est déjà présente par d'autres sources de données, notamment dans les enquêtes origine-destination.

Il est donc pertinent de s'interroger sur la faisabilité de fusionner les données provenant des cartes à puce et des enquêtes origine-destination afin de représenter plus exactement les habitudes de déplacements. Le cas échéant, il faut se questionner aussi sur la manière dont ces bases de données pourront être intégrées l'une à l'autre.

Ceci permettrait d'établir un portrait longitudinal des habitudes de déplacement de la population au lieu de l'image ponctuelle obtenue d'une enquête origine-destination. Par conséquent, ceci permettrait d'améliorer la planification et d'optimiser l'utilisation des ressources, humaines et matérielles, du réseau. Les enjeux sont grands et les retombées sont potentiellement très intéressantes.

Actuellement, le contexte est favorable à cette analyse puisque, depuis quelques années déjà, la société de transport de l'Outaouais (STO) récolte des données au moyen des cartes à puce implantées sur son réseau. De plus, la région desservie par la STO a récemment fait une enquête origine-destination et les dates de collecte des données des deux méthodes coïncident. C'est donc une chance unique d'étudier les deux techniques et de les comparer afin d'analyser si les données peuvent être utilisées conjointement.

1.1 Problématique

Depuis quelques années, la STO archive des données CAP et EOD au moyen d'une base de données informatisée. Une partie de ces données ont aussi la particularité d'avoir été prises simultanément avec l'enquête régionale de l'Outaouais, à l'automne 2005.

Autrement dit, ces données représentent une opportunité afin d'étudier la précision et la justesse des données en les comparant entre elles. De plus, les données recueillies par l'enquête origine-destination ont récemment été analysées par le comité TRANS, un regroupement d'organismes qui élabore les enquêtes, et, de surcroît, elles sont encore utilisées pour des fins de modélisation.

Théoriquement, les données EOD et CAP ont donc la prétention de représenter le même contexte, soit le portrait des déplacements des utilisateurs du transport en commun qui sont titulaires d'une carte à puce (Figure 1.1). Cependant, en pratique, plusieurs facteurs peuvent influencer la justesse des données. L'identification de ces facteurs est donc primordiale afin de justifier les éventuels écarts de mesure entre les données provenant des deux sources.

D'une part, les données CAP fournissent de l'information quotidiennement où il est possible d'étudier la variation dans les comportements des usagers et, de l'autre part, les données EOD estiment un jour moyen de déplacements et elles comportent plusieurs caractéristiques, notamment sur les personnes, les ménages et les déplacements. Cependant, très peu d'expérimentations de fusion de données ont été faites et les conditions permettant cette fusion sont encore vagues. De plus, avant d'en arriver là, il faut que les données soient formatées de manière à être compatibles. Dans le cas présent, les bases de données sont différentes tant en terme de format, de types de données, de nom de tables et de champs et de la structure relationnelle. Bref, la complexité et la quantité astronomique des données ne permettent pas une exploitation efficace et rendent difficile la comparaison.

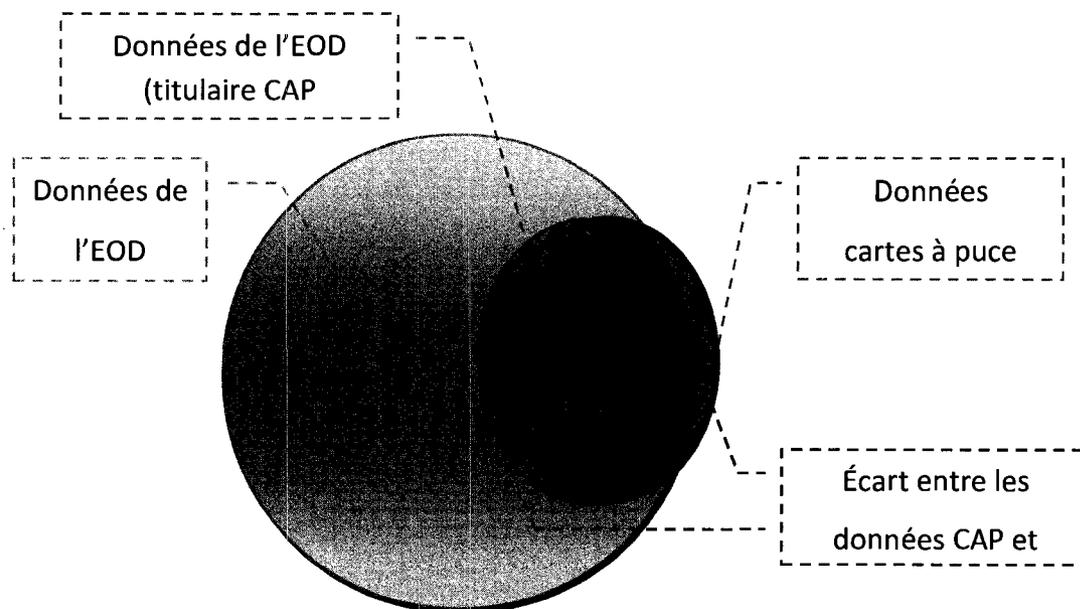


Figure 1.1: Compatibilité et écarts entre les données provenant de sources différentes (CAP et EOD)

Bref, il est difficile de saisir la nature de la possible intégration de ces bases de données. Cependant, il est évident que les forces de l'une pourraient venir combler les faiblesses de l'autre et vice-versa. Le potentiel de l'utilisation simultanée et conjointe des données CAP et EOD pourrait améliorer les techniques actuelles de planification en transport en commun et ainsi permettre une meilleure offre pour la clientèle.

1.2 Buts et objectifs

Les objectifs de ce mémoire gravitent autour de la compréhension des données utilisées afin de façonner une offre de service adéquate dans le domaine des transports en commun. Dans le cadre de cette analyse, il est question des données provenant de l'enquête origine-destination et de celles provenant des systèmes de cartes à puce pour la région desservie par la Société de transport de l'Outaouais.

Il est préalablement nécessaire de se familiariser avec le milieu des transports en commun. Pour ce faire, une revue de littérature, décrivant les concepts et techniques employés couramment, est la première contribution de ce mémoire. Cette revue porte

plus particulièrement sur les données CAP et EOD, les matrices origine-destination et la fusion de données.

L'analyse et la compréhension des différences et des similitudes entre les données provenant d'un système de collecte par cartes à puce et celles provenant d'une enquête origine-destination demeurent le principal objectif de ce mémoire. Naturellement, cette analyse débute par l'étude des systèmes d'informations ayant servi à la collecte des données. Une contribution attendue est la construction d'un nouveau système d'information regroupant les informations, des deux sources de données, qui seront nécessaire pour la suite de l'analyse. Ce système d'information se doit d'être à la fois facile à interpréter et complet (sans perte d'informations). La démarche utilisée afin d'extraire les données recherchées est détaillée. Une méthode de comparaison doit aussi être élaborée et justifiée. Cette méthode se doit d'être à la fois adaptable à d'autres données et assez précise afin de pouvoir tirer des conclusions instructives.

1.3 Structure du mémoire

Outre la section d'introduction ci-présente, le mémoire est composé de six grandes sections. La revue de littérature, deuxième section de ce mémoire, porte sur trois éléments principaux, soit les enquêtes origine-destination, les cartes à puce et la fusion des données (Figure 1.2).

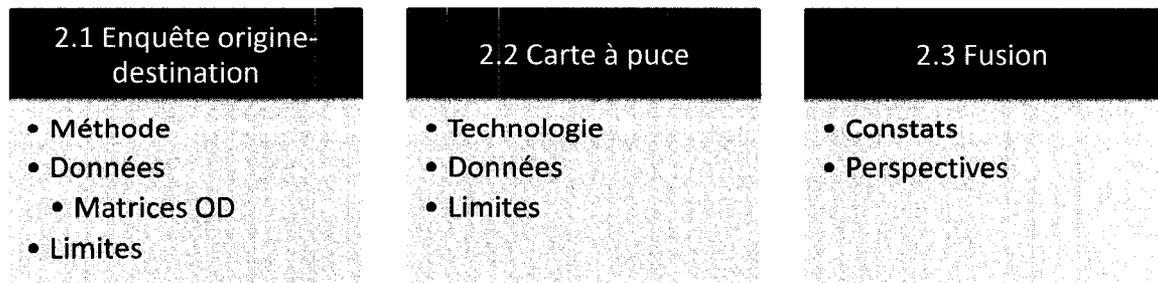


Figure 1.2: Principaux thèmes de la revue de littérature

La troisième section est consacrée à la description de la méthodologie adoptée. S'ensuit à la quatrième section une première analyse des données. En fait, une explication du contexte de la STO est faite en tout premier lieu. Ensuite, une description des données EOD et CAP ainsi que quelques explications concernant les bases de données complètent cette section.

La cinquième section est entièrement dédiée aux analyses et aux résultats. Ces analyses peuvent être divisées en trois grandes catégories, soit l'analyse globale, l'analyse segmentée et l'analyse par réseau. Chacune des analyses est détaillée individuellement et elles comportent une discussion sur les résultats obtenus.

Une discussion relative à l'enquête origine-destination suit à la sixième section. Cette discussion porte sur le rôle des facteurs d'expansion lors de l'interprétation des données de l'enquête. Cette section comportera aussi quelques commentaires sur le potentiel de fusion des données ainsi que sur des améliorations possibles pour les deux méthodes de collecte d'information. Une conclusion, incluant des ouvertures de recherches, terminera ce mémoire à la septième section.

CHAPITRE 2

REVUE DE LITTÉRATURE

Ce mémoire est élaboré sur une problématique touchant trois champs d'études distincts, soit les enquêtes origine-destination, les systèmes de cartes à puce et la fusion de données. La compréhension de ces sujets passe donc par une revue pertinente de la littérature afin, non seulement de se familiariser, mais aussi d'approfondir les concepts et ainsi avoir les connaissances nécessaires pour résoudre la problématique énoncée précédemment. Les trois thèmes, l'EOD, les CAP et la fusion des données, sont donc abordés indépendamment dans cette section.

2.1 Les enquêtes transport

Les méthodes d'enquêtes dans le transport en commun sont variées et combinent des besoins spécifiques en information. Ces méthodes permettent d'obtenir des données nécessaires, entre autres, à la planification des horaires, l'aménagement du territoire, l'estimation des coûts ou encore l'évaluation de la demande. Au Québec, plusieurs techniques d'enquête ont été testées, mais une se démarque des autres par la qualité et la variété de l'information qu'elle permet d'obtenir. Cette méthode, couramment appelée enquête origine-destination, permet d'établir le portrait des déplacements sur un territoire donné, tout en comportant plusieurs informations sur les habitudes de déplacement et sur les caractéristiques démographiques de la population.

L'enquête origine-destination se base sur les principes statistiques d'échantillonnage afin d'estimer l'état ou le portrait de la population ciblée. Naturellement, la grandeur de l'échantillonnage influence directement la précision des données. En fait, Stopher & Greaves (2007) résument cette pensée en disant que : « It is axiomatic that a model can never be better than the data from which it is estimated ». Lors des enquêtes origine-destination de Montréal en 1998 (Morency & Chapleau, 2002) et en 2003 (Secrétariat

aux enquêtes origine-destination métropolitaines, 2003) ainsi que pour la région de la capitale nationale en 2005 (iTRANS Consulting inc., 2006), un échantillonnage d'environ 5 % des ménages a été utilisé. Afin d'arriver à rejoindre autant de ménages, les entrevues des dernières EOD sont faites par téléphone et elles sont assistées par des logiciels, « computer-assisted telephone interview » (CATI), qui permettent une meilleure saisie de l'information et s'assurent d'un certain niveau de qualité des données recueillies (Chapleau, Allard, Trépanier, & Morency, 2001) (Bayart, Bonnel, & Morency, 2008).

Des études (Bonnel & Le Nir, 1998)(Stopher, 1998)(de Leeuw E. D., 1992) ont été faites sur la performance des entrevues téléphoniques, fréquentes en Amérique du Nord, comparativement aux interviews en personne, plus fréquentes en Europe et en Australie. Il ressort que les entrevues téléphoniques sont plus flexibles et moins dispendieuses, mais qu'il est de plus en plus difficile d'obtenir une liste de numéros de téléphone pertinents puisque certains numéros ne sont tout simplement pas disponibles (Bonnel & Le Nir, 1998) ou ils font parti de liste « à ne pas appeler » (Stopher & Greaves, 2007). Actuellement, la popularité des téléphones cellulaires fait aussi en sorte que plusieurs ménages n'ont pas de ligne à domicile et la plupart des enquêtes ne permettent pas encore d'appeler sur les cellulaires (Stopher & Greaves, 2007). Aussi, de plus en plus de mécanismes permettant de filtrer les appels sont accessibles, tel que le répondeur. Ceci limite naturellement l'échantillonnage et, par conséquent, peut amener un biais dans les données enquêtées. C'est néanmoins l'entrevue téléphonique qui demeure la principale méthode utilisée pour les enquêtes origine-destination québécoises.

Ce choix est, entre autres, basé sur les avantages suivants :

- Meilleur suivi et évaluation du personnel engagé pour interviewer (de Leeuw & Van der Zouwen, 1992) (de Leeuw E. D., 1992),

- Tous les employés sont situés au même endroit, il est donc plus facile d'évaluer leurs compétences. De plus, l'utilisation du téléphone permet à un supérieur d'écouter les conversations et ainsi se former une meilleure idée de la compétence de l'interviewer.
- Meilleur suivi des données collectées (de Leeuw & Van der Zouwen, 1992) (de Leeuw E. D., 1992)(Trépanier, Chapleau, & Morency, 2006),
 - Surtout avec l'utilisation de logiciels spécialisés, comme la gamme de logiciels MATIDUC, la vérification des données se fait tout au long de l'interview. En fait, la vérification de la validité des chaînes de déplacements par ménage peut être faite au fur et à mesure que l'interview progresse. De plus, l'origine et la destination des chaînes de déplacements sont maintenant géocodées en coordonnées XY (en mètres), et ce, automatiquement par les logiciels de type « CATI ».
- Distribution plus uniforme des interviewers sur tout le territoire (Bonnell & Le Nir, 1998),
 - Les enquêteurs n'étant pas attirés à une région particulière, il est plus aisé de répartir le biais, causé par des techniques de questionnement douteuses, sur toute la zone enquêtée.
- Accès aux données plus rapidement (de Leeuw & Van der Zouwen, 1992) (de Leeuw E. D., 1992),
 - Les données sont immédiatement inscrites dans une base de données ou un registre. De plus, elles sont vérifiées et validées lors des entretiens téléphoniques, ce qui permet de les obtenir plus rapidement.
- Des coûts moins importants comparativement aux enquêtes ménages faites en personne (Bonnell & Le Nir, 1998),
 - Les enquêteurs n'ont pas besoin de se déplacer, ceci limite donc les coûts et augmente la productivité en réduisant le temps requis pour

rejoindre un répondant. Les enquêteurs sont donc plus efficaces puisqu'ils peuvent contacter plus de ménages, et ce, en moins de temps. Par le fait même, il n'est pas nécessaire d'avoir autant d'enquêteurs pour accomplir le même travail.

Habituellement, les enquêtes québécoises s'étalent sur une période de trois à quatre mois (Trépanier & Chapleau, 2001a) et elles ont lieu principalement à l'automne puisque l'achalandage sur le réseau routier est à son plus haut niveau et relativement stable (Transport Québec, 2007). Généralement, seules les journées en semaine, soit du lundi au vendredi, sont incluses dans l'enquête, bien que des essais aient été faits pour les journées de la fin de semaine. Les enquêtes se font, au Québec, toujours en partenariat avec différents paliers gouvernementaux et organismes municipaux, tel que les sociétés de transport. Ces différents acteurs agissent donc dans le but commun de mieux observer la mobilité des personnes sur un territoire donné. Plus précisément, lors de l'enquête origine-destination montréalaise de 1998, les objectifs élaborés dès le début de l'enquête concernant les applications les plus pertinentes des données à recueillir étaient (Bergeron & Grondines, 1998):

- mesurer des facteurs de mobilité et leur évolution,
- évaluer et analyser le comportement des usagers,
- alimenter les modèles de prévision,
- analyser les interactions entre l'aménagement urbain et les déplacements,
- réaliser des plans de transport,
- réguler et orienter la demande de déplacement,
- gérer les équipements collectifs.

La méthode d'échantillon stratifié systématique en grappe permet de choisir, de manière aléatoire, les numéros à appeler selon des groupes d'âge et des zones

préalablement définies (Transport Québec, 2007). C'est cette méthode qui est utilisée pour les enquêtes ménages du Québec.

Lors des entrevues, il est fréquent qu'une seule personne (le répondant) représente l'ensemble des personnes du ménage (Trépanier, Chapleau, & Morency, 2006). Cette personne sera questionnée sur les déplacements faits, la veille, par l'ensemble du ménage. Au cours de l'entretien, trois dimensions sont enquêtées, soit les déplacements, les personnes et les ménages. Les principales caractéristiques étudiées lors de l'enquête montréalaise de 1998 ainsi que les précédentes sont (Trépanier & Chapleau, 2001a) :

- **Ménages** : Le lieu de résidence, le nombre de personnes résidant à cet endroit, le nombre d'automobiles;
- **Personnes** : L'âge, le sexe, la possession d'une automobile et d'un permis de conduire;
- **Déplacements (effectués la veille)**: le moyen utilisé pour le déplacement, soit
 - Le transport en commun : Modes utilisés, lignes utilisées, points de jonction intermodaux;
 - L'automobile : Pont utilisé (si applicable).

Ces caractéristiques sont amenées à évoluer selon les besoins et les enjeux propres à la période enquêtée. D'autres informations peuvent être demandée telles que, l'heure du déplacement, le titre de transport utilisé, le motif du déplacement, etc. (Trépanier, Chapleau, & Morency, 2006).

L'utilisation de ces données permet, selon Bonnel & Le Nir (1998), de faire des plans de transport, de gérer le système routier, de régulariser la demande sur les axes routiers, de mieux organiser les déplacements entre les différents modes de transport et d'analyser les interactions entre les planifications en transport et les plans d'urbanisme.

Traditionnellement, le but derrière la collecte de ces données est donc d'établir une journée moyenne (Axhausen, Zimmerman, Schönfelder, Rindsfuser, & Haupt, 2002) afin d'analyser les comportements de la population. Cette analyse se fait à l'aide de modèles, bien souvent agrégés, comme la procédure séquentielle classique (Stopher & Greaves, 2007). Mais des approches différentes ont été soumises, dont l'analyse orientée objets et totalement désagrégée (Trépanier & Chapleau, 2001a). Cette approche permet de conserver une précision et une rigueur dans les analyses puisqu'elle n'est pas sujette à des pertes d'information dues à l'agrégation des données.

2.2 L'approche totalement désagrégée

Lors des enquêtes origine-destination, les données brutes recueillies se distinguent par leur relation à l'individu. C'est cette particularité que l'approche totalement désagrégée tente d'exploiter. En fait, les données ne sont pas analysées selon des moyennes zonales, comme pour les approches agrégées, mais bien de manière distincte et ponctuelle. Il est donc possible d'obtenir de l'information sur des déplacements propres à un individu.

L'approche totalement désagrégée permet alors d'obtenir des données différentes des approches classiques. En fait, les données ont la particularité (Trépanier & Chapleau, 2001b):

- d'être codées selon des coordonnées X-Y;
- d'utiliser des échelles de temps, d'espace et de caractéristiques sociodémographiques (âge, sexe, etc.) beaucoup plus précises;
- de garder les déplacements au niveau de l'individu tout en conservant l'information relative au ménage, à la personne, au motif et aux autres variables;

- d'être utilisées à d'autres fins, telles que la cartographie, la planification, les études marketing, etc.

Par conséquent, cette approche nécessite beaucoup plus d'informations et elle exige une qualité supérieure de données. En fait, ce type d'analyse est beaucoup plus sensible aux erreurs de saisie, aux oublis ou au manque d'information (Stopher & Greaves, 2007). Il est donc naturel que la première phase de cette méthode soit la validation des données.

La deuxième phase de cette méthode est le traitement de l'information. Celui-ci se fait en fonction des déplacements faits par des individus. En fait, toutes les caractéristiques déclarées, par exemple l'heure, le motif ou le mode, demeurent associées aux déplacements (Trépanier, Chapleau, & Morency, 2006). Aucune agrégation n'est faite aux données. Ceci permet entre autres d'ajouter et d'améliorer des caractéristiques, tel que le géocodage des origines, des destinations, des ménages et des jonctions (Trépanier, Chapleau, & Morency, 2006).

Plus précisément, le traitement de l'information consiste en une création de liens entre les différentes déclarations d'un individu. Par exemple, lors de l'enquête, les répondants donnent des indicateurs de lieu, comme un code postal, un générateur, une intersection ou une adresse, pour l'origine et la destination d'un déplacement et, ceux-ci, peuvent être traduits en coordonnées cartésiennes afin d'augmenter la précision des données (Trépanier & Chapleau, 2001a). D'autres traitements peuvent avoir lieu, par exemple la traduction des déplacements faits en transport en commun (TC) en itinéraires. En fait, les indicateurs de lieu des déplacements TC sont souvent différents des accès au réseau de transport en commun. Lorsque ces indicateurs sont à proximité d'un accès au réseau, alors il est possible d'unir ces lieux afin d'utiliser un seul et unique nom. Ceci augmente la précision et la clarté des données et permet une uniformisation dans les indicateurs de lieu. Seulement après ce traitement, il est

possible d'établir une corrélation afin d'obtenir une séquence de « nœuds-lignes » qui représentent un itinéraire cohérent et compréhensible (Trépanier & Chapleau, 2001a).

Il en découle donc des fichiers où tous les déplacements sont distincts et auxquels se rapportent des caractéristiques sur le ménage et la personne. Ces fichiers représentent énormément de données puisque celles-ci ne sont pas agrégées; Ces fichiers ne sont pas biaisés par des regroupements d'information et demeurent une représentation fidèle des observations recueillies via l'enquête origine-destination tout étant beaucoup plus précis.

2.3 Modélisation orientée objets

Dans le milieu de l'informatique, plusieurs langages de programmation sont dit orientés objets, tels que Visual Basic. Ces langages se démarquent des autres par le fait qu'ils sont composés d'objets ou, à proprement parler, d'idées, de solutions, de représentations virtuelles d'objets concrets. À ces objets sont attachées certaines propriétés qui décrivent ou explique l'état de l'objet. De plus, ces objets peuvent interagir entre eux au moyen de méthodes.

En transport, une approche basée sur le même principe a fait son apparition, soit l'analyse orientée objet (Trépanier & Chapleau, 2001a). Similairement au milieu informatique, ce type de modélisation inclut des objets, des propriétés et des méthodes. Quatre grandes catégories d'objets ont été identifiées (Chapleau, Trépanier, & Allard, 1998), soit :

- Les objets statiques : ce sont des objets qui ne varient pas dans le temps et l'espace.
 - Exemple : Les codes postaux.
- Les objets dynamiques : ce sont des objets qui créent le mouvement ou le besoin de se déplacer.

- Exemple : Les personnes.
- Les objets cinétiques : ce sont des objets qui ont pour but de décrire comment le déplacement a lieu.
 - Exemple : Le trajet.
- Les objets systémiques : ce sont des objets qui alimentent, aident ou répondent aux déplacements.
 - Exemple : Le réseau routier (opérationnel), les enquêtes (informationnel) ou les banlieues (multifonctionnel).

Les fichiers obtenus suite à l'enquête sont totalement désagrégés et, par conséquent, il est facile d'appliquer les principes et les concepts de l'approche orientée objet. En fait, la Figure 2.1 résume les principaux objets présents dans les fichiers.

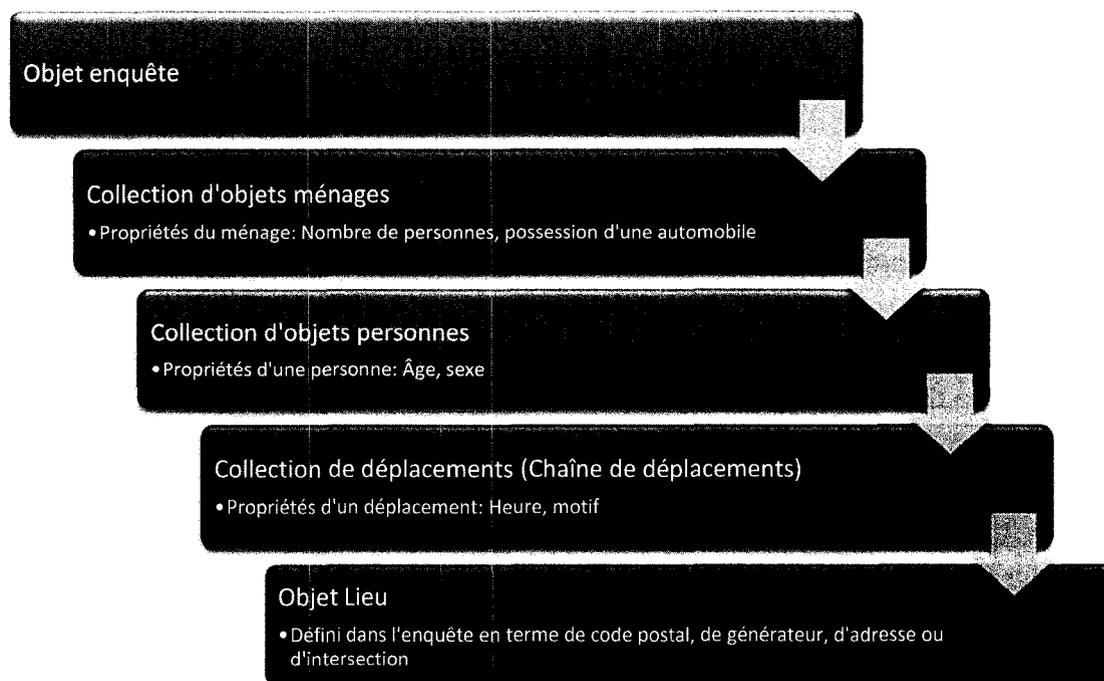


Figure 2.1: Objets et collection d'objets de l'enquête origine-destination (inspiré de Trépanier & Chapleau (2001a))

L'objet enquête est a priori défini par plusieurs objets de type ménage. Ce regroupement d'objets est couramment appelé collection. La collection d'objets

ménages représente donc l'ensemble des ménages enquêtés. Chacun de ces objets a des propriétés spécifiques, par exemple le nombre de personnes dans le ménage ou encore le nombre d'automobiles. Pour un ménage donné de la collection, une collection d'objets personnes peuvent le décrire. En fait, il y a autant d'objets personnes que le nombre de personnes dans le ménage. On parle ici de relation d'un à plusieurs. Certaines propriétés décrivent aussi l'objet personne, tel que l'âge ou le sexe.

Naturellement, une personne peut faire ou ne pas faire de déplacement au cours d'une journée. Par conséquent, la relation entre un objet personne et la collection de déplacements est de type un à soit zéro ou soit un (0-1). Cette collection de déplacements, communément appelée chaîne de déplacements, comprend autant d'objets que le nombre de déplacements que la personne a déclaré. Chacun des objets déplacements est donc caractérisé par des propriétés comme l'heure ou le motif. Finalement, l'objet déplacement a une relation de type un à deux avec des objets de type lieux. Ceci est dû au fait que pour tous les déplacements, une origine et une destination sont exigées lors de l'entrevue. Ces objets lieu représentent donc une donnée servant à identifier spatialement le déplacement.

L'avantage d'utiliser une méthode orientée objet est qu'il est facile d'enrichir les propriétés des objets. En fait, plusieurs enrichissements sont possibles, notamment (Trépanier & Chapleau, 2001a):

- le géocodage des déplacements;
- la validation des informations saisies;
- la détermination de l'accès au TC et de chargement;
- la simulation d'itinéraires.

Ceci permet d'obtenir des données, uniformes et précises, correspondant mieux aux besoins des analyses subséquentes. Lors de l'enrichissement des propriétés, il est

possible que des objets soient créés. En fait, il existe trois types de nouveaux objets, soit (Trépanier & Chapleau, 2001a) :

- les objets créés;
- les objets dérivés;
- les objets associés.

Les objets créés sont des objets qui préalablement n'existaient pas et qui sont faits dans le but d'apporter de l'information supplémentaire ou encore préciser davantage l'information. Les objets dérivés sont des objets qui s'imposent, suite à une analyse des données, comme étant importants, par exemple les générateurs de déplacement. Finalement, les objets associés représentent des objets qui servent de liens ou de supports à d'autres objets, comme le réseau routier.

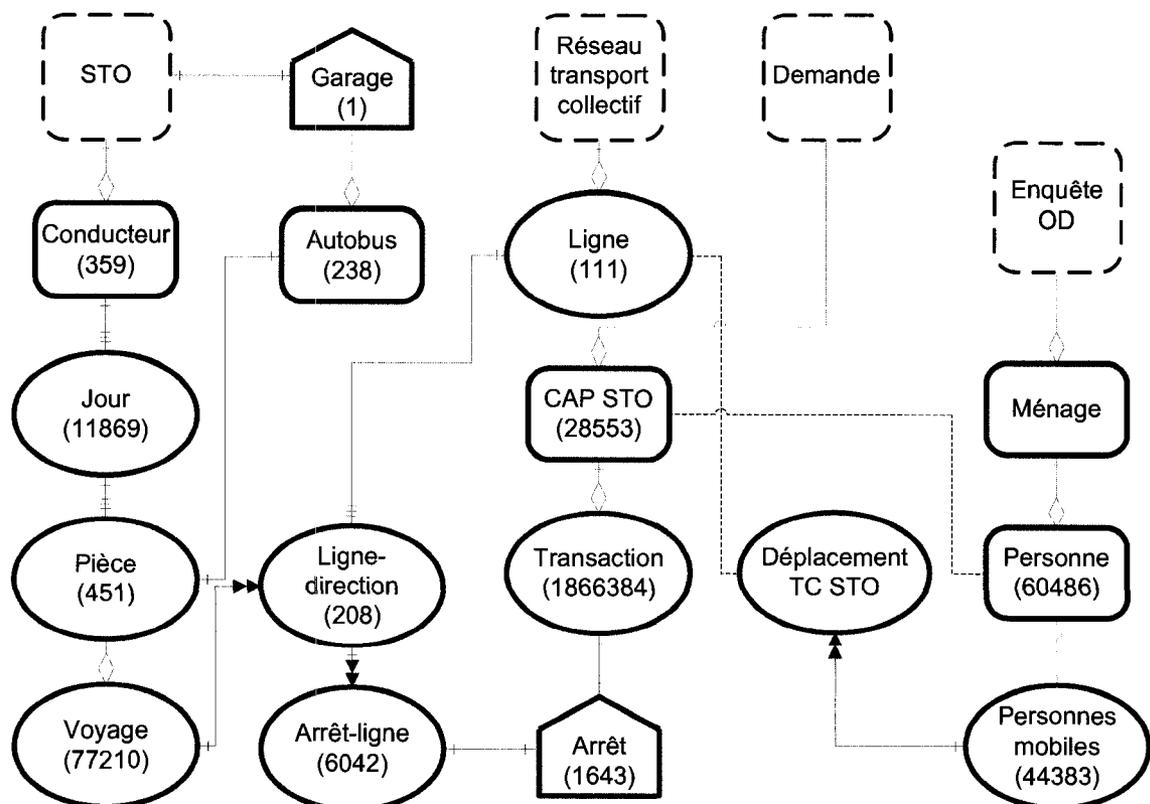


Figure 2.2: Modèle orienté objets de la STO (inspiré de Trépanier & Chapleau (2001a))

La Figure 2.2 représente les différents objets étudiés et créés à partir des données de l'enquête origine-destination faite sur le territoire de la STO en 2005. Sur cette figure, on remarque que les objets qui sont entourés de traits pointillés représentent des objets dits systémiques. Les cercles représentent des objets cinétiques, tandis que les rectangles aux coins arrondis indiquent des objets dynamiques. Les objets « garage » et « arrêt », entourés d'un pentagone, sont définis comme étant des objets statiques.

Ce modèle, bénéficiant de l'approche totalement désagrégée, améliore donc énormément la qualité et la pertinence de données des enquêtes. En fait, cette méthode a non seulement fait ses preuves à plusieurs reprises au Québec (Trépanier, Chapleau, & Morency, 2006), mais elle s'avère une technique hautement performante et elle laisse une porte ouverte à la venue de nouvelles sources d'informations, comme les données provenant de cartes à puce.

2.4 Carte à puce

Les débuts de la carte à puce remontent à aussi loin que 1947, alors qu'un dénommé Geoffrey Dummer, un ingénieur britannique, créa une mémoire portative (Electronic product news, 2005). L'idée a ensuite été améliorée par plusieurs chercheurs, dont les allemands Dethloff et Gröttrup (Trépanier, Chapleau, & Tranchant, 2007) ainsi que le français Moreno (Portalis, 2007) (Shelfer & Procaccino, 2002). C'est d'ailleurs ce dernier qui fera breveter, en 1974, un prototype de carte à puce ayant des moyens inhibiteurs totalement logés sur la puce. Par contre, la commercialisation prendra encore quelque temps avant de se concrétiser. Les deux principales entreprises au cœur de cette commercialisation sont Honeywell Bull, avec le SPOM (« Self Programmable One-chip Microcomputer ») et Motorola, avec le CP8 (« Circuit Portatif des années 80 ») (Guillou, 2004).

Aujourd'hui plusieurs entreprises commercialisent des cartes à puce et elles se destinent à de multiples secteurs de l'industrie. En fait, les cartes à puce sont utilisées entre autre lors de transactions financières, de la vérification d'identité ainsi que lors d'accès à des renseignements médicaux (Blobel, Pharow, Spiegel, Engel, & Engelbrecht, 2001). C'est la facilité et la rapidité avec laquelle les cartes à puce peuvent traiter des quantités importantes d'information qui les rendent si intéressantes. D'ailleurs, c'est pourquoi elles sont de plus en plus présentes dans le milieu du transport en commun. Les cartes à puce permettent, non seulement, d'améliorer la vitesse à laquelle les transactions s'effectuent, mais elles fournissent aussi de nouveaux outils pour la planification et la collecte de données (Allison, Iseki, Taylor, & King, 2006).

Il existe trois grandes catégories de cartes à puce, soit les cartes sans contact, avec contact et les hybrides. Les cartes sans contact, comme leur nom l'indique, permettent un échange d'information, entre un lecteur et la carte, sans que ces deux entités se touchent. La lecture de la carte à puce se fait au moyen d'ondes radio encodées communément appelées « RFID » (« radio frequency identification »). Ces ondes sont envoyées par le lecteur vers la puce si celle-ci se retrouve à une distance spécifique (variant entre 10 cm et plus de 50 cm). Ce signal est ensuite traduit en courant continu et en courant alternatif qui iront alimenter d'une part la puce (courant continu) et d'autre par l'horloge de synchronisation (courant alternatif). La puce se synchronisera au lecteur durant tout l'échange d'information. Cet échange se fait très rapidement si les calculs sont peu complexes. On parle de l'ordre d'un cinquième de seconde (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003). De plus, avec l'avancement des microprocesseurs, ces cartes à puce sont en mesure de compléter rapidement tous les calculs exigés, même les calculs de sécurité imposés par les banques (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003).

En fait, bien que la technologie sans contact satisfasse les exigences de sécurité des banques, la majorité de celles-ci utilisent encore uniquement le second type de carte, soit les cartes à puce avec contact. Ces cartes nécessitent donc que le détenteur insère dans un lecteur sa carte afin de permettre l'échange des informations. Cette carte possède donc plusieurs « doigts en or » qui permettent le contact entre le lecteur et la puce. Le terminal peut donc lire et écrire librement de l'information sur la carte. Naturellement, ceci nécessite que l'utilisateur insère la carte dans le lecteur et la retire dans les délais convenus par les calculs requis. Ceci convient particulièrement aux longs calculs puisque le lecteur ne remet pas la carte avant que tous les calculs aient été complétés, contrairement au premier type de technologie où l'échange d'information peut se rompre avant la fin des calculs si la personne sort du champ d'action du terminal.

Il existe aussi un type de carte à puce appelé hybride. Cette carte peut avoir deux architectures différentes, soit une seule puce composée d'un processeur double ou soit deux puces distinctes composées d'un microprocesseur chacune. Naturellement, le but de ce croisement entre les deux technologies précédentes est de permettre un échange d'information avec contact et sans contact sur la même carte. Le fonctionnement de chacune de ces technologies demeure le même.

Pour toutes les cartes comprises dans les trois grandes catégories décrites précédemment, il existe deux modèles de base, soit la carte mémoire et la carte microprocesseur (McDonald, 2000). Le type « mémoire » sert uniquement à stocker de l'information sans toute fois être en mesure de la traiter sur la carte directement (des calculs peuvent, par contre, être fait sur l'ordinateur relié au système de lecture). Le type « microprocesseur » diffère par son autonomie à gérer les fichiers et les programmes sur la carte. En fait, ce type de carte peut facilement contenir plusieurs informations sécurisées, provenant de sources différentes, qui sont gérées et traitées par des programmes roulant localement (Blythe P. , 2004).

Les applications des cartes à puce sont nombreuses et touchent à plusieurs secteurs de l'économie. Les entreprises en télécommunication ont été des pionnières en termes de carte à puce. En fait, depuis les années 1980, plusieurs entreprises de téléphonie utilisent les cartes à puce comme moyen pour effectuer les transactions avec leurs clients qui désirent utiliser les téléphones publics.

Les banques, principalement européennes et asiatiques, ont emboîté le pas à l'industrie des télécommunications peu de temps après. Les cartes à puce ont donc comme rôle de permettre les transactions financières afin de débiter les comptes lors d'achats, de paiement, etc.

Les cartes à puce sont aussi maintenant grandement utilisées afin d'identifier des personnes et de restreindre l'accès à certains endroits. Plusieurs entreprises utilisent donc les cartes à puce comme système permettant l'identification de leurs employés et, selon les accès programmés sur la puce, leur donne le droit d'accéder à différents secteurs de l'entreprise. L'utilisation des cartes à puce comme gestionnaire des feuilles de temps est aussi très fréquent dans les entreprises. En fait, pour les entreprises, la carte à puce permet l'identification, la restriction ainsi que le suivi, et ce, grâce à un seul et unique système de reconnaissance.

Un peu à l'image des compagnies de téléphone, les sociétés de transport se servent aussi de plus en plus des cartes à puce afin d'améliorer leurs systèmes de paiement. De plus, ces cartes leur permettent d'obtenir plusieurs informations pertinentes sur les déplacements des titulaires. Il est aussi de plus en plus fréquent de voir cette technologie créer des ponts entre différentes sociétés de transport et ainsi améliorer l'interopérabilité des systèmes de paiement.

Les cartes à puce sont donc omniprésentes dans les entreprises et dans la vie quotidienne de bien des personnes. Mais adopter ce type de technologie comporte des enjeux importants qui souvent viennent justifier ce choix.

2.4.1 Enjeux et intérêts

Les principaux enjeux et intérêts des cartes à puce, comparativement aux cartes à bande magnétiques et autres technologie de type « mémoire portative », sont :

- la flexibilité du traitement de l'information (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003) (Iseki, Allison, & Taylor, 2007);
- le traitement multi tâche et la rapidité des transactions (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003)(Blobel, Pharow, Spiegel, Engel, & Engelbrecht, 2001);
- le stockage de données (Allison, Iseki, Taylor, & King, 2006)(Blythe P. , 2004)(Bagchi & White, 2005);
- la sécurité et la protection des informations contenues sur la carte à puce (Blythe P. , 2004)(Clarke, 2001);
- le coût des infrastructures et des cartes (Turban & Brahm, 2000)(Allison, Iseki, Taylor, & King, 2006)(Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003);
- la réutilisation ou la reprogrammation des cartes à puce afin de prolonger leur durée de vie (Business Wire, 1999)(Baladan, 2002)(Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003);
- l'intégration possible à d'autres technologies de pointe (Blythe P. , 2004)(Card Technology Today, 2002);
- l'interopérabilité des systèmes de paiement entre différentes sociétés de transport partageant une même région ou ayant des points de correspondance avec d'autres réseaux (Allison, Iseki, Taylor, & King, 2006)(Blythe P. , 2004)(Iseki, Allison, & Taylor, 2007);
- la fiabilité de la technologie et les possibles erreurs de fonctionnement (Allison, Iseki, Taylor, & King, 2006)(Blythe P. , 1998).

Les cartes à puce, de par leur architecture physique, permettent une très grande flexibilité dans le traitement des informations. En fait, la mémoire, accessible en lecture et en écriture, est la composante maîtresse qui fait la grande polyvalence de cette technologie et qui permet autant d'utilisations différentes. Cette polyvalence est d'autant plus importante que les besoins des entreprises et des sociétés sont continuellement en changement.

Sa flexibilité associée à une grande capacité de traitement de l'information augmente considérablement l'intérêt des milieux d'affaires envers cette technologie. En fait, avec l'avancement des microprocesseurs, les cartes sont de plus en plus performantes et peuvent, du même coup, travailler sur plusieurs tâches simultanément. C'est d'autant plus important puisque plusieurs tâches sont souvent nécessaires, tel que l'identification, l'écriture et la lecture de données, le décryptage, etc. Ce traitement des calculs et l'échange des informations se fait très rapidement facilitant ainsi l'implantation de cette technologie dans des milieux où la vitesse des transactions est directement reliée à la fiabilité du service offert, par exemple en transport en commun (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003) ou dans les milieux médicaux (Blobel, Pharow, Spiegel, Engel, & Engelbrecht, 2001).

Dans le même ordre d'idée, la grande capacité de stockage de données permet l'implantation de différentes stratégies tarifaires, de systèmes de sécurité accrue ou encore de suivi de transactions. Par exemple, des tarifications en fonction du temps, de l'espace, du mode de transport et des réseaux utilisés peuvent facilement être gérées sur une seule et unique carte à puce (Allison, Iseki, Taylor, & King, 2006). Cette grande capacité de stockage de données peut aussi être utilisée afin de fidéliser la clientèle (Blythe P. , 2004) ou encore afin de connaître davantage les habitudes d'utilisation de la carte. Ceci est le cas en transport en commun puisque plusieurs données concernant les déplacements des titulaires de carte sont inscrites au fur et à mesure dans le

système. Ces données représentent une nouvelle source d'informations et peuvent venir compléter ou même supplanter les méthodes traditionnelles de collecte de données (Bagchi & White, 2005).

La sécurité de l'information qui transige par les cartes à puce est aussi un enjeu important puisque de plus en plus le respect des lois sur la protection de la vie privée est un sujet grandement étudié et débattu dans notre société. La sécurité peut être décomposée en trois catégories (Blythe P. , 2004), soit :

- physique : ingénierie inversée (*reverse engineering*) des systèmes de collecte de données, la carte elle-même, erreurs de manipulation;
- logique : l'encryptage;
- interface : lecteurs truqués, interception de transfert,

Heureusement, les cartes à puce possèdent plusieurs composantes de sécurité qui sont parmi les plus sophistiqués sur le marché (Blythe, 2004). Cependant, ce n'est pas tant le vol d'identité qui inquiète que l'utilisation faite avec les données. En fait, les cartes à puce doivent être réglementées, par différentes instances dont les gouvernements, puisqu'elles ont le potentiel de retracer et d'identifier des individus, grâce à leurs habitudes de déplacements (Clarke, 2001). Par conséquent, des accords de confidentialité très stricte doivent être élaborés afin de protéger ces données.

Le coût associé aux systèmes de carte à puce est une barrière importante à leur implantation. De plus, l'analyse des coûts et des bénéfices est parfois difficile à réaliser et à justifier (Turban & Brahm, 2000) et la littérature est encore pauvre en analyse des retombés économiques possibles de l'implantation de cette technologie (Allison, Iseki, Taylor, & King, 2006). Plusieurs éléments font varier les coûts d'implantation, tel que la dimension du réseau, le nombre d'utilisateurs par journée, le type de service offert, etc. Par exemple, le projet *SmarTrip* de Washington, implanté en 1999, a été évalué à 20 millions \$US uniquement pour le capital investi tandis que le projet *GoVentura* a

couté 1,8 million \$US (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003). Il faut aussi prendre en considération d'autres coûts, tel que l'entretien, la formation des employés et la désuétude de l'ancien matériel. Ces coûts diffèrent énormément d'une société de transport à l'autre et, par conséquent, il est ardu de prévoir les bénéfices ou les pertes monétaires que l'implantation d'un tel système pourrait avoir. De plus, certains coûts peuvent être considérés comme inévitables, par exemple, un système de paiement, qui devait être changé préalablement à l'implantation des cartes à puce, ne pourra pas être entièrement imputé aux coûts du nouveau système. Somme toute, de plus en plus de sociétés implantent cette technologie malgré les coûts inhérents et, par conséquent, de plus en plus d'études sur les bénéfices font leur apparition. Un exemple est l'implantation des cartes à puce aux Pays-Bas où « l'évaluation avec des hypothèses plausibles démontrent que l'introduction des cartes à puce peut être un investissement profitable » (Cheung, 2006).

Les bénéfices que peut apporter une telle technologie sont aussi difficiles à cerner qu'à chiffrer. En effet, il est problématique de quantifier les revenus associés, entre autres, aux nouvelles données qui serviront principalement à améliorer la planification et, par le fait même, la fiabilité et l'efficacité du réseau. Ce sont ces ambiguïtés ainsi que les différences entre les sociétés de transport qui font en sorte que le calcul des coûts et bénéfices est si complexe et imprécis. Cependant, plusieurs sociétés de transport ont adopté cette technologie, malgré ces coûts importants.

Un autre point débattu est la réutilisation ou le recyclage des cartes. Le coût d'achat des cartes à puce varie entre 1,50 \$US et 10 \$US comparativement aux cartes magnétiques, couramment utilisées en TC, qui coutent entre 0,02 \$US et 0,10 \$US (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003) (Balaban, 2007). Plusieurs pratiques sont donc mises de l'avant afin que les sociétés de transport puissent offrir les mêmes services sans toutefois augmenter leur

frais d'exploitation. Par exemple, la régie autonome des transports parisiens, la RATP, a implanté des billets, les « C tickets », s'apparentant aux cartes à puce sans contact. En fait, ces billets sont destinés à des utilisations de courtes durées étant donné qu'ils ne possèdent qu'une petite unité de mémoire. Les avantages de ce type de billets est qu'ils sont sans contact et, surtout, coute une fraction du prix des cartes à puce, soit entre 0,20 \$US et 0,40\$US (Baladan, 2002). D'autres sociétés de transport ont développé des méthodes de recyclage (Business Wire, 1999), mais il est certain que tant que les prix ne diminueront pas, la question de la réutilisation des cartes à puce sera débattue.

L'intégration de technologies de pointe représente un défi considérable pour les fabricants de cartes à puce et un attrait important pour le transport en commun. D'une part, il y a les producteurs de cartes à puce qui n'intègrent pas nécessairement des technologies compatibles aux cartes à puce (Blythe P. , 2004) et d'autre part, il y a les sociétés de transport qui désirent offrir à leurs clients plusieurs moyens différents pour payer leur solde, acheter des titres, etc. En fait, l'utilisation conjointe de technologie demeure en progression et certains leaders dans le domaine de la fabrication de carte à puce, comme Walter Bonneau de Cubic, pense que l'utilisation de site web et de téléphones cellulaires permettra sous peu d'effectuer des transactions avec les systèmes de cartes à puce (Card Technology Today, 2002).

Il est fréquent que des sociétés de transport distinctes offrent des services sur des territoires appartenant à un autre réseau de transport en commun. Dans certain cas, ces sociétés gagnent à développer des ententes, sur la technologie utilisée ou encore sur les prix, afin de faciliter les correspondances entre les différents réseaux qui desservent la même région. Comme il a été mentionné précédemment, certaines cartes à puce sont dotées de microprocesseurs très performants qui permettent plusieurs traitements simultanés et, par conséquent, peuvent être des outils importants afin de faciliter cet interopérabilité entre différentes sociétés de transport

(Iseki, Allison, & Taylor, 2007). Par contre, les technologies utilisées par les cartes à puce sont pratiquement aussi variées que le nombre d'entreprises produisant des cartes (Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc., 2003). Cette diversité n'est cependant pas un obstacle insurmontable pour parvenir à une interopérabilité entre réseaux. Un brillant exemple de coopération est celui de l'Angleterre où, en 1999, a été créé la « ITSO » ou « Integrated Transport Smart-Card Organization » qui a pour but d'unir les différents acteurs au sein d'un même groupe et de permettre une uniformisation des technologies et un environnement plus propice à l'interopérabilité (Blythe P. , 2004). Il est évident que les cartes à puce ne peuvent pas à elles seules assurer la compatibilité entre les réseaux puisqu'il existera toujours des divergences institutionnelles (Allison, Iseki, Taylor, & King, 2006). Mais cette technologie a le bénéfice de simplifier grandement tout le côté technique et, par le fait même, d'améliorer les échanges entre les sociétés et de faciliter les déplacements des usagers entre les réseaux.

La fiabilité demeure un élément important de tout système de paiement. La littérature suggère que la fiabilité globale du système de paiement par carte à puce est fonction de la fiabilité de la carte (Blythe P. , 1998) elle-même et du matériel (logiciels, lecteurs, etc.) permettant les transactions (Allison, Iseki, Taylor, & King, 2006). En fait, les répercussions d'un problème technique ou informatique relié au système de paiement par carte à puce sont nécessairement plus coûteuses, en terme de manque à gagner (\$), comparativement aux systèmes plus traditionnels (Allison, Iseki, Taylor, & King, 2006). Ceci est principalement dû à la centralisation des opérations. Les systèmes traditionnels n'ont pas ce problème puisque, pour chaque autobus ou station de métro, le paiement se fait à l'aide d'un système qui est totalement indépendant des autres. Ceci n'est évidemment pas le cas des systèmes de paiement par carte à puce puisque tous les lecteurs, qui permettent les transactions, sont reliés entre eux par un système de gestion. Si ce système de gestion tombe en panne, comme à Londres

(Allison, Iseki, Taylor, & King, 2006), tous les autobus ou toutes les stations comportant des lecteurs seront affectés. Heureusement, ces systèmes sont considérés par plusieurs auteurs, dont Allison, Iseki, Taylor, & King (2006), comme étant très fiables. De plus, l'adoption de systèmes totalement décentralisés, où les cartes, les lecteurs et les serveurs sont totalement indépendants les uns des autres, permet d'isoler les pannes et ainsi d'augmenter considérablement la fiabilité globale du système de paiement (Shiibashi, Yajima, Xiaodong, & Mori, 2006).

2.4.2 Données TC

Au fil des années, plusieurs sociétés de transport à travers le monde ont décidé de se tourner vers l'implantation de systèmes de carte à puce afin d'améliorer leurs méthodes de paiement ou encore d'élargir leur offre (McDonald, 2000). Parallèlement à ces améliorations, ces sociétés de transport se retrouvent avec de nouvelles données, sur les déplacements de leurs usagers, provenant des enregistrements sur les cartes à puce. Ces données brutes ont donc la particularité d'être prises en continue et sur de très longues périodes de temps. Cependant, leur potentiel est encore peu connu et il y a toujours place à des améliorations d'élargir l'utilité de ces données (Bagchi & White, 2005).

Selon Bagchi et White (2005), les données provenant des cartes à puce sont caractérisées par :

- de très grand volume de données;
- la possibilité de lier les données aux utilisateurs ou aux cartes;
- la possibilité de collecter les données sur de longues périodes et en continue;
- la possibilité de reconnaître les utilisateurs les plus fréquents.

Ces caractéristiques peuvent différer légèrement selon le type d'implantation réalisé. Par exemple, l'utilisation des cartes à puce de la société de transport *Chicago Transit*

Authority nécessite une inscription. Lors de cette inscription, des données personnelles, tel que l'âge, le sexe et les revenus, sont recueillies (Utsunmiya, Attanucci, & Wilson, 2006). Ceci procure des informations démographiques sur les titulaires des cartes à puce, ce qui n'est pas le cas pour d'autres sociétés de transport. En fait, Blythe (2007) estime que les sociétés de transport pourraient bénéficier de cette information afin d'établir un meilleur portrait de la demande et ainsi promouvoir l'utilisation du transport en commun.

Plusieurs auteurs considèrent donc l'apport actuel des cartes à puce comme une source de données complémentaire aux techniques actuelles de planification. En fait, les données provenant de cartes à puce permettent un suivi continu des déplacements et elles ont le potentiel de décrire la variabilité quotidienne sur le réseau (Morency, Trépanier, & Agard, 2007). Il est aussi envisageable de segmenter les données de manière à obtenir des groupes homogènes qui permettent une analyse plus rigoureuse. Le cas étudié par Morency, Trépanier & Agard (2007) montre bien que ce potentiel n'est pas utopique puisqu'ils arrivent à établir des liens entre les données tout en faisant ressortir des indicateurs pertinents sur la variabilité des déplacements quotidiens des usagers, et ce, avec uniquement les montées à bord des autobus de la société de transport de l'Outaouais.

Ces données permettent aussi de retracer les déplacements d'un individu pour toute une journée en particulier, malgré le fait que souvent seules les origines des déplacements soient disponibles. En fait, le niveau de détail de l'information présent dans les bases de données permet de reconstruire fidèlement les voyages effectués. Un traitement préalable des données est, par contre, nécessaire avant d'estimer les destinations grâce à des algorithmes performants (Trépanier, Chapleau, & Tranchant, 2007), et ce, même si les destinations ne font pas partie intégrante des données recueillies.

Bien que les bases de données provenant de cartes à puce soient souvent riches en détail sur les déplacements, elles sont néanmoins très pauvres en informations concernant directement l'utilisateur, telles que le motif du déplacement (Bagchi & White, 2005)(Bryan & Blythe, 2007). C'est donc là où l'utilisation parallèle d'autres sources de données prend tout son sens.

2.5 La fusion des données

Il existe plusieurs méthodes permettant de recueillir des données sur les habitudes de déplacement. Ces données se distinguent les unes des autres de par leur taille d'échantillonnage, leur population visée, leur complexité, leurs besoins à combler, etc. Autrement dit, il existe des sources de données différentes qui apportent des informations bien souvent complémentaires. Cependant, ces données sont habituellement analysées indépendamment et il est difficile de les lier ensemble. Ceci soulève bien des interrogations au niveau de l'intégration de ces données. Il est donc tout à fait naturel d'évaluer la faisabilité et la pertinence de fusionner ces données afin d'obtenir des informations bonifiées, précises et plus complètes. Il n'existe cependant que très peu d'informations, en transport en commun, sur la méthodologie à adoptée afin de fusionner les données. Par conséquent, les prochaines sections seront plutôt consacrées à l'étude du potentiel de la fusion et aux perspectives d'utilisation.

2.5.1 Constats

Les modèles de planification d'aujourd'hui nécessitent énormément de données sur des aspects sociaux et des déplacements (Bayart, Bonnel, & Morency, 2008). Ce besoin grandissant en données exige des méthodes d'enquêtes encore plus rigoureuses. Cependant, plusieurs auteurs (Stopher & Greaves, 2007)(Bonnel & Le Nir, 1998)(Venigalla, 2004) notent qu'il y a davantage de non-réponses lors des enquêtes. En fait, il a été mentionné précédemment que plusieurs moyens existent maintenant pour ne pas répondre aux enquêtes et qu'ils sont fréquemment utilisés. D'autres facteurs

expliquent aussi l'augmentation de la non-réponse, par exemple la peur de divulguer des informations personnelles (Bayart, Bonnel, & Morency, 2008), le manque de temps, le désintéressement ou encore parce que les personnes ne sont pas présentes à leur domicile (Stopher & Greaves, 2007). Par conséquent, ceci laisse un vide à combler et, selon Bayart, Bonnel & Morency (2008), tout indique que cette tendance va se poursuivre.

Ceci pose un problème important, puisque le taux de non-réponse influence directement la précision des enquêtes (Stopher & Greaves, 2007). Pour compenser cette tendance, plusieurs auteurs croient que la fusion de sources de données différentes permettrait de combler les besoins actuels en matière d'information sur les déplacements d'une population.

À première vue, cette intégration semble pertinente, d'autant plus que la fusion de données est couramment utilisée dans d'autres domaines, tel que la micro simulation. Mais des questions sont soulevées sur la manière de procéder. Les différences entre les données peuvent être énormes et ceci vient compliquer davantage l'intégration. En fait, Venigalla (2004) souligne qu'il existe souvent des différences sur la composition, l'échantillonnage et le contenu des sources de données et que plusieurs tests statistiques devront être appliqués sur les données afin de s'assurer qu'un biais n'est pas introduit. Venigalla (2004) souligne aussi que si les deux bases de données ont des éléments en communs ils risquent fort de ne pas être identiques et des calculs de normalisation seront nécessaires. Cela vient complexifier davantage l'intégration des sources et ne fait que souligner qu'une méthodologie rigoureusement doit d'abord être élaborée et testée afin de mieux définir les règles d'intégration des données.

2.5.2 Perspectives

Venigalla (2004) définit la fusion des données comme étant un processus d'intégration de deux ou plusieurs bases de données afin d'obtenir les caractéristiques voulues ou

une seule et unique base de données. De cette définition, on peut déduire que des sources d'informations, provenant de contextes non reliés au transport en commun, peuvent être utilisées conjointement aux données TC. En fait, Bayart, Bonnel & Morency (2008) croient que des données provenant de sources, telles que le suivi des téléphones cellulaires, pourraient être intégrées aux données TC, si les accords de confidentialité le permettent.

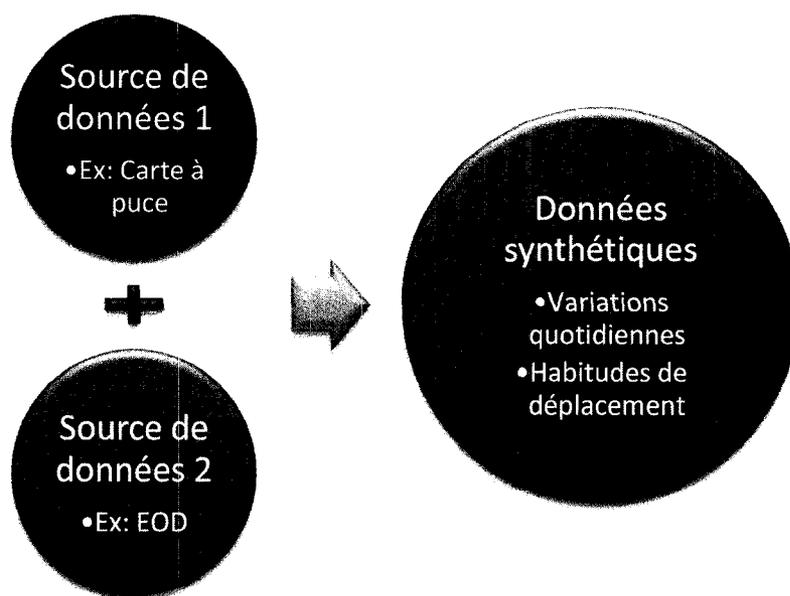


Figure 2.3: Intégration de sources de données distinctes

Par exemple, la Figure 2.3 montre que la fusion de données CAP, riches en informations sur les déplacements quotidiens, et la source de données EOD, comportant énormément d'information sur les habitudes de déplacements et sur la démographie de la région étudiée, pourrait aboutir à une source de données, créée artificiellement, ayant les caractéristiques recherchées des deux sources de données.

La fusion des données s'impose donc comme une solution qui pourrait, dans un avenir rapproché, venir combler un manque d'information et ainsi permettre d'améliorer les outils de planification en transport en commun.

CHAPITRE 3

MÉTHODOLOGIE

La méthodologie utilisée afin d'atteindre les objectifs et de développer les contributions de ce mémoire est présentée dans cette section. Les concepts généraux derrière chacune des étapes de la méthodologie sont décrits à la Figure 3.1. On y retrouve aussi les numéros de sections et de sous-sections correspondants.

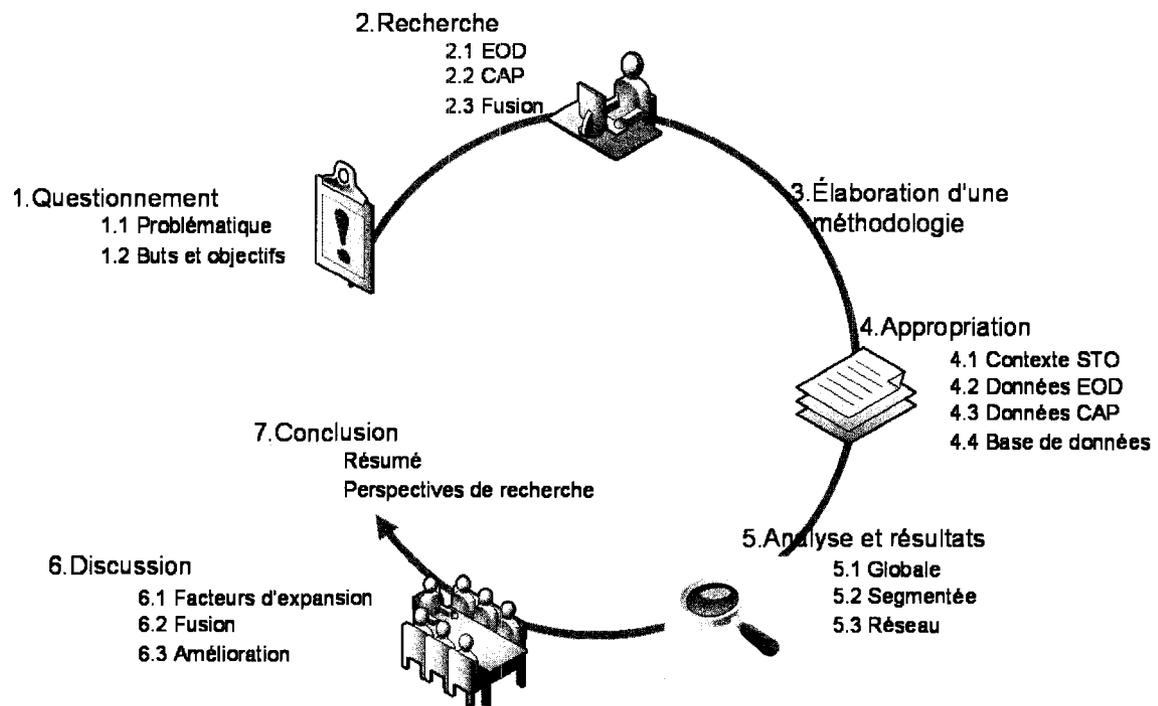


Figure 3.1: Concepts de la méthodologie adoptée

L'identification et la description de la problématique sont le point de départ de ce mémoire. Mais afin de l'énoncer clairement, une première recherche d'informations est nécessaire. C'est donc après s'être familiarisé avec le contexte et la terminologie entourant le transport en commun, que l'élaboration de la problématique a eu lieu. La section 1 décrit tous les éléments qui seront à la base de la réflexion faite dans ce mémoire.

Cependant, une simple familiarisation avec le contexte n'est pas suffisante. Il faut approfondir et élargir le champ d'études afin de bien comprendre les enjeux et ainsi obtenir une réflexion plus détaillée et pertinente. L'étude des données carte à puce et des enquêtes origine-destination demandent aussi de bien maîtriser les concepts et les termes propres au transport en commun et plus particulièrement à la planification et à l'organisation des réseaux. C'est donc pour répondre à ces exigences qu'une revue de littérature a été faite afin de clarifier les termes utilisés dans ce mémoire, ainsi que d'ajuster la ligne directrice des analyses pour mieux correspondre aux besoins actuels des sociétés de transport.

La problématique et le contexte étant maintenant établis, la prochaine étape consiste en l'analyse des données. Les bases de données provenant de l'enquête OD et des systèmes de cartes à puce, gracieusement fournies par la STO, regorgent d'informations concernant les déplacements et les modes de transport pour la grande région de Hull/Gatineau. Cependant, les deux bases de données sont totalement différentes de par leurs structures, leurs tables et la diversité des informations qu'elles contiennent. Il est donc important de bien identifier les données pertinentes et nécessaires à la suite de l'analyse. La méthodologie suivante a donc été adoptée afin de développer les données requises.

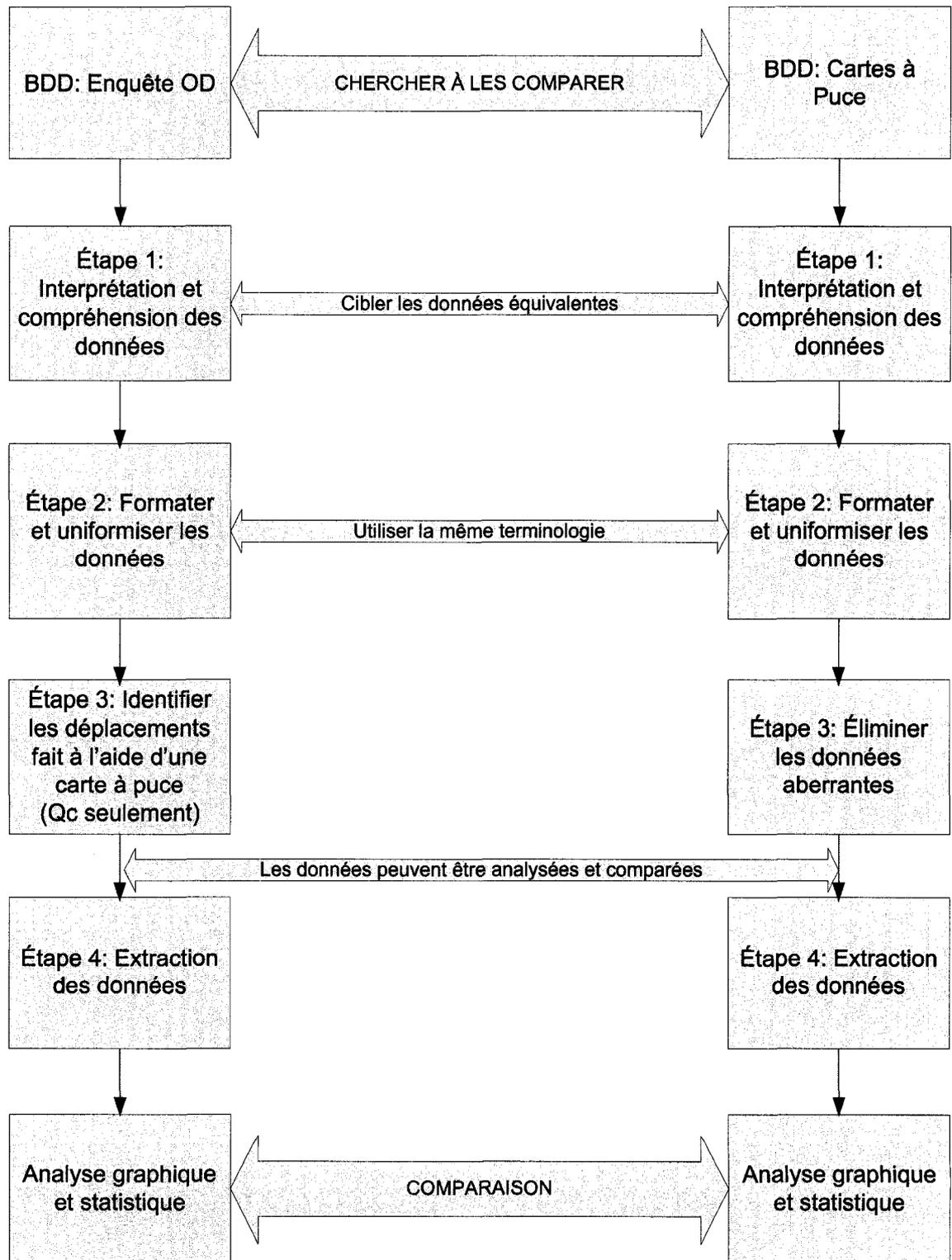


Figure 3.2: Méthodologie employée pour la création des fichiers comparables d'analyse

La première étape de cette méthodologie est de se familiariser avec les tables et le modèle relationnel des deux bases de données (Figure 3.2). Cette étape est importante puisqu'elle permet d'établir une équivalence entre les bases de données. La section 4.4 résume les principales tables étudiées.

La même nomenclature n'est pas nécessairement utilisée dans les deux bases de données. Il est donc naturel d'utiliser une notation commune et facilement interprétable pour chacun des champs à l'étude (deuxième étape). Cette notation commune, en plus d'améliorer grandement la lisibilité et la compréhension des bases de données, est un premier pas vers la fusion des données. Cette analyse est faite à la section 4.4.1.

L'identification des déplacements faits à l'aide des cartes à puce est indubitablement la phase critique avant l'extraction des données. Dans un premier temps, la base de données provenant des cartes à puce doit être nettoyée du maximum de codes d'erreurs. Bien que certains enregistrements erronés aient leurs intérêts lors d'analyses sur la fiabilité des systèmes de cartes à puce, il est clair que pour des fins de comparaisons, ces données doivent être ignorées, corrigées ou encore réparties équitablement.

Dans un deuxième temps, la base de données provenant de l'enquête origine-destination propose un défi totalement différent. Par la structure complexe des enquêtes OD, la quantité d'informations présente dans les bases de données est substantiellement plus importante et touche autant les déplacements que les différents modes utilisés ainsi que des données sociodémographiques. Autre particularité, le contexte géopolitique de la région de Gatineau fait en sorte que deux sociétés de transport en commun sont présentes, soit la STO pour la région québécoise et OC Transpo pour l'Ontario. La première étape est donc d'identifier les déplacements faits en transport en commun, autant ceux sur les circuits de la STO que d'OC Transpo.

Ensuite, il est nécessaire de ne retenir que les déplacements faits au Québec, puisque nous disposons uniquement des données cartes à puce pour la STO. Bien que les cartes mensuelles, utilisant uniquement des cartes à puce, soient le mode de paiement privilégié par les utilisateurs des autobus de la STO, plusieurs déplacements se font aussi sans cartes à puce. Il faut donc distinguer les déplacements selon la méthode de paiement. La Figure 3.3 représente cette définition de l'univers des déplacements permettant d'obtenir uniquement des déplacements fait par des détenteurs d'une carte à puce. La section 4.4.2 présente la démarche utilisée ainsi que les principaux résultats.

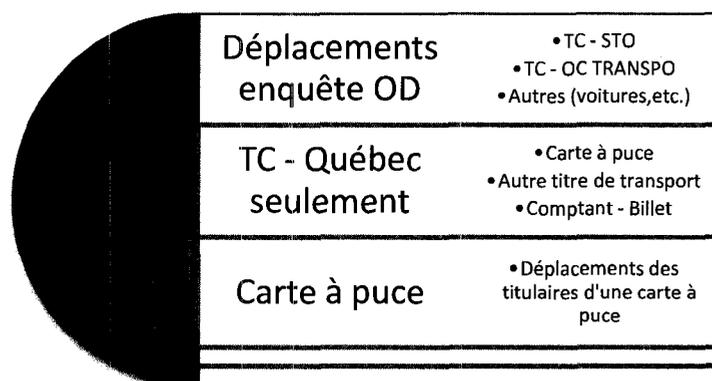


Figure 3.3: Définition de l'univers d'étude

À ce stade, les données sont nommées significativement et elles représentent uniquement les déplacements CAP. Cependant, une dernière étape doit être faite avant de procéder à l'analyse des données. Cette étape, nommée l'extraction des données (Figure 3.2), comprend d'autres formatages qui ont pour but de s'assurer que les données soient équivalentes en tous points. Par exemple, dans la base de données de l'enquête origine-destination, le champ « HEURE » est de type numérique, tandis que le champ équivalent dans la base de données CAP est de type chaîne de caractères. Si aucune modification n'est apportée, il y aura nécessairement une erreur lors de la jonction (« UNION ») des requêtes. La section 0 explique tout le processus de traitement des données et justifie les différentes étapes.

Finalement, ces quatre étapes permettent de cibler les données utiles et, surtout, d'assurer une compatibilité qui permettra d'analyser et de comparer les données sans biais dû à la manipulation des bases de données. Le schéma suivant (Figure 3.4) représente le traitement des données à l'aide des logiciels adoptés.

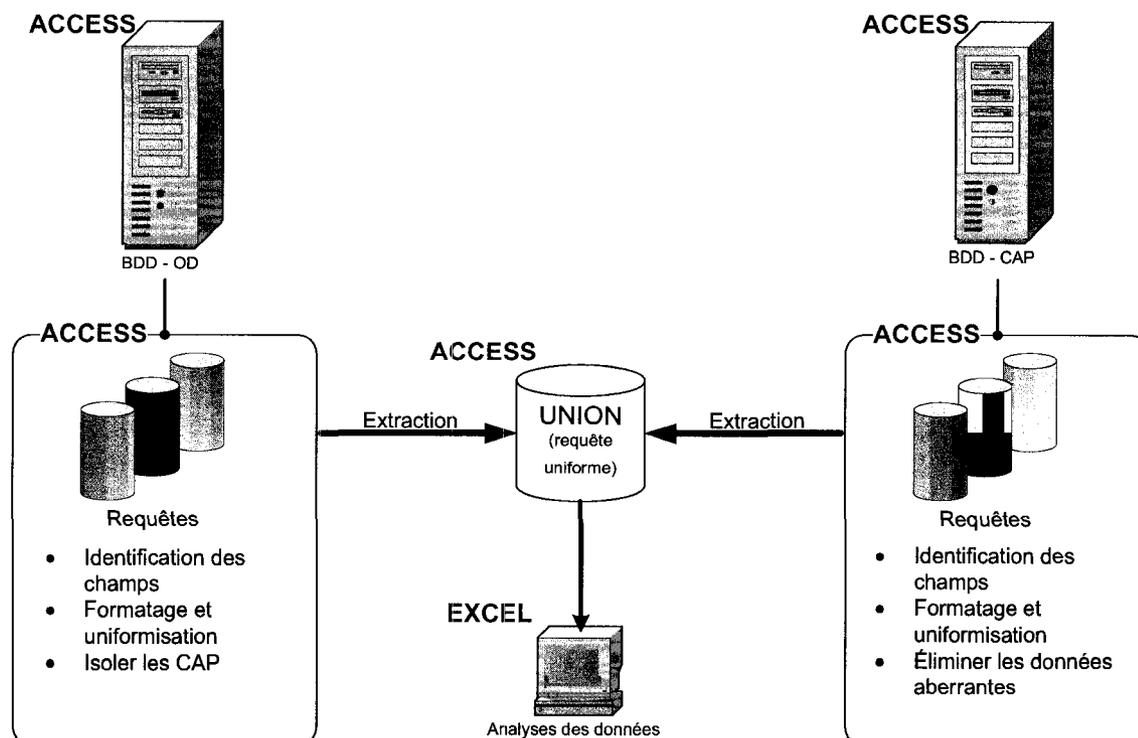


Figure 3.4 : Méthode d'extraction des données

L'utilisation d'un programme de traitement de larges volumes de données est ici essentielle. En fait, le gestionnaire de base de données Access 2007 a été utilisé afin d'effectuer toutes les manipulations précédentes.

Suite à ces manipulations, il faut trouver un moyen permettant d'effectuer des analyses et des comparaisons. Pour des raisons pratiques, tous les traitements analytiques des données sont faits avec Excel 2007. En fait, l'utilisation d'Excel 2007 se justifie par la grande capacité de traitement de l'information (plus de 1 million de lignes) ainsi que par sa polyvalence (tableaux dynamiques croisés, solveur, Visual Basic pour

Applications, etc.). L'emploi d'Excel 2007 permet aussi de regrouper les données CAP et EOD formatées sous un seul et unique fichier. Ceci facilite grandement les prochaines analyses.

Le point de départ des analyses consiste en l'étude de paramètres statistiques globaux. Il faut donc rechercher des indicateurs pertinents et communs aux deux bases de données. Ces indicateurs se doivent d'être très généraux et de permettre d'avoir une vue d'ensemble sur les habitudes de déplacements. Cette analyse globale sert donc de première comparaison des données et permet de faire un portrait des contextes CAP et EOD.

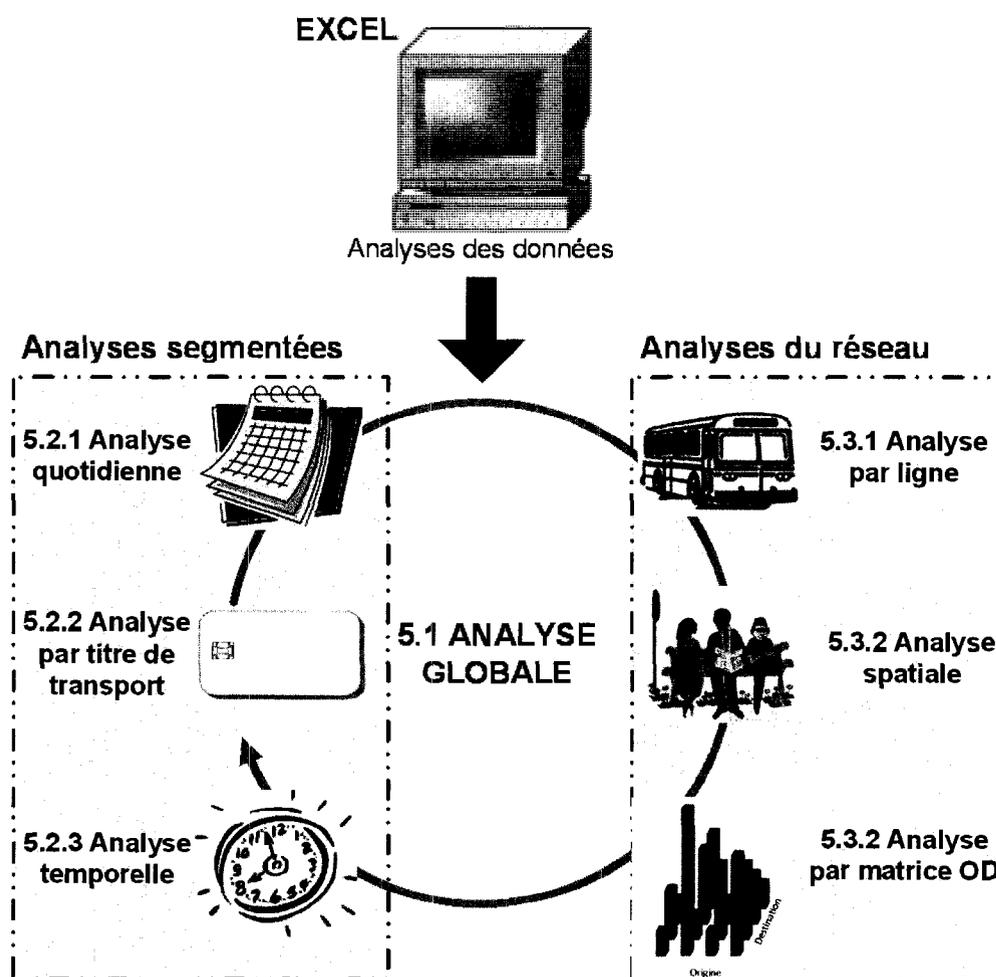


Figure 3.5 : Résumé des analyses

Bien que cette analyse globale soit très instructive, il faut cependant procéder à une étude plus approfondie des données. Cette étude peut être divisée en deux grandes catégories, soit les analyses segmentées et les analyses du réseau (Figure 3.5). La catégorie des analyses segmentées contient :

- les analyses par titre de transport : adultes, étudiants, aînés, etc.;
- les analyses temporelles : par demi-heure, par heure de pointe et par groupe d'heure;
- les analyses quotidiennes : par jour et par type de jour (lundi, mardi, etc.).

La catégorie des analyses du réseau contient :

- les analyses spatiales : par arrêts d'autobus;
- les analyses par lignes : par type de ligne et par ligne;
- les analyses par matrice OD : par origine-destination.

Toutes ces analyses sont faites en trois étapes. La première étape concerne uniquement les données EOD. Lors de cette étape, les données de l'enquête origine-destination seront étudiées au moyen de graphiques et d'indicateurs statistiques. La deuxième étape est identique à la première, sauf que les données analysées proviendront uniquement des bases de données CAP. La dernière étape constituera la partie de comparaison des données et elle fera le lien entre les analyses faites lors des deux premières étapes. Chacune des grandes catégories d'analyses sera ensuite suivie d'explications sur leurs limites et, s'il y a lieu, quelques remarques seront faites.

Une discussion sur les résultats obtenus suit à la section 6. Plus particulièrement, elle porte sur les facteurs d'expansion définis dans l'enquête origine-destination, sur le potentiel de fusion des données et sur les améliorations souhaitables pour obtenir des données plus précises et plus compatibles.

Enfin, le mémoire est complété par une conclusion qui résumera les principaux faits saillants et servira d'ouverture sur les possibilités d'analyses à venir.

CHAPITRE 4

LES SYSTÈMES D'INFORMATIONS

La société de transport de l'Outaouais dispose de deux sources d'informations, soit les données de l'enquête origine-destination et les données collectées sur son réseau à l'aide de systèmes par cartes à puce. Ces données sont complexes et elles nécessitent de plus amples explications avant de procéder à leurs analyses. Les prochaines sections sont donc consacrées à l'étude du contexte de la STO et à la description des bases de données.

4.1 Analyse du contexte de la STO

La Société de Transport de l'Outaouais a été fondée en 1971, sous le nom de la Commission de Transport de la Communauté régionale de l'Outaouais (CTCRO), suite à l'adoption de la Loi sur la Communauté régionale de l'Outaouais de 1969. Auparavant, le service de transport en commun était assuré par trois sociétés privées, soit le Transport urbain de Hull, le Transport Hull métropolitain et le Transport d'écoliers du nord de l'Outaouais. Ces sociétés comptaient 42 autobus afin de desservir une population de 2,5 millions de personnes. L'achat de ces sociétés a été fait en 1972 au coût de 2,25 millions de dollars. La même année, la CTCRO s'est dotée d'un symbole et de couleurs qui, encore aujourd'hui, caractérisent la STO.

En 1973, la CTCRO va de l'avant avec un ambitieux projet d'interchangeabilité des correspondances avec la société de transport ontarienne, OC Transpo. Cette entente permet aux utilisateurs une plus grande flexibilité lors des déplacements dans la capitale fédérale. Toujours avec le souci d'offrir un meilleur service à ces utilisateurs, la CTCRO ouvrira un service de renseignements en 1977. En 1980, la société instaure son premier titre mensuel, la carte « Passe-Partout ». Bien que de prime abord ce titre

semble anodin, ce sera une véritable révolution autant pour les utilisateurs que pour la société de transport.

En 1981, la CTCRO abandonne le transport scolaire afin de concentrer ses activités sur le transport en commun pour le grand public. La société se donne donc comme mandat de connaître les habitudes de déplacements de la population et, par conséquent, elle réalise sa première enquête origine-destination en 1983¹. Cette enquête, peu conventionnelle, se tient à même les autobus et a pour unique but de connaître les caractéristiques de déplacements des utilisateurs du transport en commun. Toujours en 1983, la STO commence l'implantation d'un système informatique du réseau en octroyant un contrat, à la firme Fischback and More, de plus de 1,5 million de dollars. Elle poursuit avec la mise en service en 1985 du système Sagepas qui permet de gérer, de façon automatisée, le parc d'autobus. Des investissements majeurs, de 4,95 millions de dollars, ont lieu au cours de cette même année afin d'agrandir le centre administratif et la capacité du garage.

Le comité TRANS entreprend, en 1986, la tâche ardue de répertorier tous les déplacements faits durant une journée de semaine pour la grande région de la capitale nationale. Cette enquête se fait à l'aide de formulaires qui doivent être retournés par la poste. Le taux d'échantillon est faible² et, par conséquent, la méthode utilisée pour enquêter est remise en question.

Une modification importante au système de paiement est faite en 1987 lorsque la CTCRO interdit le paiement par monnaie de papier et de pièces d'un cent. Avec un certain recul, cette interdiction a probablement été un facteur déterminant pour le succès des cartes mensuelles et, aujourd'hui, c'est grâce à ce succès que la société peut aller de l'avant avec d'ambitieux projets comme les cartes à puce. D'ailleurs, dès 1990,

¹ La première enquête origine-destination pour la région de la capitale fédérale a eu lieu en 1963, cependant, celle de 1983 est la première réalisée par la CTCRO ou, maintenant, la STO.

² 3% pour la population urbaine et 6% pour la population rurale

la CTCRO commence à étudier le potentiel des systèmes de contrôle et de perception des titres de transport.

En 1991, la CTCRO adopte un nouveau nom, soit la Société de Transport de l'Outaouais. La STO continuera son développement jusqu'en 1995 où aura lieu une autre enquête origine-destination. Cette enquête viendra compléter et mettre à jour celle de 1986. Cette fois-ci l'enquête se déroulera par entrevue téléphonique assistée par ordinateur. Les résultats obtenus sont beaucoup plus encourageants puisqu'un peu moins de 25 000 ménages ont pu être rejoints, ce qui représente environ 6% des ménages de la région (Québec et Ontario).

Le projet des cartes à puce fait un pas de plus vers l'avant lorsqu'en 1997 la STO annonce fièrement la mise en œuvre du système informatisé de validation des titres de transport. De plus, cette même année, la STO choisit son fournisseur pour le matériel informatique. L'année suivante, la STO se voit remettre le prix du réseau de transport en commun de l'année au Canada par l'Association canadienne de transport urbain. Fort de cet honneur, la STO poursuit sa progression en allant même en France afin d'étudier des idées novatrices qui pourraient être appliquées sur leur territoire.

La STO a ensuite été innovatrice en implantant des cartes à puce à l'échelle de son réseau. Auparavant, aucune autre société de transport au Canada n'avait été aussi loin dans l'utilisation des cartes à puce. Ceci a d'ailleurs valu à la STO le prix de l'innovation remis par l'association canadienne du transport en urbain (ACTU) en 2003.

En 2005, une autre enquête origine-destination est faite. Cette enquête vis non seulement à mettre à jour les données de l'enquête précédente, mais elle incorpore de nouvelles questions et, par le fait même, cherche à faire un portrait encore plus complet des habitudes de déplacement de la population de la grande région de la capitale nationale. Cette enquête se déroula du 21 septembre 2005 au 29 novembre

2005, soit 48 jours si on exclut les deux journées fériées du 10 octobre et du 11 novembre 2005.

La Société de Transport de l'Outaouais jouit donc d'une place enviable dans le système de transport au Québec, grâce à son passé riche en innovations ainsi que son contexte géopolitique particulier.

4.2 Présentation des données de L'EOD

L'enquête origine-destination est un outil utilisé au Québec depuis de nombreuses années afin d'analyser les habitudes de déplacement de la population. Au fil des ans, les techniques ont évolué, passant des comptages à même les autobus aux formulaires postaux. Actuellement, les enquêtes se font principalement par entretiens téléphoniques assistés par ordinateur. Bien que la ligne directrice des enquêtes demeure l'évaluation de la mobilité des personnes, il est évident qu'elles ont dû s'adapter aux tendances de leur époque et répondre à des besoins en informations de plus en plus importants. Par conséquent, la prochaine section est consacrée aux différentes informations présentes dans l'enquête origine-destination de la grande région de la capitale nationale pour l'année 2005.

4.2.1 Informations sur la population

Lors de l'enquête de 2005, 23 912 ménages ont été questionnés, ce qui représente un échantillon de 5,1% sur tout le territoire (Figure 4.1). Seuls les déplacements de la population de plus de 11 ans ont été enquêtés.

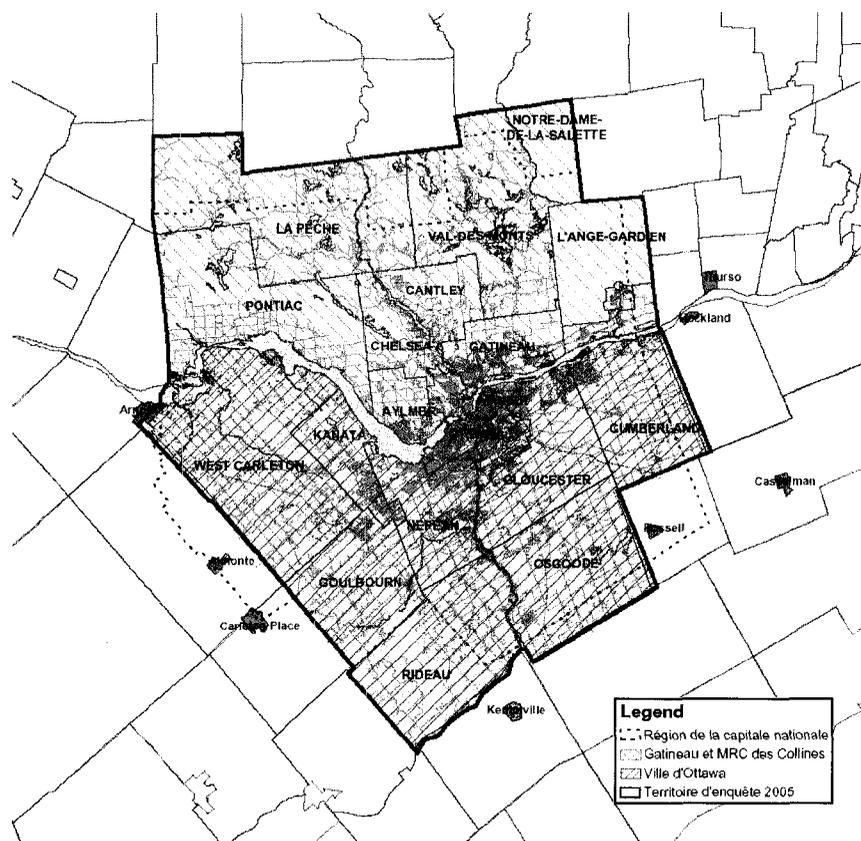


Figure 4.1: Secteurs enquêtés (source : iTRANS Consulting inc. (2006))

L'enquête origine-destination 2005, menée par le comité TRANS³, touche les secteurs québécois, autant qu'ontariens, de la grande région de la capital nationale. Ce territoire s'étend sur 5126 km², dont 2 227km² au Québec. Bien que le territoire québécois représente près de la moitié du territoire enquêté, la population de résidents du Québec correspond à un quart de la population de la région enquêtée (1 150 600 personnes). Les prochaines sections sont dédiées à l'étude plus approfondie des statistiques démographiques, géographiques et sociales de la région québécoise enquêtée.

³ Le comité TRANS est un regroupement d'organismes gouvernementaux et municipaux tel que la société de transport de l'Outaouais et le Ministère du Transport du Québec.

Plusieurs statistiques démographiques, de la région de Gatineau, peuvent être calculées à partir de l'enquête origine-destination de 2005. Ce sont des statistiques qui aident à mieux comprendre les enjeux spécifiques à cette région et, par conséquent, il est nécessaire d'en expliquer quelques-unes.

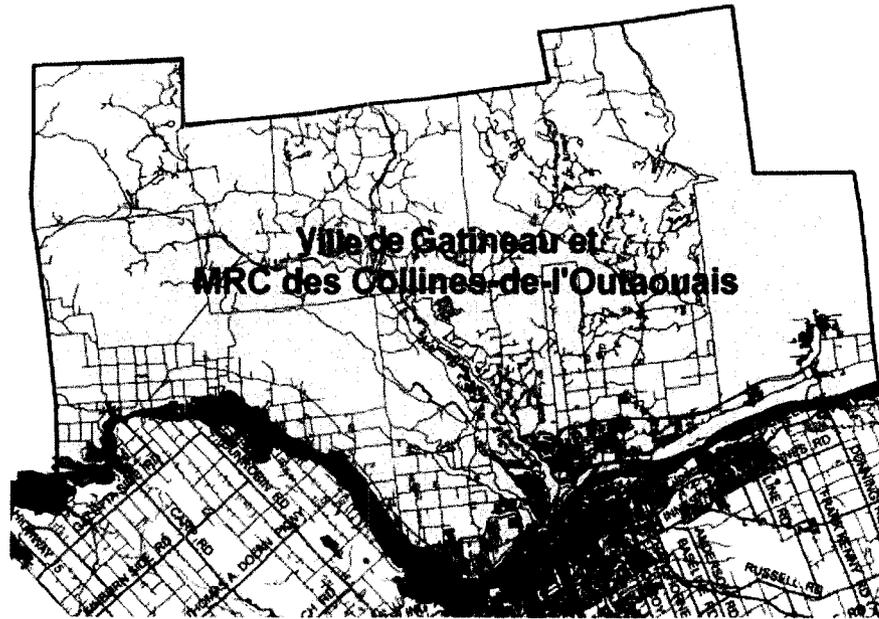


Figure 4.2: Territoire québécois de l'enquête origine-destination 2005 (source: iTRANS Consulting inc. (2006))

Selon le ministère des Transports du Québec, la région de Gatineau comporte en moyenne 2,42 personnes par ménage en 2005, ce qui est près de la moyenne canadienne du recensement de 2006, soit 2,5 personnes par ménage (Statistique Canada, 2008). Le taux moyen de déplacement de la population de Gatineau et des environs québécois est de 2,57 déplacements par jour et par personne. Environ 249 000 personnes sont âgées de 11 ans et plus. La distribution la population selon l'âge est présentée à la Figure 4.3. On constate que les groupes d'âge les mieux représentés sont les 40 ans à 55 ans. En fait, il y a une démarcation nette entre ce groupe d'âge et le nombre de personnes plus vieilles et plus jeunes. Le nombre d'enfants de moins de 15 ans est très petit comparativement aux nombres d'adultes de

plus de 30 ans. Bref, la population de Gatineau est principalement représentée par des adultes s'approchant ou étant dans la cinquantaine.

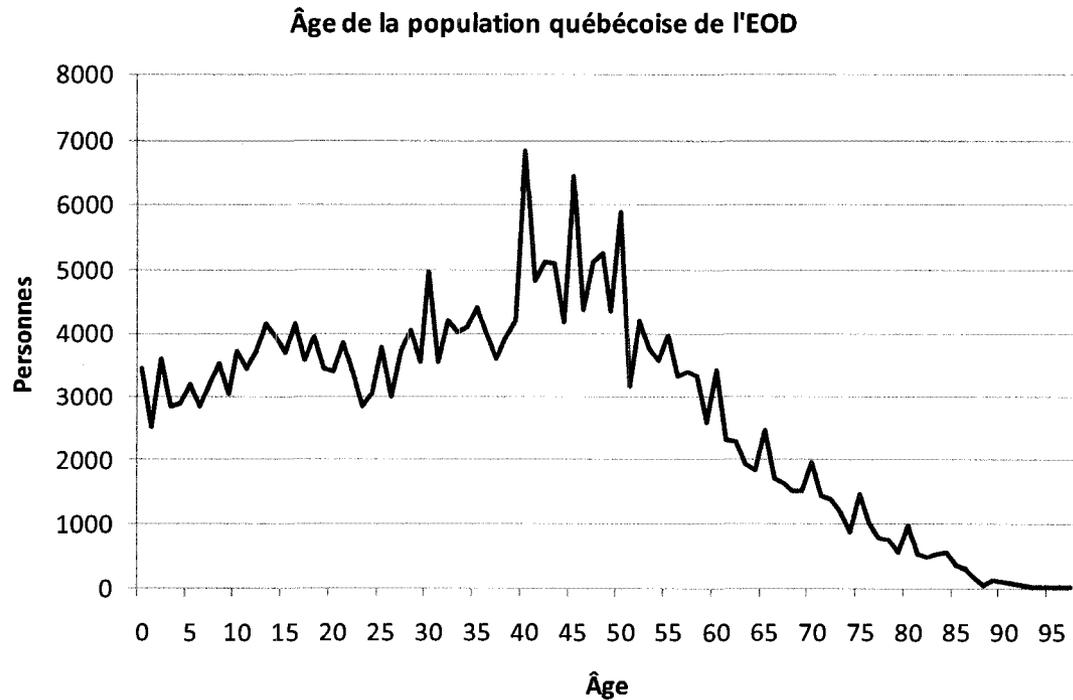


Figure 4.3: Distribution de la population selon l'âge

De la population de plus de 11 ans, 131 900 personnes sont des travailleurs à temps plein et 10 100 personnes sont des travailleurs à temps partiel. La Figure 4.4 montre les autres occupations ainsi que la proportion de population de plus de 11 ans associée. On remarque que le nombre d'étudiants et de retraités est sensiblement le même, soit une représentation de 17 %.

Occupation de la population de Gatineau

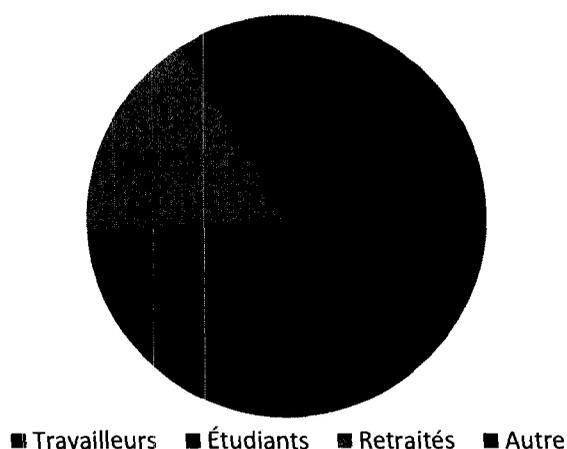


Figure 4.4: Occupation de la population de Gatineau (Transport Québec, 2007)

Il est aussi possible de répéter la même analyse, mais cette fois en ne regardant que les personnes possédant une carte à puce. La Figure 4.5 montre donc les occupations des titulaires d'une carte à puce selon les données de l'EOD⁴. On remarque que la proportion d'étudiants augmente considérablement lorsque l'on compare l'ensemble de la population qui se déplace à la population qui utilise une carte à puce. En fait, cette augmentation du nombre d'étudiants qui utilise le transport en commun (35 %, augmentation de 18 %) semble principalement se faire aux dépens de la proportion de retraités qui se déplace grâce au TC (3 %, chute de 14 %). Néanmoins, la majorité des personnes qui se déplacent grâce aux CAP demeurent des travailleurs, à temps plein ou partiel (59 %).

⁴ La procédure d'extraction des données CAP de l'EOD est décrite à la section 0 4.4.2 Recherche des déplacements CAP dans la base de données EOD

Occupation des titulaires d'une CAP

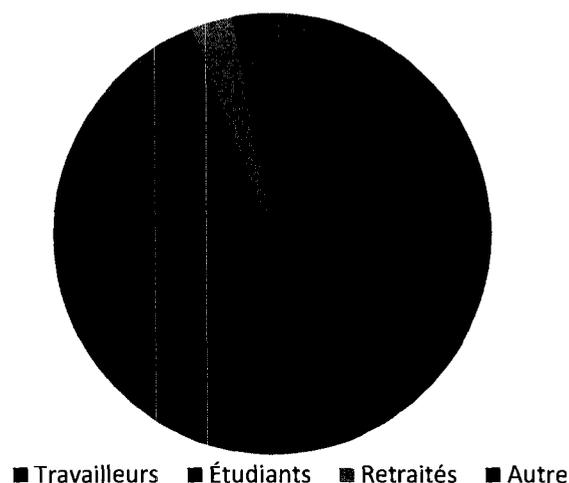


Figure 4.5 Occupation des titulaires d'une carte à puce

L'analyse des modes de déplacements privilégiés par la population de Gatineau est aussi très importante afin de mieux comprendre les enjeux et les bénéfices possibles des analyses de ce mémoire. Donc, c'est sans grande surprise que l'automobile est le mode le plus souvent utilisé pour se déplacer dans la région de Gatineau. En fait, 62,9 % des déplacements sont faits à titre de conducteur et 13,6 % des déplacements sont faits comme passager dans une automobile. La part des déplacements faits sur le réseau de la société de transport de l'Outaouais est de 8,5 %. Cependant, la part totale des déplacements en transport en commun est la somme des déplacements sur le réseau de la STO (8,5 %) et de l'OC Transpo (1,3 %), soit une représentation de 9,8 % de l'ensemble de déplacements faits par la population de Gatineau et des environs québécois. Les autres modes de déplacements sont résumés dans le tableau 4.1.

Tableau 4.1: Modes de déplacement de la population de Gatineau (Source MTQ)

Mode	Gatineau (%)
Auto-conducteur	62,9
Auto-passager	13,6
STO	8,5
Marche	7,5
Autobus scolaire	4,0
Vélo	1,1
OC Transpo	1,3
Autres	1,0

L'étude des principaux axes de déplacements de la population permet de connaître les origines et les destinations privilégiées. Sur la Figure 4.6 on peut voir les lignes de désir où l'épaisseur du trait est proportionnelle à l'achalandage, en termes de personnes, sur cette ligne. En fait, les principales lignes de désires s'orientent vers le centre-ville d'Ottawa puisque c'est le centre d'emploi le plus important de la région. On remarque aussi que certaines banlieues d'Ottawa génèrent davantage de déplacements, comme la région d'Orléans ou celle de Hunt Club. On peut aussi remarque que les grandes lignes de désire se situent principalement du côté de la région d'Ottawa plutôt que de celle de Gatineau.

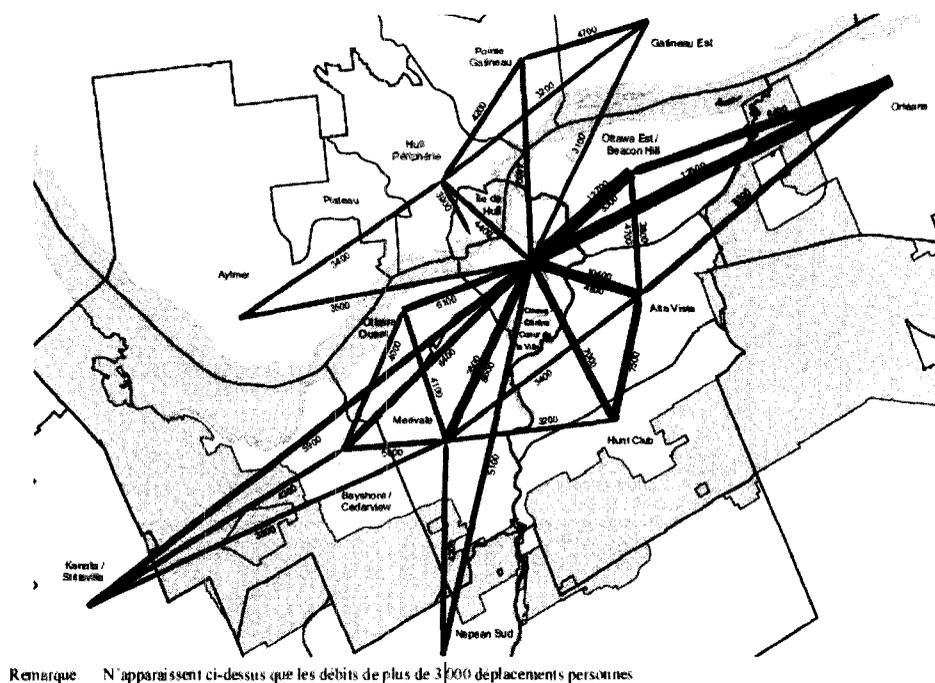


Figure 4.6: Principales lignes de désir pour tout le territoire enquêté (Source: iTRANS Consulting inc. (2006))

Ceci représente un aperçu des habitudes de déplacement de la population de Gatineau. Ces statistiques et ces indicateurs seront repris, à la section 5.1 **Reference source not found.**, pour des fins d'analyses et de comparaisons avec les données CAP, qui sont présentées à la prochaine section.

4.3 Présentation des données CAP

Les déplacements effectués grâce aux cartes à puce sont enregistrés en continu dans le système d'informations de la STO et de manière totalement désagrégée. En fait, un enregistrement dans la table des transactions équivaut à un déplacement fait par un usager du transport en commun. En même temps que le déplacement est enregistré, d'autres informations, tel que le numéro de la carte, l'heure et la date sont aussi sauvegardées dans la base de données.

Pour la période s'étendant du 21 septembre 2005 au 29 novembre 2005⁵, 1 842 666 enregistrements de déplacement ont été faits. Ces déplacements ont été effectués par 28 539 titulaires de carte à puce, ce qui représente environ 80 % des usagers.

Tous les enregistrements ont un champ qui indique si l'utilisation de la carte est valide ou non. Le Tableau 4.2 montre les messages d'erreur ainsi que le nombre d'enregistrements ayant le code d'erreur. Les enregistrements n'ayant aucune erreur ou un supplément exigé peuvent être considérés comme des déplacements effectués. En fait, dans le cas des suppléments exigés, il est possible que le titulaire du titre de transport décide de ne pas payer le supplément et ainsi ne pas se déplacer sur le réseau de la STO. Par contre, il serait surprenant que cela arrive fréquemment et, par conséquent, ces déplacements sont considérés comme valides.

Tableau 4.2: Messages d'erreur pour les enregistrements de la base de données CAP

Message d'erreur	Nombre d'enregistrements (déplacements)
Aucune erreur	1 805 417
Carte refusée (1)	77
Carte refusée (2)	69
Carte invalide	445
Supplément exigé	33 133
Carte expirée	1 410
Titre invalide	2 096
Ligne invalide	3
Période invalide	16

Les déplacements ayant d'autres messages d'erreur sont tout simplement ignorés lors des analyses subséquentes. Ceci se justifie par le fait que l'ensemble de ces

⁵ Ceci exclut les fins de semaine et les deux journées fériées du 10 octobre 2005 et du 11 novembre 2005

déplacements représente environ 0,2 % de tous les déplacements. Ceci équivaut à un peu plus de 80 déplacements erronés par journée et, par conséquent, n'influence pratiquement pas les résultats obtenus.

Les principales variables disponibles dans la base de données, sous format Access, sont :

- Date et heure de l'embarquement;
- Type de tarif (régulier, express, interzone, étudiant, aînés, etc.);
- Numéro et direction de la ligne empruntée;
- Numéro de l'arrêt où a eu lieu la transaction;
- Numéro de l'arrêt estimé où a eu lieu la descente (obtenu par application de l'algorithme de Trépanier, Chapleau, & Tranchant (2007));
- Données opérationnelles telles que le numéro du conducteur, de l'autobus, du voyage, etc.

4.4 Préparation des données

La comparaison des données OD et CAP commence par l'identification et la compréhension des données fournies par la STO (Figure 3.2). Cette section est donc consacrée à l'étude des champs présents dans les bases de données

4.4.1 Les objets étudiés

Les champs utilisés sont le cœur même de l'information présente dans les bases de données. Ces champs contiennent des enregistrements, qui eux, représentent des déplacements, des heures, des dates, etc. Ces champs sont parfois présents uniquement pour l'information qu'ils contiennent, mais ils peuvent aussi servir à assurer l'intégrité des tables, par exemple, en ne permettant pas de doublons.

L'analyse des bases de données commence par l'identification des champs pertinents qui sont dépendants de l'information recherchée. La Figure 4.7 et la Figure 4.8 montrent les champs utilisés et jugés pertinents.

Table: DEPLACEMENT	Table: PERSONNES	Table: Logis
<ul style="list-style-type: none"> • CLEDEPLACEMENT (Numérique): Identificateur unique d'un déplacement qui n'est comptabilisé qu'une seule fois. • HEURE (Numérique): Heure d'arrivage au lieu. • GRPH (Numérique): Groupe d'heures du déplacement. • MODEL 1,3,4,5 (Numérique): Mode de transport utilisé à l'origine du déplacement. • LIGN 1,3,4,5 (Numérique): Parcours d'autobus utilisé à l'origine du déplacement. • PAETC (Numérique): Mode de paiement pour le transport en commun. • SDORAG07 (Numérique): Code du secteur de recensement d'origine du déplacement version 2007. • SDDEST07 (Numérique): Code du secteur de recensement de destination du déplacement version 2007. • FacExp (Numérique): Facteur d'expansion du déplacement. 	<ul style="list-style-type: none"> • CLEPERSONNE (Numérique): Identificateur unique à l'assuré qui n'est comptabilisé qu'une seule fois. • PAIEMEN (Numérique): Possibilité d'un laissez-passer ou d'un abonnement. • PAIEMENTP (Numérique): Type de laissez-passer ou d'abonnement. 	<ul style="list-style-type: none"> • CLELOGIS (Numérique): Identificateur unique à l'assuré qui n'est comptabilisé qu'une seule fois. • DateEnq (Date): Date des déplacements rapportés. • JourEnq (Numérique): Jour de l'enquête. • sdlogis05 (Numérique): Subdivision de recensement du logis version 2005. • F3_s (Numérique): Correction pour le jour enquête. • FacLog (Numérique): Facteur d'expansion du logis.

Figure 4.7: Tables et champs utilisés dans la base de données EOD

Les tables, nommées en haut des tableaux sur les figures ci-dessus, représentent tout simplement un regroupement de champs. Les types des champs sont indiqués entre parenthèses. La table et les champs commençant par « CB_ » ne proviennent pas de la STO et ils servent uniquement à diminuer le temps des requêtes et faciliter l'interprétation des données. Par exemple, il existe seulement 8 titres de transport différents dans la base de données EOD, tandis que la base de données CAP en comporte 38 différents. La création de la table « CB_Titre » a donc permis d'utiliser la même notation, soit les 8 titres de transport présents dans l'enquête origine-destination, pour les deux bases de données. La correspondance entre les titres de transport présents dans la base de données CAP et ceux de la base de données EOD sont présentés à l'annexe A.1.

Table: sivi_transactionBus	Table: sivi_Jours	Table: CB_Titre
<ul style="list-style-type: none"> • Identificateur (Numérique): Identificateur unique pour le déplacement. • dateComp (Date): Date de l'embarquement. • NUM_U (Numérique): Parcours d'autobus utilisés. • numJour (Numérique): Numéro correspondant au jour du déplacement (voir table sivi_Jours). • typeTran (Numérique): Type de transaction (par exemple, validation ou correspondance). • NUM_TITRE (Numérique): Numéro correspondant au type de titre de transport de l'utilisateur (voir table CB_Titre). • HeureComp (Texte): Heure de l'embarquement (voir table CB_30min). 	<ul style="list-style-type: none"> • Num_Jour (Numérique): Identificateur unique pour un type de journée (corresponds au champ numJour de la table sivi_transactionBus). • Libelle (Texte): Type de journée. 	<ul style="list-style-type: none"> • CB_Num_Titre (Numérique): Identificateur unique pour un type de titre de transport (corresponds au champ NUM_TITRE de la table sivi_transactionBus). • CB_Libell_Titre (Texte): Type de laissez-passer ou d'abonnement.

Figure 4.8: Tables et champs utilisés dans la base de données CAP

4.4.2 Recherche des déplacements CAP dans la base de données EOD

Tous les déplacements de la base de données EOD ne sont pas des déplacements faits par des titulaires d'une carte à puce. Il faut donc isoler les déplacements CAP des autres déplacements. Pour ce faire, le premier critère retenu est le mode de transport. Les champs « mode 1 », « mode 2 », « mode 3 », « mode 4 » et « mode 5 »⁶ indiquent le mode de transport utilisé pour le déplacement ou, plus particulièrement, pour une partie du trajet du déplacement. Lorsque le mode de transport est égal à « 4 », cela signifie que le déplacement s'est fait sur un autobus de la STO. Dès qu'un des cinq champs « mode » est égal à « 4 », alors le déplacement peut être considéré comme un déplacement TC. La Figure 4.9 montre que 59 441 déplacements ont nécessité au moins un déplacement grâce à un autobus de la STO. Cependant, ces déplacements comprennent tous les utilisateurs du transport en commun. Il est donc nécessaire de raffiner davantage les critères afin d'isoler les titulaires d'une carte à puce.

⁶ Pour la définition complète des modes, il est possible de se référer à l'annexe A.3

Le champ « PasMen » de la table « PERSONNES » permet d'identifier les détenteurs de titres mensuels. Étant donné qu'à la STO tous les titres mensuels sont, depuis 2001, des cartes à puce, il suffit de retenir que les déplacements ayant le champ « PasMen » égal à « 1 ». Un champ « PasMen » égal à « 1 » signifie que la personne interrogée lors de l'enquête origine-destination a déclaré qu'elle possédait un titre de transport mensuel. Autrement dit, si le déplacement s'est fait sur un autobus de la STO (« mode » égal à « 4 ») et que la personne est titulaire d'un titre mensuel (« PasMen » égal à « 1 »), alors ce déplacement peut être considéré comme un déplacement CAP. 50 299 déplacements correspondent à cette définition.

Cependant, seules les personnes habitant au Québec peuvent être en possession d'une carte à puce de la STO. Par conséquent, le champ « sdrlogis05 » permet de connaître la subdivision de recensement (version 2005) où habite la personne interrogée. De plus, les subdivisions des régions québécoises débutent toutes par « 248 ». Donc, en appliquant les deux premiers facteurs et en limitant le champ « sdrlogis05 » aux enregistrements qui commencent par « 248 », on obtient alors, en moyenne par journée selon l'enquête origine-destination, 47 373 déplacements faits par des titulaires d'une carte à puce STO.

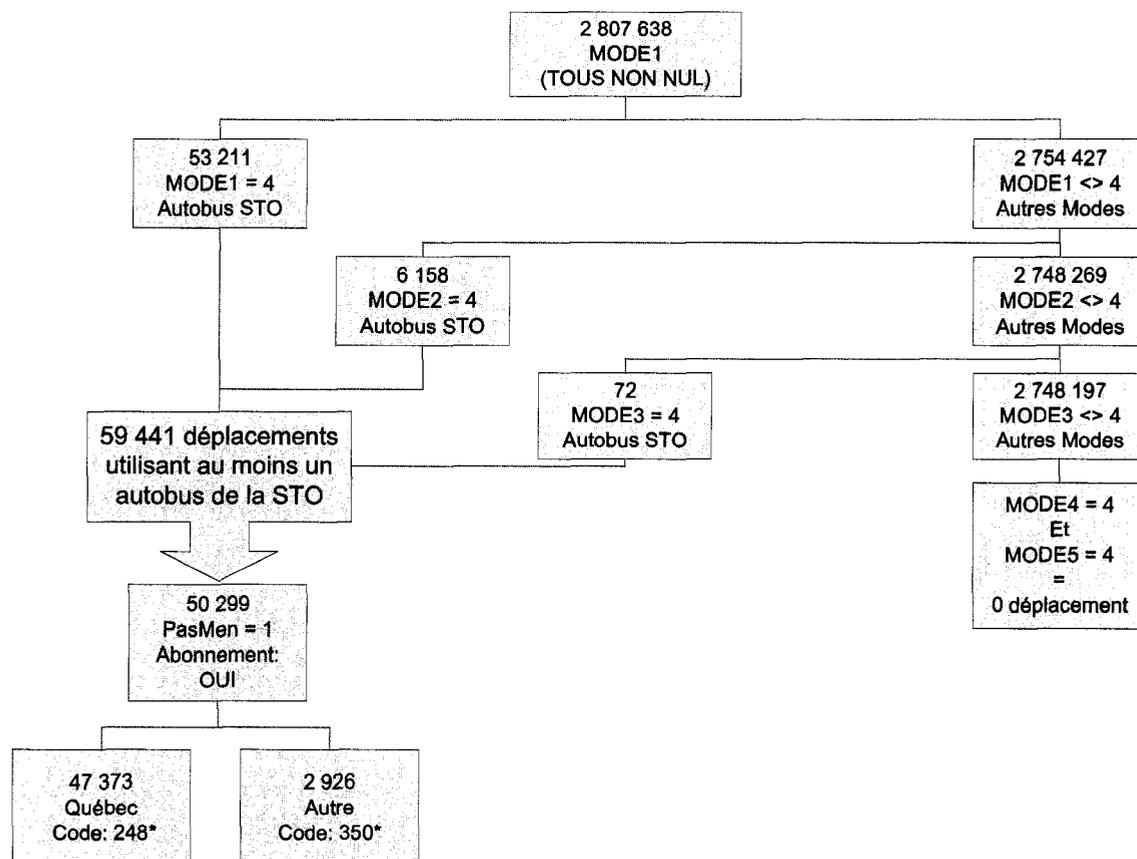


Figure 4.9: Déplacements CAP de la base de données EOD

Sur les 2926 déplacements faits grâce à au moins un autobus de la STO par des personnes résidant en Ontario et possédant un titre mensuel, 1045 déplacements sont questionnables (Figure 4.10). En fait, 699 de ces déplacements ont été faits uniquement sur des autobus de la STO (tous les modes égaux à 4) et 346 déplacements ont utilisés plus souvent les autobus de la STO que tout autres modes. De prime abord, il semble aberrant qu'une personne utilise principalement les autobus de la STO et que son titre de transport ne provienne pas de la STO. De plus, ces déplacements ne représentent qu'environ 2 % des déplacements CAP déjà identifiés. C'est donc pourquoi ces données ne seront pas considérées comme des déplacements CAP sur le réseau de la STO.

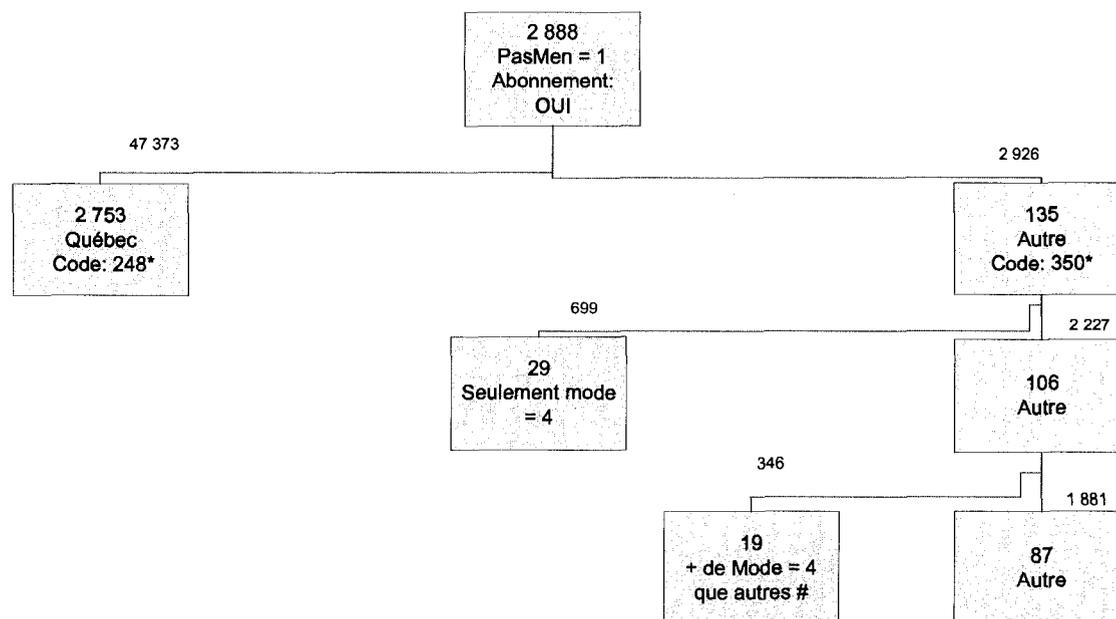


Figure 4.10: Anomalies des déplacements CAP pour les résidents hors Québec

4.4.3 Les hypothèses posées et leurs justifications

Certaines irrégularités apparaissent dans les bases de données de l'enquête origine-destination et des cartes à puce. Ces irrégularités doivent être analysées et traitées afin de s'assurer de l'intégrité des données utilisées pour des fins de comparaisons. Les irrégularités remarquées sont ici présentées et les hypothèses ainsi que les décisions qui leur sont associées sont détaillées.

La table « sivt_Transactions » comporte deux champs, « DateComp » et « numJour » qui sont parfois erronés. En fait, il arrive que pour une date du champ « DateComp », il y ait plus d'un « numJour » associé. Autrement dit, le 25 septembre 2005, qui est un dimanche (donc « numJour » égal à « 7 ») apparaît aussi comme un lundi (« numJour » égal à « 1 ») pour certains enregistrements. Cependant, en regardant tous les enregistrements qui comportent cette anomalie, on peut remarquer que la valeur du champ « HeureComp » est toujours supérieure à 24 heures. Autrement dit, la valeur de « numJour » est correcte, mais la date ne s'est pas ajustée en conséquence.

Cependant, les données de l'enquête origine-destination sont elles aussi sujettes à ce type d'anomalie puisque les heures enregistrées pour les déplacements peuvent aller jusqu'à 28 heures. Bref, étant donné que les deux bases de données ont le même type d'approche vis-à-vis des heures des déplacements, aucune mesure n'a été prise pour modifier la date du déplacement.

Un autre formatage est aussi nécessaire. Celui-ci touche les dates enquêtées. En fait, l'enquête origine-destination se déroule du 21 septembre 2005 au 29 novembre 2005 et exclut les fins de semaines. De plus, le 10 octobre 2005 et le 11 novembre 2005 sont des journées fériées, donc elles ne sont pas enquêtées. Il est donc essentiel de restreindre les données CAP aux dates enquêtées afin de garder une compatibilité complète des données.

4.4.4 Traitement des données sous Excel 2007

Cette section sera consacrée à l'étude des requêtes sous Access ainsi que des tableaux et du traitement des données sous Excel. Ceci est la dernière phase de préparation des données avant la comparaison et, potentiellement, la fusion.

Les données importées dans Excel sont à la fois nombreuses et complexes à organiser. De plus, lors de la création de tableaux ou de graphiques, les erreurs de manipulations peuvent être fréquentes et difficiles à cerner. Il est donc nécessairement d'automatiser toutes les manipulations qui pourront être faites sur les données.

Par conséquent, un outil a été créé dans Excel afin d'afficher des indicateurs, des tableaux ou des graphiques sans manipuler directement les données importées. Cet outil fonctionne uniquement sous Excel 2007 et il a la particularité d'utiliser à la fois les tableaux dynamiques croisés et du code VBA. Cet outil est composé d'une interface où l'utilisateur choisit des options afin d'afficher les données et graphiques qu'il souhaite étudier.

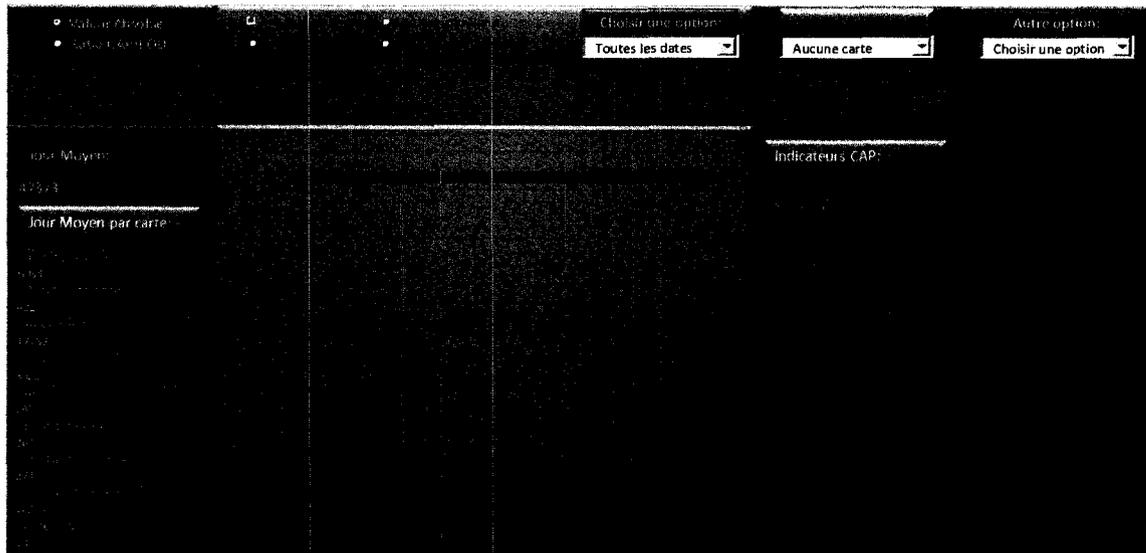


Figure 4.11: Interface de l'utilisateur du fichier de comparaison des données

Parallèlement à cette interface, plusieurs pages contiennent les données sous forme de tableaux dynamiques. Lorsque l'utilisateur choisit une option, une macro Visual Basic Application (VBA) s'exécute afin de modifier les tableaux dynamiques pour afficher les valeurs souhaitées. L'outil permet d'afficher quatre types d'analyse des données et plusieurs analyses plus fines, soit :

- L'analyse par date,
 - Type de carte,
- L'analyse par heure,
 - Date,
 - Type de carte,
- L'analyse par ligne,
 - Date,
 - Type de carte,
- L'analyse par direction,
 - Date,
 - Type de carte.

Avec ces analyses, l'utilisateur peut, par exemple, sortir un graphique des montées par lignes d'autobus pour une journée moyenne et un titre de transport en particulier, et ce, simplement en cliquant sur des boutons. Ceci limite donc les manipulations des données et assure, par le fait même, que les analyses sont toujours justes. Il est aussi possible d'ajouter des données provenant de l'enquête origine-destination. En fait, plusieurs boutons permettent d'ajouter des statistiques EOD sur les graphiques CAP, par exemple, la journée moyenne EOD. Il est beaucoup plus commode de comparer les données EOD et CAP lorsque les informations se retrouvent sur le même graphique. Plus de détails concernant l'utilisation de cet outil sont présentés à l'annexe B.1.

CHAPITRE 5

ANALYSES ET RÉSULTATS

Les données provenant de l'EOD et des systèmes de CAP permettent d'étudier certains objets TC communs de manière totalement indépendante. Les estimations de la valeur de ces objets sont aussi très différentes puisque d'un côté on parle de jour moyen fait à partir de facteurs d'expansion et de l'autre côté il est question de moyenne sur des données quotidiennes. Ceci pose un défi de taille afin d'établir des comparaisons. C'est pourquoi une approche graduelle est utilisée dans cette section afin de commencer par des analyses globales afin de se familiariser avec les données pour ensuite finir avec des analyses plus fines et contextuelles.

5.1 Analyse globale

Il a été mentionné précédemment que les données CAP et EOD (si l'on considère uniquement les titulaires d'une carte à puce) sont des estimations d'une réalité ou d'un univers pratiquement identique. Mais, jusqu'à quel point ces estimations représentent-elles la même réalité ou quelle est la fiabilité de ces estimations? Une première ébauche de réponse vient de l'évaluation et de la comparaison d'indicateurs statistiques communs aux deux ensembles de données⁷.

Un premier indicateur est le nombre moyen de déplacements fait par journée pour tous les titulaires d'une carte à puce, et ce, sur tout le réseau. Ce jour moyen représente, selon les données de l'EOD, 47 373 déplacements par jour qui sont faits sur au moins un autobus de la STO par les détenteurs d'une carte à puce. Comparativement, le nombre de déplacements moyen par jour, estimé par le système de collecte de données CAP, est de 32 689 déplacements. En somme, ceci représente

⁷ Rappel : Seulement les données du 21 septembre 2005 au 29 novembre 2005 ont été utilisées. De plus, les fins de semaine ainsi que les journées du 10 octobre 2005 et du 11 novembre 2005 n'ont pas été étudiées puisqu'elles n'ont pas été enquêtées.

une différence de 14 684 déplacements ou une diminution des déplacements de 31 % par rapport aux estimations fixées par l'EOD. Il est certain que la différence est importante entre ces estimations, mais cet indicateur ne peut à lui seul déterminer la compatibilité des données ou encore le potentiel de fusion. Cependant, il faut retenir, de cette différence importante entre les deux indicateurs, qu'il existe probablement un biais, systématique ou non, lors des estimations de l'EOD ou des CAP. Il faut toutefois souligner que même si les deux indicateurs avaient donné un résultat similaire, il aurait été impossible, sans une exploration plus approfondie des données, de déclarer que les deux univers sont totalement identiques. Bref, cet indicateur démontre donc qu'il faudra maintenant analyser non seulement la répartition des données, mais aussi chercher des causes d'erreurs lors de l'estimation des univers.

Un deuxième indicateur est le nombre moyen de montées par journée pour tous les autobus du réseau. Cet indicateur comprend toutes les montées, donc les premiers déplacements sur le réseau et les correspondances. Pour les données CAP, ce nombre est équivalent aux 32 689 déplacements précédents plus les 5 700 déplacements étiquetés comme correspondance, donc, ceci représente 38 389 montées par journée. Lorsque l'on compile les données de l'EOD, on obtient pour le même indicateur un total de 58 330 montées par journée. L'estimation des montées faite par l'EOD est encore à largement supérieure aux 38 389 montées CAP. Si l'on considère uniquement les correspondances, on remarque un écart tout aussi grand puisque l'EOD estime à environ 11 000 correspondances faites par journée, tandis que les données CAP indiquent deux fois moins de correspondances, soit 5 700. Encore là, peu d'information peut être tirée de ces indicateurs sinon qu'il y a aussi une différence importante pour les estimations des correspondances.

Les deux premiers indicateurs concernent les déplacements faits par les usagers tandis que le prochain indicateur ne considère que les détenteurs. Cet indicateur est le nombre moyen d'usagers ayant une carte à puce qui se déplacent quotidiennement

grâce aux autobus de la STO. Selon l'EOD, 24 257 personnes détiennent une carte à puce et se déplacent à toutes les journées sur le réseau. L'estimation CAP, calculée selon la moyenne sur les 48 jours enquêtés, est de 17 716 personnes par jour avec un écart type de 1069 personnes par jour (coefficient de variation de 6%). La Figure 5.1 montre la variabilité du nombre de personnes, selon les données CAP, qui utilisent le transport en commun pour toutes les journées enquêtées.

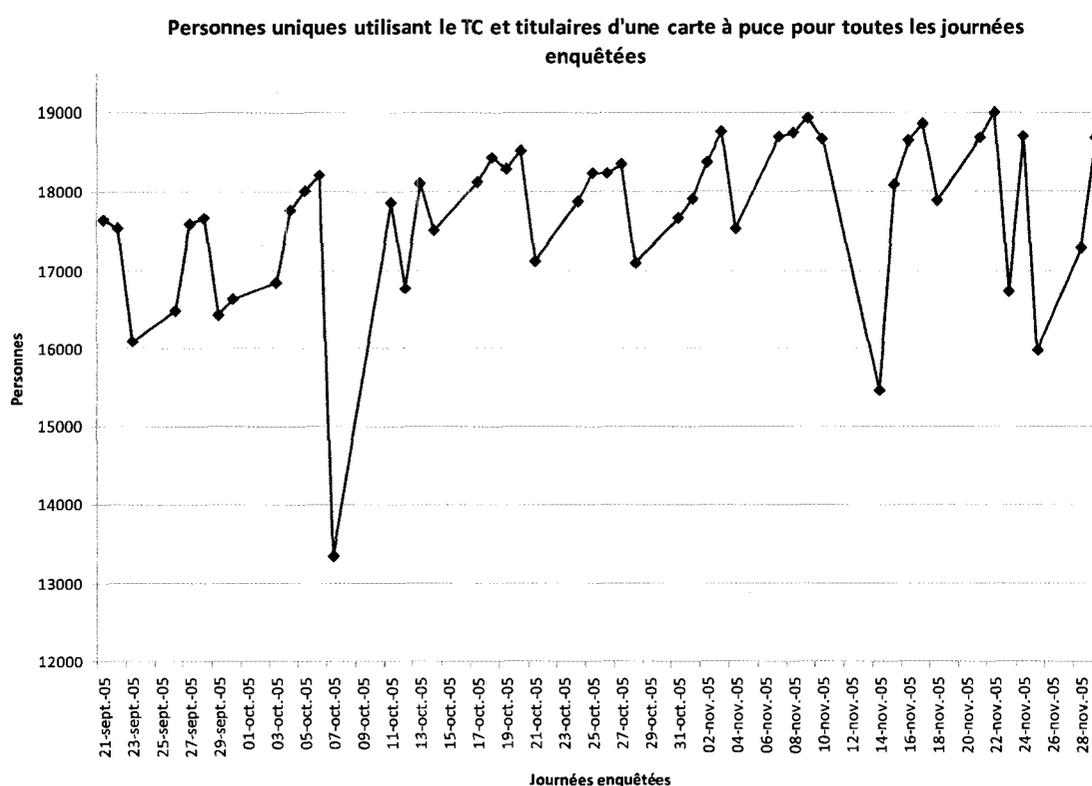


Figure 5.1: Analyse du nombre de personnes utilisant le transport en commun (CAP)

Il serait possible d'étudier plusieurs autres indicateurs de ce type, mais les informations que l'on peut en sortir sont agrégées et elles ne permettent de tracer qu'un portrait très général et peu précis des univers en présence. Il est toutefois possible, grâce à ces indicateurs, d'obtenir une estimation des différences globales entre les univers. De plus, ces indicateurs aident grandement à établir la ligne directrice des recherches plus approfondies qui seront faites par la suite. En fait, dans le cas de cette étude, il est clair

qu'il existe un biais que ce soit dans les données, dans la manière d'enquêter ou de collecter les données, ou encore dans l'attribution des facteurs d'expansion. Bref, ce sont ces indicateurs qui guideront l'analyse désagrégée des données et c'est cette dernière qui permettra de tirer des conclusions valides sur les univers étudiés.

5.2 Analyse segmentée

Les prochaines sections seront consacrées à l'étude des comportements et des déplacements des utilisateurs selon des critères spécifiques, tels que le type de carte ou encore l'heure du déplacement.

Par analyse segmentée, il est sous-entendu que les déplacements des usagers du transport en commun de la STO ne seront plus analysés globalement, mais bien en se concentrant sur des contextes particuliers. Autrement dit, ces analyses seront plus fines puisqu'elles segmenteront les données selon des critères qui peuvent être contenus à la fois dans les deux bases de données. Ces regroupements contextuels pourront donc être comparés et interprétés afin de tirer des conclusions sur les similitudes des données et sur le potentiel de fusion.

5.2.1 Par journée

Les habitudes de déplacement d'une population varient en fonction de plusieurs variables géographiques, sociales ou temporelles. Une de ces variables est le type de journée. En fait, on peut facilement penser que la population se déplace différemment les mercredis que les vendredis, et ce, à cause de facteurs sociaux ou autres. Par conséquent, le Tableau 5.1 montre la variation du nombre de déplacements en fonction de la journée pour les données CAP et EOD. La moyenne CAP est calculée à partir des sommes de déplacements par type de journée (lundi, mardi, mercredi, jeudi et vendredi). Afin de calculer les cinq jours moyens de l'EOD, il faut d'abord déterminer la proportion de la somme des facteurs d'expansion par journée. Cette proportion est

ensuite appliquée au jour moyen de l'EOD, représentant 47 373 déplacements, afin d'obtenir les jours moyens pour chacune des journées de la semaine.

En analysant uniquement les données CAP du Tableau 5.1, on relève trois phénomènes distincts, soit :

- la baisse du nombre de déplacements le vendredi;
- la constance dans les déplacements pour les journées de milieu de semaine (mardi, mercredi et jeudi);
- la variabilité dans les habitudes de déplacements lors des journées du lundi et du vendredi.

La baisse du nombre de déplacements le vendredi est en moyenne de l'ordre de 5 % à 10 % par rapport aux autres journées de la semaine. Similairement, la constance des déplacements pour les journées de milieu de semaine est imputable à des moyennes s'approchant, dont l'écart n'est jamais plus important que 700 déplacements (moins de 2 % d'écart), et à des coefficients de variation bas, surtout pour la journée du mardi.

Tableau 5.1: Analyse statistique selon les journées (CAP et EOD)

Jour (jour)	Moyenne ¹ (déplacement)	Écart type ¹ (déplacement)	Coefficient de variation (%)	Moyenne EOD ¹ (déplacement)
Lundi (9)	32237	2118	6,57	47269
Mardi (10)	33716	918	2,73	45233
Mercredi (10)	33035	1882	5,7	47929
Jeudi (10)	33707	1500	4,45	47632
Vendredi (9)	30485	2727	8,95	48806

¹ Les nombres sont arrondis à l'entier supérieur

La baisse des déplacements le vendredi n'est cependant pas présente lorsqu'on analyse uniquement les données EOD. En fait, c'est tout le contraire qui arrive, car le vendredi est la journée de la semaine où les déplacements sont le plus nombreux. Les moyennes de l'EOD indiquent aussi que pour les journées de milieu de semaine, il n'y a pas nécessairement de constance comme pour les données CAP. En fait, le mardi a une

baisse de près de 6 % des déplacements par rapport au mercredi et au jeudi. Malheureusement, on n'est pas en mesure d'établir des coefficients de variations pour les données de l'EOD, mais il est tout de même possible d'observer que la journée du mardi, qui a le coefficient de variation CAP le plus bas, est le jour où l'écart entre les moyennes de l'EOD et des CAP est le plus petit. Inversement, la journée CAP avec le coefficient de variation le plus élevé, soit le vendredi, est aussi la journée où l'écart entre la moyenne de l'EOD et des données CAP est le plus grand.

Ce phénomène peut en partie s'expliquer par la plus grande difficulté d'estimer une moyenne de déplacements, lors de l'EOD, lorsque les habitudes de la population sont sujettes à de grandes variations.

Cependant, cette analyse démontre non seulement un écart d'environ 15 000 déplacements, mais aussi deux représentations différentes des habitudes de déplacements. On peut aussi se questionner sur le bien-fondé de la répartition des déplacements faite par l'EOD, surtout pour la journée du vendredi. En fait, ceci va à l'encontre de la croyance populaire et de la représentation des habitudes de déplacements faites par les cartes à puce qui montrent les journées du lundi et du vendredi comme ayant un plus faible achalandage.

5.2.2 Par type de carte

L'analyse par type de carte consiste en une segmentation des déplacements, sans les transferts, selon les titres de transport utilisés lors de l'enquête origine-destination⁸. La Figure 5.2 représente les déplacements faits par journée selon les différents types de carte pour les données provenant des systèmes de CAP uniquement. Une première constatation, somme toute assez évidente, est qu'il y a davantage de déplacements faits par des titulaires de titre « adulte régulier » que tout autre titre de transport. Une

⁸ L'annexe A.1 contient toutes les informations pertinentes concernant les correspondances entre les titres de transport de l'EOD et les données CAP.

deuxième constatation est qu'il y a des baisses importantes des déplacements autour des journées du 10 octobre 2005 et du 11 novembre 2005. Ceci est facilement explicable puisque plusieurs personnes prennent congé avant et après les journées fériées, qui ont eu lieu cette année-là le 10 octobre et le 11 novembre.

L'allure des courbes « adulte régulier » et « adulte express » nous indique aussi que le nombre de déplacements tend à augmenter tout au long des 48 jours enquêtés. Ceci est particulièrement difficile à expliquer puisque l'automne est réputé comme une période de l'année où les déplacements sont plutôt constants (Transport Québec, 2007) et aucun événement majeur ayant un impact sur l'achalandage n'a été répertorié.

La courbe « étudiant régulier », contrairement aux adultes, tend à diminuer vers les dernières journées du mois de novembre. Plusieurs hypothèses peuvent être avancées pour expliquer cette baisse. Entre autres, il est fort possible que vers la fin des sessions d'école, les étudiants se rendent moins à l'école puisque certains cours ont déjà pris fin ou tout simplement puisque certains étudiants ont abandonné des cours.

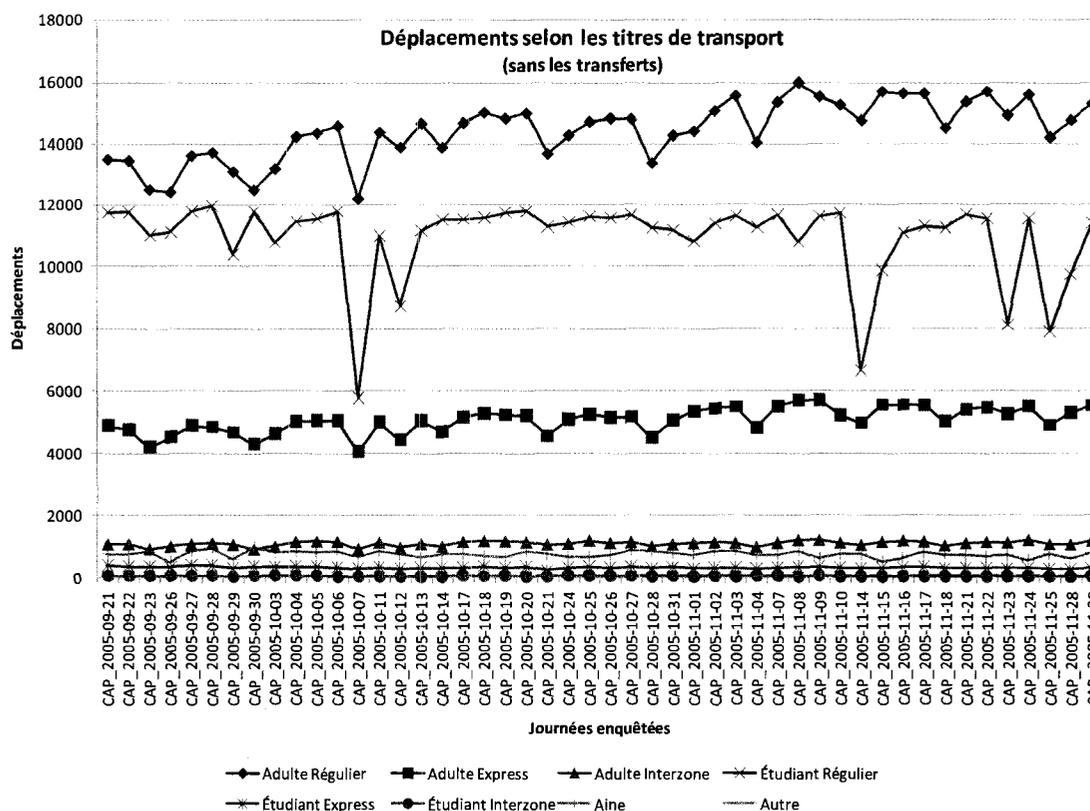


Figure 5.2: Déplacements par titre de transport pour toutes les journées (CAP)

La comparaison de ces données avec l'enquête origine-destination est résumée dans le Tableau 5.2. On peut y voir que l'écart entre les déplacements EOD et CAP est parfois énorme, comme pour les « étudiants réguliers » où il y a une différence de plus de 5 000 déplacements. On constate aussi que les coefficients de variation sont plus élevés que lors des analyses précédentes. En fait, plusieurs titres de transport ont des coefficients de variation tout près de 10 % et parfois tout près de 20 %.

La variation des déplacements change moins, d'une journée à l'autre, pour les titulaires de carte « adultes réguliers » que tout autre titre. Ceci n'est guère surprenant puisque ce groupe comporte beaucoup de travailleurs qui normalement se déplacent tous les jours de semaine à des heures fixes. Contrairement, il n'est pas surprenant de voir un coefficient de variation élevé pour les aînés qui se déplacent selon des habitudes irrégulières souvent en fonction de facteurs extérieurs, tels que la température.

Il est intéressant de noter que tous les titres de transport, à l'exception du titre « adulte interzone », sont surestimés par l'EOD. Dans certains cas cette surestimation est très minime, par exemple pour les titres « adulte régulier » et « adulte express », mais dans les autres cas la surestimation est très importante.

Tableau 5.2: Analyse des déplacements selon les titres de transport

Titre de transport	Moyenne ¹ (Dép.)	Écart type ¹ (Dép.)	Coefficient de variation (%)	Moyenne EOD ¹ (Dép.)	Écart entre EOD et CAP sur EOD (%) ²
Adulte régulier	14 433	959	6,65	17 738	18,64
Adulte express	5 062	400	7,91	6 934	27
Adulte Interzone	1 094	79	7,23	491	-122,82
Étudiant régulier	10 938	1333	12,19	16 374	33,2
Étudiant express	329	29	8,82	863	61,88
Étudiant interzone	44	8	18,19	275	84
Aîné	749	106	14,16	3 922	80,91
Autre	43	6	13,96	779	94,49

¹ Les nombres sont arrondis à l'entier supérieur

² Cet écart est calculé avec le jour moyen EOD et la valeur CAP la plus près comprise dans l'intervalle d'un écart type

La Figure 5.3 représente graphiquement l'écart entre les données CAP et EOD. Les colonnes (vertes) représentent les données EOD tandis que les points représentent les données CAP. L'incertitude des données CAP est déterminée par la valeur moyenne plus ou moins l'écart type (petite ligne noire). On remarque que même en tenant compte de cette incertitude sur les données CAP, l'estimation de l'EOD demeure éloignée. En fait, aucune valeur de l'EOD n'est comprise dans l'intervalle d'un écart type. Il faut cependant noter que si les données CAP ont une incertitude, il en est de même pour les données de l'EOD. Mais cette incertitude est difficile à évaluer et, par

conséquent, l'écart type calculé grâce à ces données sera considéré comme une approximation de cette incertitude (Tableau 5.2).

En observant ces résultats, on constate que les estimations de l'EOD devraient avoir une incertitude très grande afin d'être similaire aux estimations des CAP. Cependant, l'indice de confiance des données de l'EOD est particulièrement bon puisque près de 5 % de la population est échantillonnée.

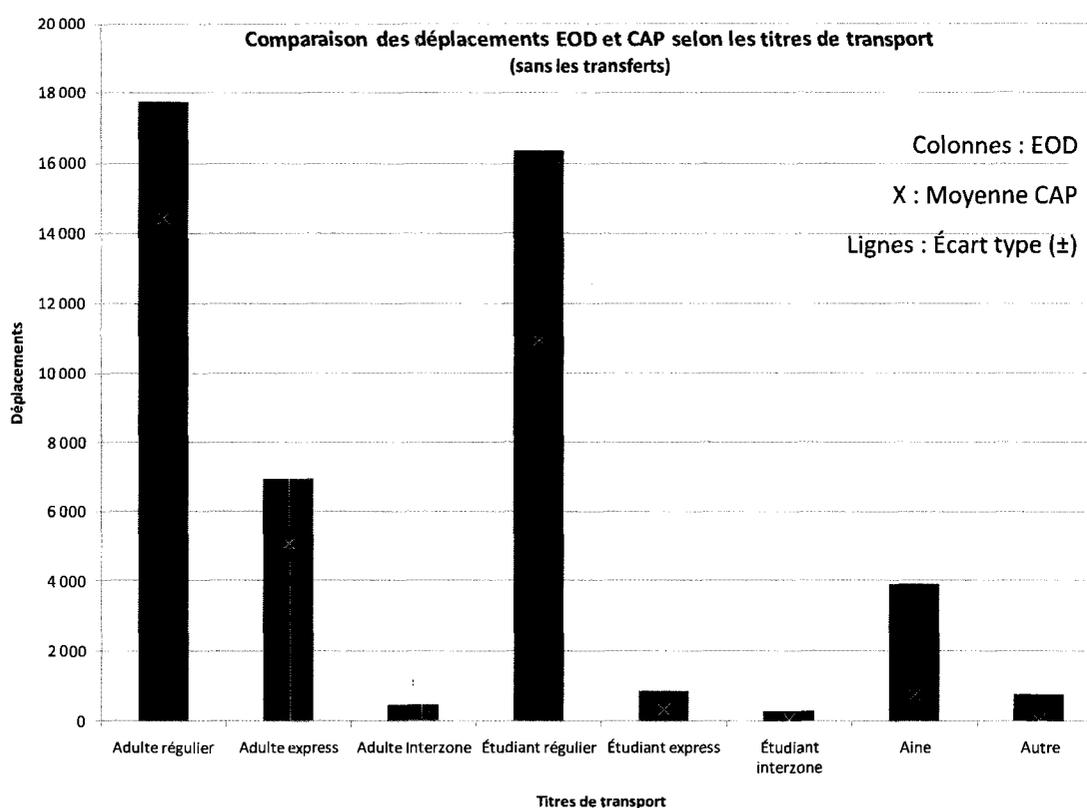


Figure 5.3: Comparaison des déplacements selon les titres de transport (CAP et EOD)

Il est donc envisageable que les facteurs d'expansion des données EOD ne représentent pas exactement la réalité, comme il est aussi envisageable que les données CAP comportent une incertitude puisque la base de données contient des codes d'erreurs, preuve que ce système n'est pas infallible. Cependant, un cas demeure intrigant, soit les « étudiants réguliers ». Lors de l'enquête origine-destination, les étudiants ont été

estimés à un nombre similaire aux « adultes réguliers ». Par contre, on remarque que les estimations de l'EOD et des CAP sont, somme toute, assez comparables pour les adultes, mais que les deux estimations plus distancées pour les étudiants, et ce, malgré la proportion importante des étudiants interrogés lors de l'EOD. Par conséquent, il est nécessaire d'analyser ces deux titres de transport afin de comprendre la divergence des estimations.

La Figure 5.4 indique le nombre de déplacement journaliers, selon l'EOD et les CAP, faits par les titulaires d'une carte « adulte régulier ». La droite (rouge) représente donc le jour moyen EOD pour les titulaires d'une carte « adulte régulier » et les points bleus représentent le nombre de déplacements CAP comptés pour toutes les journées enquêtées.

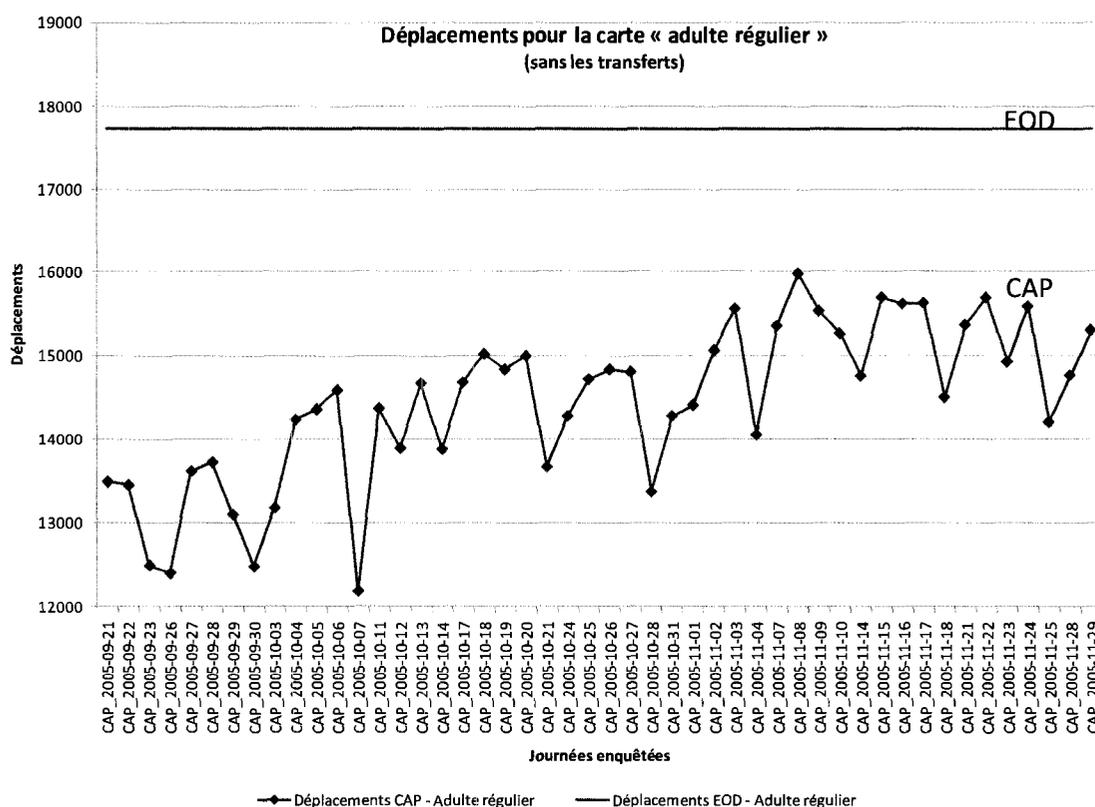


Figure 5.4: Déplacements des « adultes réguliers » pour toutes les journées (CAP et EOD)

Cette figure montre que même si le coefficient de variation est l'un des plus bas parmi tous les titres de transport, il n'en demeure pas moins que, parfois, les données fluctuent considérablement. En fait, l'augmentation progressive, de septembre à novembre, du nombre de déplacement semble accentuer ce phénomène. Il n'en demeure pas moins que ce groupe d'individus a des habitudes stables.

Une analyse plus approfondie s'impose afin de mieux comprendre cette variabilité et cette augmentation progressive. Le Tableau 5.3 montre des statistiques par mois ainsi que par type de journée pour les « adultes réguliers ». Ce tableau confirme que le nombre de déplacements entre septembre et novembre a augmenté (une augmentation d'environ 16% par rapport au mois de septembre 2005). Il vient aussi appuyer l'hypothèse qu'il y a plus de déplacement les mardis, mercredis et jeudis. Le coefficient de variation indique d'ailleurs qu'il y a une plus grande fluctuation dans les déplacements des usagers pour les journées du lundi et du vendredi, ce qui était aussi le cas lors de l'étude précédente.

Tableau 5.3: Analyse des déplacements des « adultes réguliers »

Période (jours)	Moyenne ¹ (dép.)	Écart type ¹ (dép.)	Coefficient de variation (%)	Moyenne EOD ¹ (dép.)
Septembre (8)	13096	553	4,23	17 738
Octobre (20)	14240	711	4,99	17 738
Novembre (20)	15162	552	3,64	17 738
Lundi (9)	14338	977	6,81	15911
Mardi (10)	14902	765	5,13	16083
Mercredi (10)	14625	738	5,05	19515
Jeudi (10)	14760	877	5,94	18287
Vendredi (9)	13430	844	6,29	18892

¹ Les nombres sont arrondis à l'entier supérieur

On constate aussi que les coefficients de variation par mois sont inférieurs au coefficient de variation pour toute la durée de l'enquête (6,64 %). Autrement dit, pour un mois donné, le nombre de déplacements quotidiens ne varie que très peu, mais il y a de grandes fluctuations entre les mois. Cette augmentation du nombre de

déplacements à chaque début de mois implique probablement qu'il y a eu une augmentation des ventes de titre de transport puisque les usagers achètent habituellement leurs cartes le premier de chaque mois.

La Figure 5.5 illustre la distribution des déplacements des « étudiants réguliers » recueillis par le système de collecte de carte à puce (points bleus, sur l'axe secondaire) et compare ces données à la journée moyenne EOD (ligne rouge, sur l'axe primaire) toujours pour le même groupe d'usagers. On remarque que les déplacements sont stables sur pratiquement tout le mois d'octobre, contrairement aux déplacements qui ont lieu en novembre. On peut aussi observer une légère diminution du nombre de déplacements plus on s'approche du mois de novembre. Mais contrairement aux adultes, on ne perçoit pas de démarcation évidente du nombre de déplacements entre les mois.

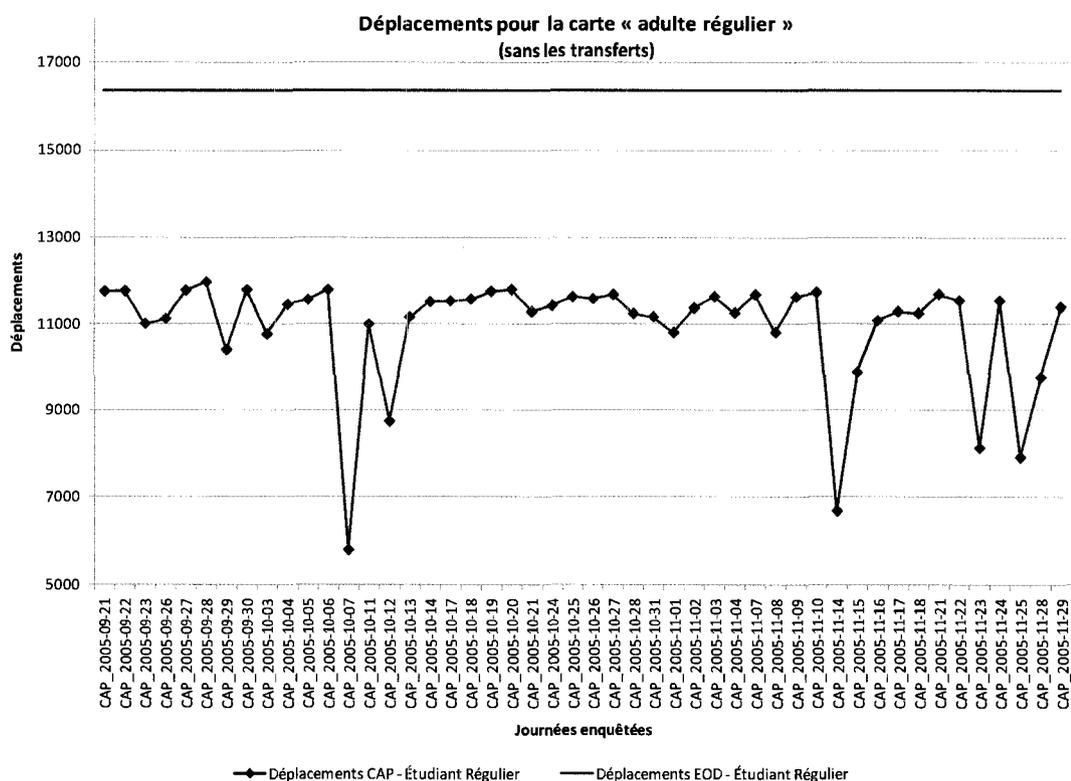


Figure 5.5: Déplacements des « étudiants réguliers » pour toutes les journées (CAP et EOD)

Le Tableau 5.4 analyse les déplacements des « étudiants réguliers » selon les mois et les jours autant pour l'EOD que les données CAP. Premièrement, on remarque que l'écart entre les données EOD et CAP est toujours très grand, peu importe le mois ou la journée. Deuxièmement, les coefficients de variation sont plus élevés pour les étudiants que pour les adultes (Tableau 5.3), et ce, pour pratiquement toutes les périodes de temps analysées. Ceci indique que les données des « étudiants réguliers » varient davantage que celles des adultes. On observe aussi que les coefficients de variation, des étudiants, pour les journées du milieu de la semaine sont moins constants que pour les adultes. Finalement, l'estimation du jour moyen de l'EOD le plus élevé demeure, tout comme lors de l'analyse des adultes, le vendredi. Encore là, ceci va à l'encontre des données CAP qui montrent que le vendredi est l'une des journées où les déplacements sont le moins nombreux. Étrangement, une autre journée avec beaucoup de déplacements selon l'EOD, soit le lundi, est aussi une journée ayant peu de déplacements selon les systèmes de CAP.

Tableau 5.4: Analyse des déplacements des « étudiants réguliers »

Période (jours)	Moyenne ¹ (dép.)	Écart type ¹ (dép.)	Coefficient de variation (%)	Moyenne EOD ¹ (dép.)
Septembre (8)	11449	542	4,73	16374
Octobre (20)	11021	1398	12,69	16374
Novembre (20)	10650	1459	13,7	16374
Lundi (9)	10645	1608	15,11	16694
Mardi (10)	11185	573	5,12	17267
Mercredi (10)	10953	1362	12,44	15115
Jeudi (10)	11476	430	3,75	15431
Vendredi (9)	10337	2061	19,94	17364

¹ Les nombres sont arrondis à l'entier supérieur

Donc, en analysant les déplacements selon les titres de transport, on a pu identifier des titres mensuels où le nombre de déplacements estimé, par l'EOD et les systèmes de CAP, est très similaire, par exemple les « adultes réguliers ». À l'opposé, certaines estimations, dont celles des « étudiants réguliers », sont apparues comme étant

contradictoires. Une analyse plus précise des titres de transport « adulte régulier » et « étudiant régulier » a permis de comprendre que les habitudes de déplacements de ces deux groupes sont différentes. Les adultes semblent plus routiniers dans leurs habitudes de déplacements que les étudiants, qui eux sont sujets à de fortes variations. C'est d'ailleurs pourquoi il est plus difficile d'estimer les déplacements des étudiants et, par conséquent, cela explique en partie la différence des estimations de l'EOD et des CAP pour ce groupe. Cependant, il semble que cet écart n'est pas uniquement imputable aux habitudes de déplacements puisque la différence est importante. Il faut donc poursuivre l'analyse de données afin de comprendre d'où provient cet écart entre les estimations.

5.2.3 Par demi-heure et groupe d'heure

Lors de l'enquête origine-destination, l'heure des déplacements est enregistrée⁹ et un groupe d'heures y est associé. Ces groupes d'heures représentent des périodes de la journée et facilitent l'analyse des données. Voici les groupes d'heures :

- Groupe 1 : 4 h à 6 h 59
- Groupe 2 : 7 h à 8 h 59
- Groupe 3 : 9 h à 11 h 59
- Groupe 4 : midi à 15 h 29
- Groupe 5 : 15 h 30 à 17 h 59
- Groupe 6 : 18 h à 23 h 59
- Groupe 7 : 24 h à 28 h

⁹ Le groupe d'heures associé au déplacement dans la base de données peut représenter l'heure de la montée dans l'autobus (si le premier mode de déplacement est « autobus STO ») ou l'heure du premier déplacement de la chaîne de déplacements (premier mode n'est pas sur un autobus de la STO). Pour les déplacements qui débutent sur un autre mode que « autobus STO », le groupe d'heures utilisé a été extrait de la table « JONCTION » afin que ce soit réellement le groupe d'heure représentant la montée sur un autobus de la STO. Ceci permet donc d'établir une équivalence avec les données CAP (heure de montée).

Les groupes 2 et 5 représentent donc la période de pointe de la matinée et de l'après-midi qui sont particulièrement intéressantes à analyser puisque ce sont des périodes où il y a un très grand nombre de déplacements.

La Figure 5.6 présente les données CAP pour chaque journée en fonction des groupes d'heures. Ce graphique est intéressant dans la mesure où on distingue bien la variation entre les données pour un même groupe d'heures. En fait, on peut aisément supposer que la moyenne des déplacements, pour une période donnée, se situe plus près de la valeur maximale que de la valeur minimale puisque les lignes sont concentrées près du point le plus haut.

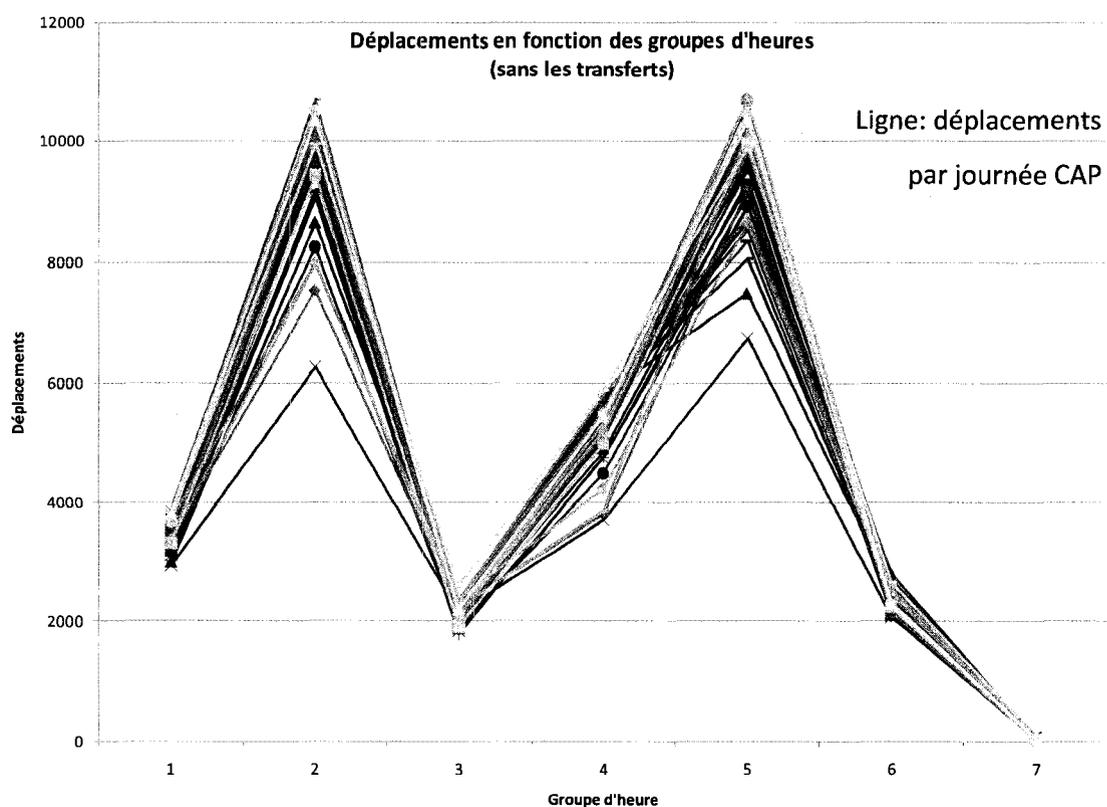


Figure 5.6: Déplacements pour toutes les journées selon les groupes d'heures (CAP)

On constate aussi que les périodes les plus achalandées sont celles de l'heure de pointe du matin (période 2) et de l'heure de pointe du soir (période 5). En fait, ces heures de

pointe sont pratiquement identiques en termes de nombre de déplacements. Il y a aussi une baisse importante du nombre de déplacements pour les périodes 3, 6 et 7 qui sont respectivement les périodes après les heures de pointe et la nuit. Il est cependant intéressant de noter que la somme des déplacements pour les périodes 1 et 3, soit l'avant et l'après heure de pointe du matin, est légèrement plus petite que la somme des déplacements pour les périodes 4 et 6.

Lors des journées moins achalandées, on observe une faible diminution du nombre de déplacements pour les périodes 1, 3, 6 et 7, mais une baisse d'achalandage plus importante pour les périodes 2, 4 et 5. Ceci laisse donc présager que les coefficients de variation pour les périodes 2, 4 et 5 seront plus élevés que pour les autres périodes.

La comparaison des estimations de l'EOD et des systèmes de CAP est faite à la Figure 5.7. Ce graphique compare le jour moyen de l'EOD (ligne mauve) à la moyenne des déplacements répertoriés par le système de CAP (ligne bleue). De plus, les valeurs maximales (ligne rouge) et minimales (ligne verte) des données CAP sont présentes sur cette figure. On peut constater que les périodes 3, 4 et 7 de l'EOD sont comprises entre la valeur maximale et minimale des données de cartes à puce. Par contre, les périodes 1 et 6 de l'EOD sont surévaluées par rapport aux données CAP. Cependant, cette surévaluation, surtout dans le cas de la période 6, est somme toute assez faible. Les périodes 2 et 5, soit les périodes de pointe, sont grandement surévaluées par l'EOD comparativement aux données CAP. En fait, l'écart atteint tout près de 6 000 déplacements lors de l'heure de pointe du soir et environ 4 000 déplacements lors de l'heure de pointe du matin.

Non seulement le nombre de déplacements pour ces périodes est surévalué par l'EOD, mais on remarque aussi que les pentes des droites reliant les sommets 2 et 5 sont beaucoup plus prononcées dans le cas de l'EOD que pour les cartes à puce. Autrement

dit, la fluctuation du nombre de déplacements pour les périodes avant et après les heures de pointe est beaucoup plus importante dans le cas de l'EOD.

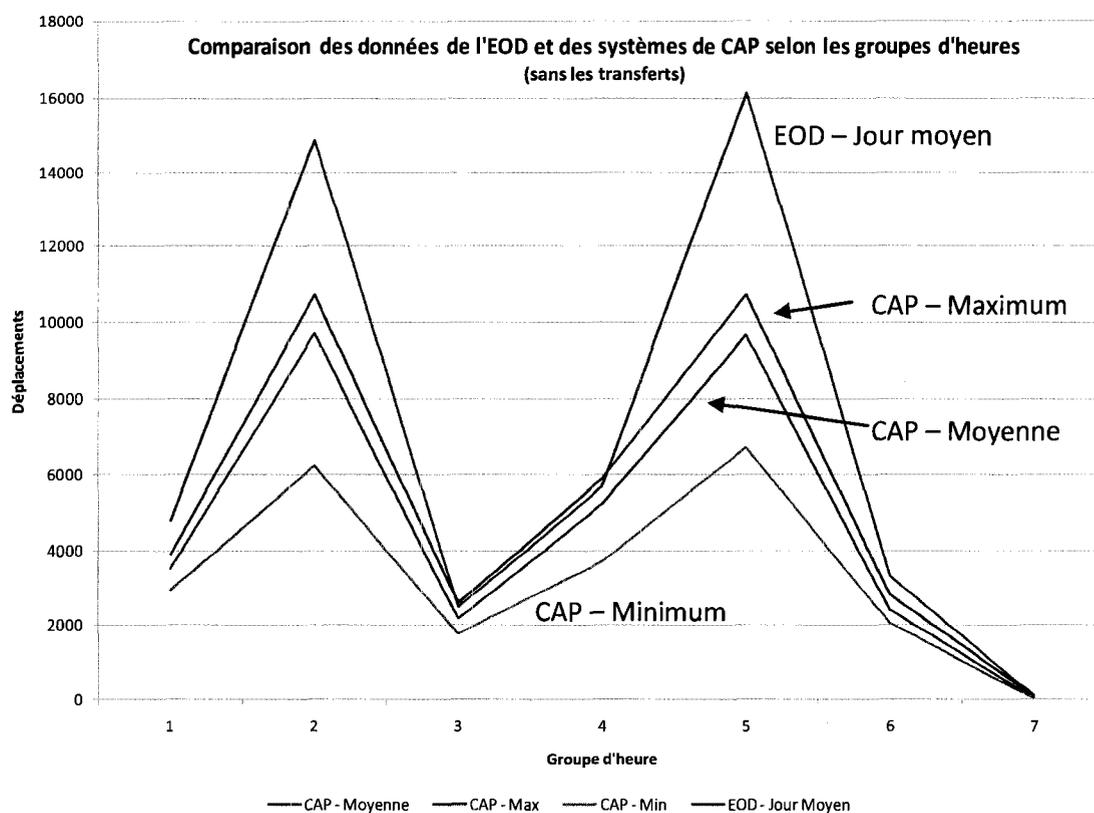


Figure 5.7: Écart entre les déplacements CAP (moyenne, max, min) et EOD (moyenne) en fonction des groupes d'heures

Il est maintenant intéressant d'étudier les déplacements quotidiens afin de comprendre davantage l'écart entre les données. Les Figure 5.8 et 5.9 montrent les données des systèmes de CAP, pour toutes les journées enquêtées, selon les groupes d'heures (lignes avec des points). Les estimations, des jours moyens, de l'EOD sont aussi présentes sur cette figure (lignes pleines). Les courbes illustrées sur la figure 5.8 montrent que les déplacements CAP lors des périodes 2 et 5 ont des fluctuations quotidiennes très similaires. Cependant, l'heure de pointe de l'après-midi (ligne bleue pâle avec des étoiles) a souvent moins de déplacements le vendredi que l'heure de pointe du matin (ligne rouge avec des carrés). Il est donc fort plausible que les usagers

débutent à la même heure pratiquement tous les matins, mais qu'ils écourtent leur journée le vendredi afin d'allonger leur fin de semaine. En fait, les déplacements faits les vendredis à la période « 4 » (figure 5.9) semblent appuyer cette hypothèse puisqu'ils sont pratiquement toujours plus nombreux cette journée-là que toute autre journée de la semaine. Cependant, les données de l'enquête origine-destination décrivent une situation totalement différente puisque la moyenne des déplacements lors de l'heure de pointe du soir (ligne bleu pâle) est plus grande que la moyenne des déplacements pour le groupe d'heure 2 (ligne rouge).

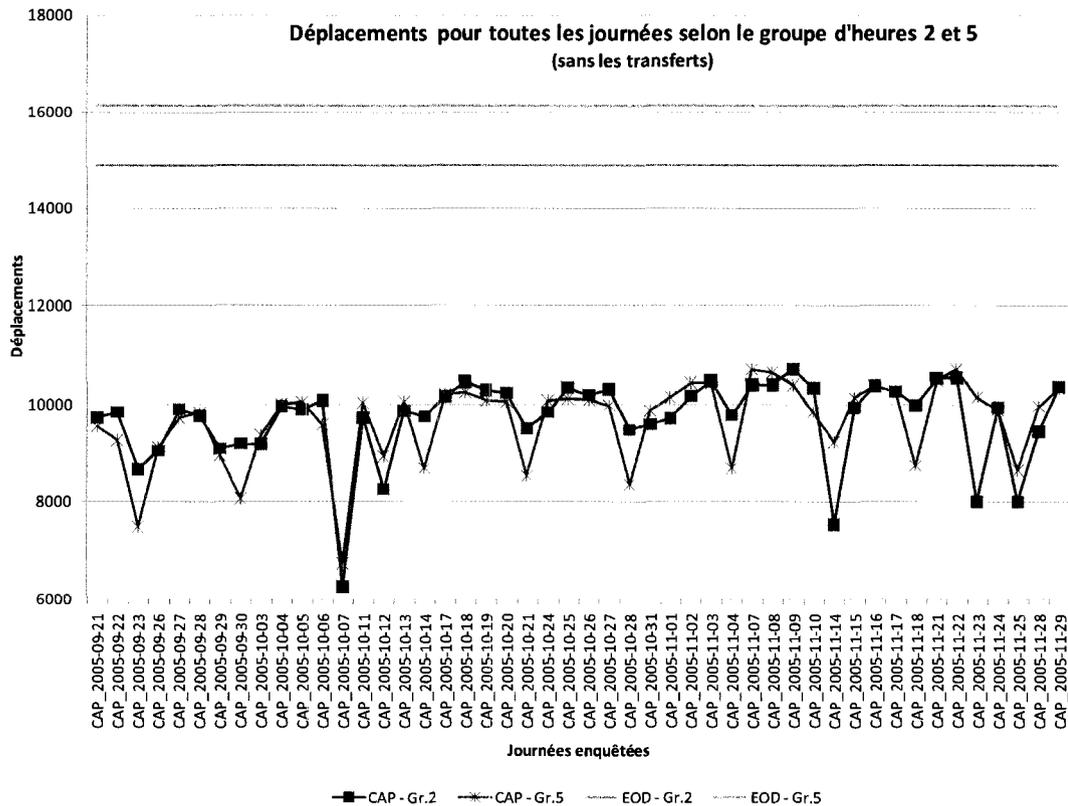


Figure 5.8: Déplacements pour toutes les journées enquêtées selon le groupe d'heures 2 et 5 (CAP et EOD)

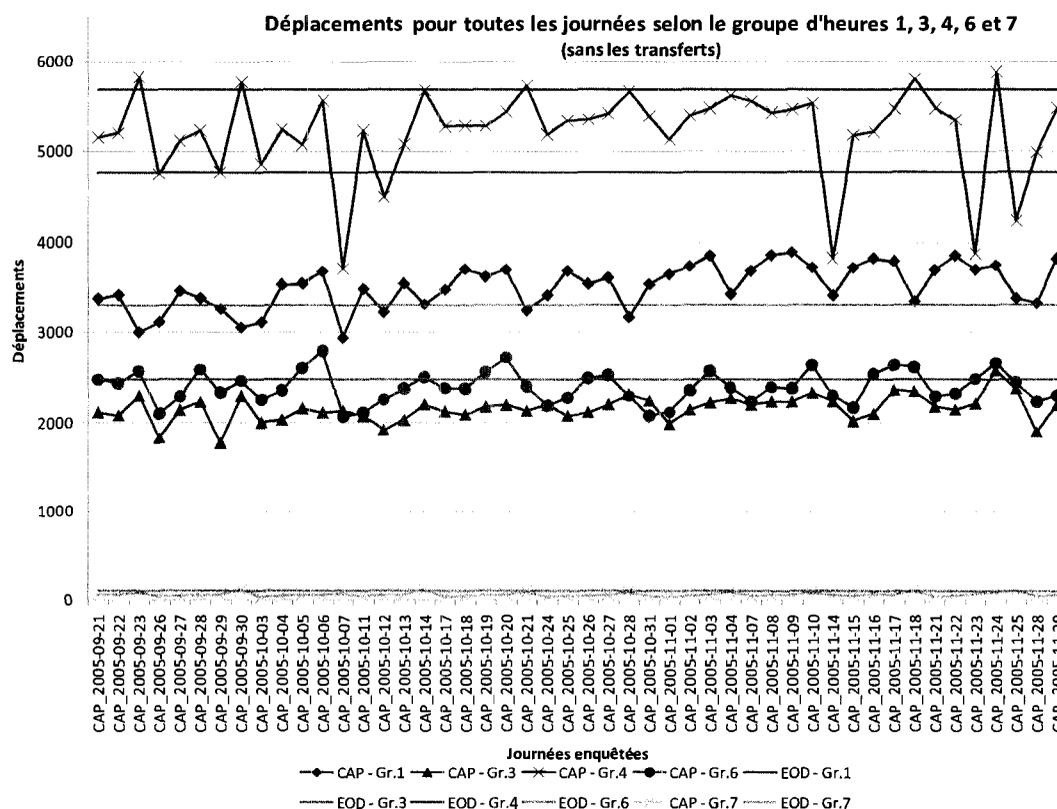


Figure 5.9: Déplacements pour toutes les journées enquêtées selon le groupe d'heures 1, 3, 4, 6 et 7 (CAP et EOD)

Les données des graphiques précédents sont détaillées davantage dans les tableaux suivants. Le Tableau 5.5 résume les moyennes, les médianes, les écarts types et les coefficients de variation des données CAP en fonction des groupes d'heures et les compare avec la moyenne des déplacements pour l'EOD. La médiane des groupes d'heure 1 à 5 est supérieure à la moyenne ce qui indique qu'une majorité de journées ont des déplacements supérieurs à la moyenne. Ceci est principalement imputable aux journées près des congés fériés et aux baisses d'achalandage des vendredis. Si on compare les données CAP à l'EOD, on constate que les estimations des périodes 3 et 4 sont pratiquement identiques. En fait, les données de l'EOD sont très près de la moyenne CAP plus ou moins un écart type. Cependant, pour les périodes de l'heure de pointe du matin et du soir, l'écart entre les estimations des systèmes de CAP et de l'EOD est important. On remarque aussi que le coefficient de variation pour les heures

de pointe est élevé, soit d'environ 9 %. On constate aussi que le nombre de déplacements effectués lors de la période 7 fluctue énormément (coefficient de variation de plus de 33 %).

Tableau 5.5: Analyse des déplacements CAP et EOD en fonction des groupes d'heures

Groupe d'heures	Moyenne ¹ (dép.)	Médiane ¹ (dép.)	Écart type ¹ (dép.)	Coefficient de variation (%)	Moyenne EOD ¹ (dép.)
1	3509	3536	247	7,02	4772
2	9701	9904	872	8,99	14895
3	2162	2170	146	6,72	2484
4	5214	5304	492	9,43	5686
5	9657	9998	860	8,9	16143
6	2395	2383	176	7,36	3297
7	55	50	19	34,21	100

¹ Les nombres sont arrondis à l'entier supérieur

Afin d'expliquer cette variation élevée, les deux tableaux suivants segmentent par mois et par type de journées les déplacements effectués lors de périodes de pointe du matin (2) et de l'après-midi (5). Le Tableau 5.6 montre que le nombre de déplacements pour les deux groupes d'heures a augmenté de septembre à novembre. Cependant, le coefficient de variation est pratiquement toujours le même sauf pour la période 2 du mois de septembre et la période 5 du mois de novembre.

Tableau 5.6: Analyse des déplacements CAP et EOD aux heures de pointe (2,5) par mois

Mois (jour)	Moyenne ¹ (dép)	Écart type ¹ (dép)	Coefficient de variation (%)	Moyenne EOD ¹ (dép)	Groupe d'heures
Sept. (8)	9406	469	4,98	14895	
Oct. (20)	9674	946	9,78	14895	
Nov. (20)	9845	919	9,34	14895	
Sept. (8)	9001	828	9,2	16143	
Oct. (20)	9560	897	9,38	16143	
Nov. (20)	10016	667	6,66	16143	

¹ Les nombres sont arrondis à l'entier supérieur

Il n'en demeure pas moins que les coefficients de variation sont très élevés si on les compare à ceux des autres analyses. En fait, ces coefficients sont très près du

coefficient obtenu au Tableau 5.1 pour la journée du vendredi (8,95 %). Autrement dit, l'impact des baisses d'achalandage le vendredi a beaucoup plus d'impact sur le nombre de déplacements aux heures de pointe que pour tout autre groupe d'heures.

Pour analyser davantage ce phénomène, le Tableau 5.7 reprend les mêmes indicateurs statistiques du tableau précédent, mais utilise les journées plutôt que les mois. On constate donc qu'effectivement le vendredi est la journée, selon les données des CAP, la moins achalandée et où les coefficients de variation sont les plus élevés. Cependant, les estimations de l'EOD semblent définir la journée du jeudi comme étant celles ayant le moins de déplacements.

Tableau 5.7: Analyse des déplacements CAP et EOD aux heures de pointe (2,5) par journée

Jour (jour)	Groupe d'heure : 2				Groupe d'heure : 5			
	Moy. ¹ (dép)	Écart type ¹ (dép)	Coef. de variation (%)	Moy. EOD ¹ (dép)	Moy. ¹ (dép)	Écart type ¹ (dép)	Coef. de variation (%)	Moy. EOD ¹ (dép)
Lundi (9)	9526	905	9,5	14826	9897	560	5,66	16605
Mardi (10)	10136	313	3,09	15142	10218	292	2,86	15449
Mercredi (10)	9743	898	9,22	15703	9998	461	4,62	17021
Jeudi (10)	10047	398	3,97	13248	9838	458	4,65	15042
Vendredi (9)	8959	1187	13,25	15554	8214	685	8,33	16600

¹ Les nombres sont arrondis à l'entier supérieur

On observe aussi que les coefficients de variations sont beaucoup plus faibles pour les journées de milieu de semaine, soit le mardi, le mercredi et le jeudi. Une exception, le mercredi en matinée a un coefficient de variation très élevé. En fait, cette exception est principalement due à deux journées, soit le 23 novembre et le 12 octobre, où les déplacements ont dramatiquement diminué. Aucune explication ne peut être fournie concernant la baisse d'achalandage pour ces deux journées. Cependant, si on néglige

ces deux données, le coefficient de variation est d'environ 3 %, donc très similaire aux autres coefficients.

Bien que ces analyses indiquent une convergence pour certaines périodes et une grande divergence pour d'autres, il est certain qu'une synthèse des déplacements par groupe d'heures est sujette à des effets frontières. En fait, les données se trouvant près de la limite entre deux périodes différentes peuvent pratiquement être considérées comme faisant partie des deux périodes et, par conséquent, cela peut grandement affecter la distribution des déplacements pour un groupe d'heures. Ceci est d'autant plus vrai que les techniques de collecte de données sont très différentes dans ce domaine et elles peuvent embrouiller davantage l'interprétation que l'on en fait. Effectivement, lors de la collecte des données CAP, l'heure exacte, à la minute près, est enregistrée lors des déplacements. Cependant, les données de l'enquête origine-destination ne sont pas nécessairement aussi précises puisqu'elles sont bien souvent arrondies aux cinq minutes près ou encore tout simplement parce que la personne ayant déclaré le déplacement le fait le lendemain et elle a juste un vague souvenir de l'heure de son déplacement. Ceci vient donc renforcer l'idée que ces analyses sont soumises à des effets frontières.

Sans se vouloir une solution pour éliminer cet effet, l'analyse par demi-heure permet une visualisation moins agrégée des données. La Figure 5.10 permet donc d'analyser la variabilité des données CAP selon les demi-heures. En fait, on remarque que pour l'heure de pointe du matin les données fluctuent davantage que lors de l'heure de pointe du soir. On peut aussi facilement constater que l'heure de pointe du matin s'étend sur une période beaucoup moins longue que l'heure de pointe du soir. En fait, l'heure de pointe du matin est de 6 h à 8 h 59 tandis que l'heure de pointe du soir peut débuter aussi tôt que 14 h 30 et elle se finit habituellement près de 18 h 30. Cependant, les déplacements sont plus concentrés à l'heure de pointe du matin, puisqu'ils se situent entre 3 000 et 3 500 déplacements lors de la demi-heure la plus

achalandée (7 h à 7 h 29) comparativement à un peu moins de 3 000 déplacements pour la période du soir la plus occupée (16 h à 16 h 29). Aussi, on observe une grande concentration des déplacements près de la valeur maximale lors des heures de pointe, autant celle du matin que du soir.

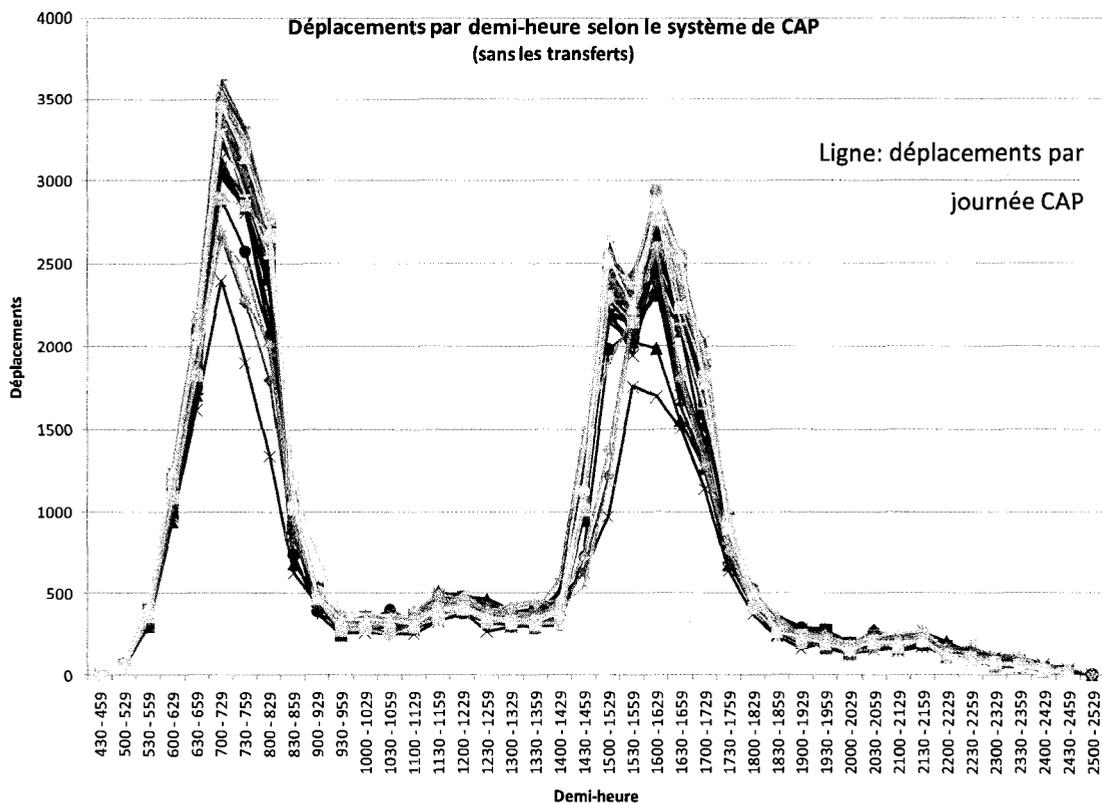


Figure 5.10: Déplacements pour chaque demi-heure de la journée et pour toutes les journées enquêtées (CAP)

La comparaison des jours moyens, divisés en demi-heure, selon les données de l'EOD et des systèmes de CAP est faite à la Figure 5.11. Hors période de pointe, les données provenant des CAP et l'EOD sont sensiblement les mêmes. En fait, pour la période débutant à 8 h 30 et se finissant à 14 h 59, les données CAP représentent pratiquement la moyenne des estimations de l'EOD. Ceci démontre donc que les deux méthodes arrivent à un résultat similaire lors de l'estimation des déplacements entre les deux heures de pointe.

Par contre, les périodes de pointe sont représentées différemment. Le nombre maximal de déplacements lors des heures de pointe est beaucoup plus grand dans le cas de l'EOD. Aussi, les données CAP sont plus étendues et varient plus doucement lors de l'heure de pointe que les données EOD. En fait, les données CAP forment presque un plateau pendant les heures de pointe, contrairement aux estimations de l'EOD qui sont plus sujettes à de grandes augmentations suivies de baisses rapides.

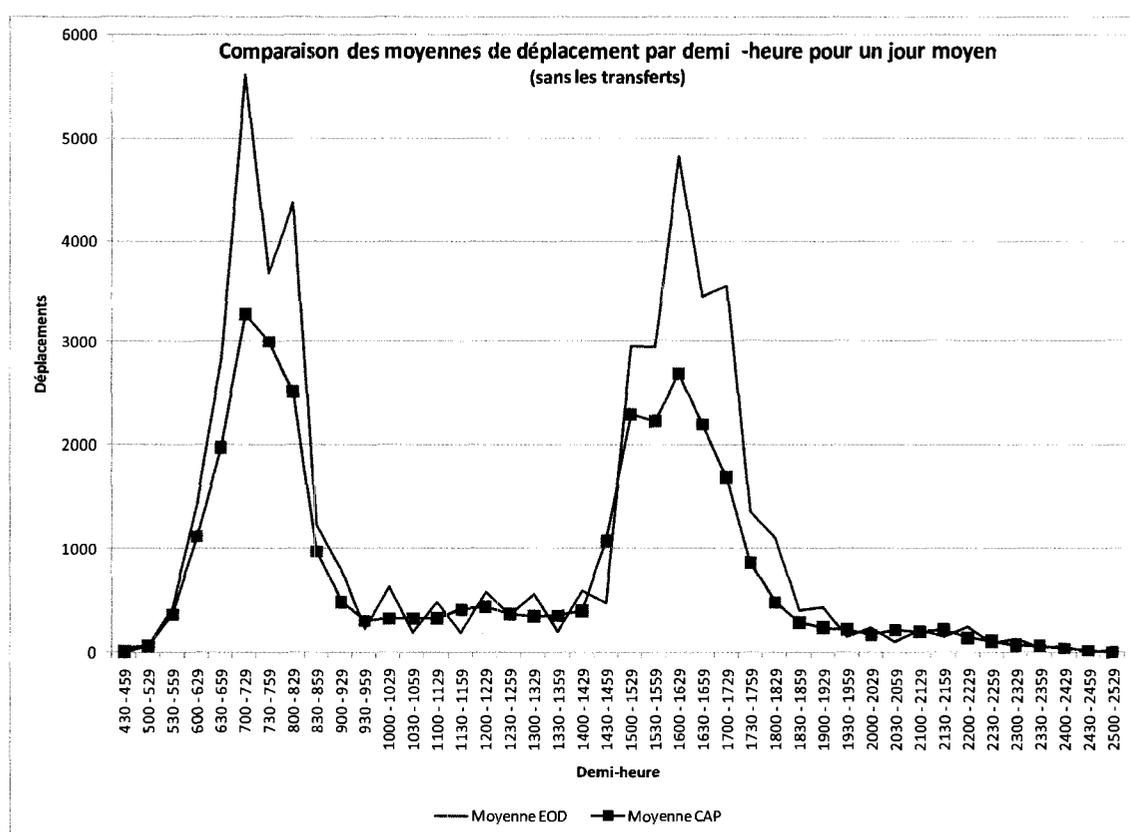


Figure 5.11: Comparaison des moyennes de déplacement des données CAP et EOD par demi-heure

Ces analyses montrent donc que l'écart entre les données provenant des systèmes de cartes à puce et de l'enquête origine-destination est particulièrement grand lors des heures de pointe, mais que les estimations hors pointe sont plutôt équivalentes. L'analyse plus approfondie des périodes de pointe pour les données des systèmes de CAP a démontré que les habitudes de déplacements des usagers sont plus instables lors de ces heures, et ce, principalement à cause des grandes fluctuations dans les

déplacements les vendredis où l'heure de pointe est devancée. Cette variation explique en partie l'écart important entre ces données et les estimations de l'EOD. Cependant, il n'en demeure pas moins qu'il existe de grandes différences entre les données et que les deux sources semblent parfois contradictoires dans leur portrait des habitudes de déplacements de la population. Donc, la prochaine section portera sur une analyse d'objets plus spécifiques au réseau de la société de transport afin d'établir d'où peut provenir ce biais entre les données.

5.3 Analyse du réseau

L'analyse segmentée vue précédemment permet de mieux comprendre la distribution des déplacements sans toutefois toucher aux particularités du réseau. La prochaine analyse intégrera donc les éléments du réseau comme les lignes d'autobus, les arrêts utilisés et les régions desservies (matrice OD).

5.3.1 Par lignes

Afin de satisfaire à la demande, la société de transport de l'Outaouais offre plusieurs services différents pouvant combler des besoins spécifiques de ses usagers. Entre autres, la STO a mis en place des lignes offrant un service régulier, aux heures de pointe (pointe¹⁰ et express¹¹) et sur un vaste territoire (interzone).

La première partie de cette analyse étudie globalement les montées sur tous les types de lignes. En fait, l'ensemble des données provenant des systèmes de CAP pour toutes les journées enquêtées sont comparées, à la Figure 5.12, au jour moyen EOD selon les lignes. Les carrés noirs sur cette figure représentent les moyennes de l'enquête origine-destination et les autres points correspondent à la somme des montées CAP pour une

¹⁰ Les lignes actives uniquement aux heures pointes sont en service entre 6 h et 9 h et entre 15 heures et 18 heures.

¹¹ Le service express se distingue du service de pointe uniquement par des déplacements plus rapides dus à des zones d'embarquement et de débarquement réservés.

journée. Aussi, plus la couleur des points tend vers le rouge, alors plus cette journée est près du 21 septembre 2005. Inversement, plus le point est bleu, alors cette journée se rapproche davantage du 29 novembre 2005.

Ce graphique a cependant nécessité une reformulation des lignes présentes dans les bases de données. En fait, certaines lignes dans la base de données des CAP ont été segmentées et il a donc fallu trouver leur équivalence parmi les lignes présentes dans les données de l'EOD. Ces équivalences sont présentées à l'annexe A.2.

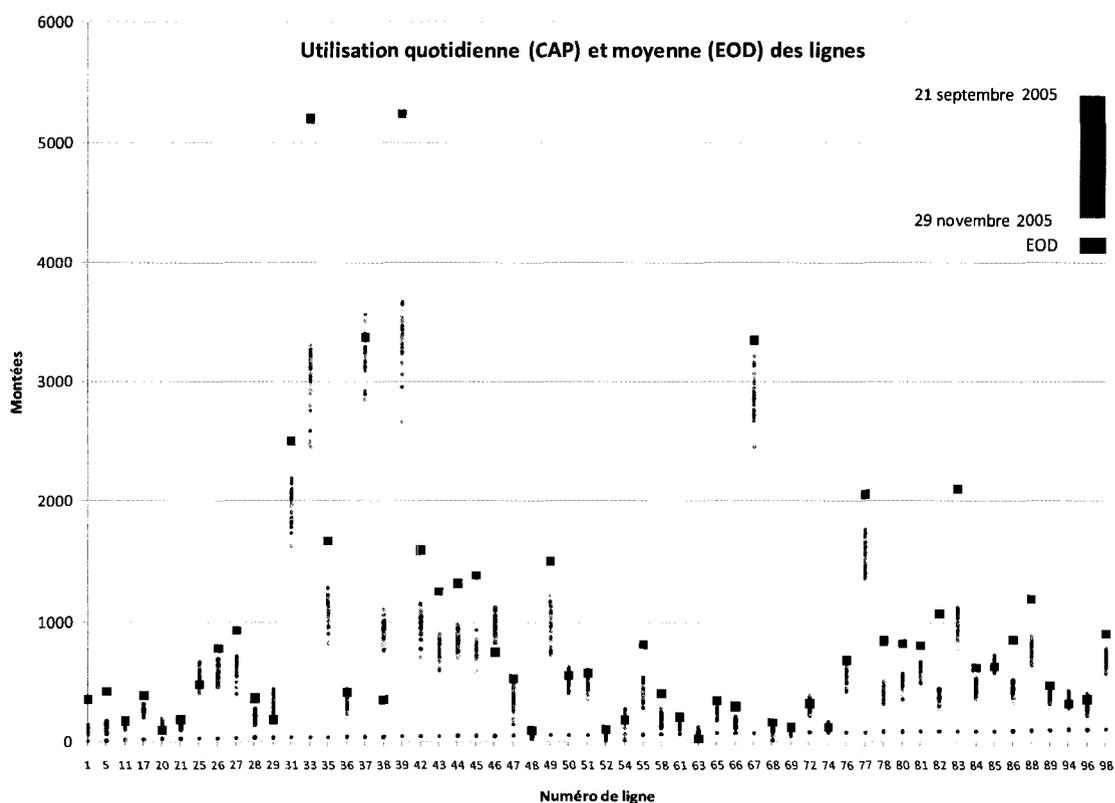


Figure 5.12: Analyse globale des lignes (CAP et EOD)

Cette figure permet, entre autres, de voir la variabilité de l'achalandage sur les lignes. En fait, plusieurs lignes, comme la ligne 33, ont beaucoup de variations dans le nombre de montées et certaines lignes semblent très stables, comme la ligne 83. Aussi, ce graphique montre que les montées effectuées lors des premières journées de

l'enquête (rouge, orange, jaune et vert pâle) ne sont pas regroupées, mais bien réparties uniformément tout comme les données des dernières journées enquêtées (vert, bleu pâle, bleu, bleu foncé). Autrement dit, il ne semble pas y avoir d'augmentation ou de diminution du nombre de montées dans le temps pour une ligne donnée.

On constate aussi un écart parfois très grand entre les données CAP et EOD, par exemple pour les lignes 33, 39 et 83, et parfois très minime, comme pour les lignes 50, 51 et 85. Une sous-estimation de l'EOD pour certaines lignes, par exemple la ligne 46, ressort aussi de cette analyse. Ceci est particulier puisque précédemment, toutes les analyses montraient une surestimation de l'EOD par rapport aux données CAP. Ce phénomène sera étudié davantage lors des analyses par lignes spécifiques qui sont décrites plus loin dans cette section.

L'analyse par ligne peut aussi être segmentée en analyses plus précises selon le type de service utilisé. En fait, les prochains graphiques traiteront les lignes selon leurs appartenances à un type de service en particulier. La première analyse concerne uniquement les lignes express. Donc, les colonnes vertes de la Figure 5.13 représentent la moyenne EOD du nombre de montées sur une ligne et les « X » blancs représentent la moyenne CAP. La ligne noire consiste en la valeur moyenne CAP plus ou moins l'écart type. Lorsque la ligne noire touche au sommet d'une colonne verte, alors cela signifie que l'estimation EOD du nombre de montées est comprise dans l'intervalle de la moyenne CAP plus ou moins l'écart type. On constate que seule la ligne 85 est comprise dans cet intervalle et que toutes les autres lignes sont surestimées par l'enquête origine-destination comparativement aux données de cartes à puce.

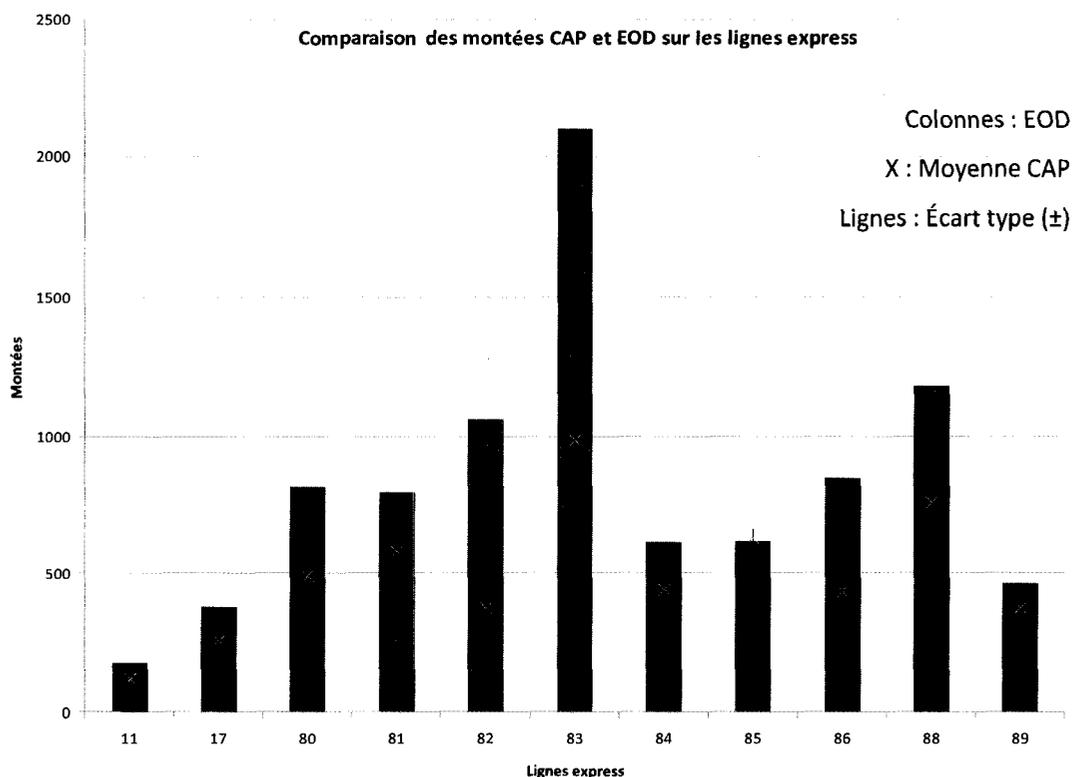


Figure 5.13: Analyse des lignes express (CAP et EOD)

Il est aussi possible de comparer la variabilité des données provenant des systèmes de cartes à puce à l'aide des écarts types présents sur ce graphique. Par exemple, si l'on considère uniquement les lignes 84 et 86, qui ont une moyenne (« X ») approximativement identique (à la même hauteur), on remarque que l'écart type pour la ligne 86 est beaucoup plus grand (ligne noire beaucoup plus longue) que pour la ligne 84. On peut donc affirmer que les données de la ligne 86 varient plus que pour la ligne 84. On peut donc affirmer que l'écart entre l'estimation de l'EOD est plus grand pour la ligne 86 que la ligne 84. On peut donc se questionner sur l'impact qu'a la variabilité, des données des CAP, sur l'écart avec les données de l'EOD. En fait, en analysant toutes ces lignes, on constate bien qu'il n'y a aucune relation entre la valeur de l'écart type et la différence des estimations de l'EOD et des CAP.

Le même exercice peut être fait avec les lignes régulières (Figure 5.14) et les lignes aux heures de pointe (Figure 5.15)¹². L'étude des lignes régulières est d'ailleurs forte intéressante puisque plusieurs de ces lignes sont les plus achalandées du réseau et, par conséquent, elles représentent des grandes lignes de désirs.

Donc, l'analyse de la Figure 5.14 démontre que quelques lignes (21, 37, 50, 72, 74 et 96) ont des estimations très similaires. Deux lignes sont aussi sous-estimées par l'EOD, soit les lignes 38 et 63, et plusieurs lignes sont surestimées. Cependant, l'écart entre l'estimation EOD et CAP est souvent petit, sauf pour les lignes 33 et 39 où cet écart est pratiquement le double de l'estimation des CAP. Les lignes 33 et 39 sont d'ailleurs particulières puisqu'elles font partie des cinq lignes les plus achalandées du réseau et elles ont de petits coefficients de variation, ce qui laisse présager que les usagers empruntant cette ligne sont très routiniers dans leurs habitudes de déplacements. Il est donc tout à fait surprenant que les estimations sur ces deux lignes soient si distantes.

On remarque aussi que, généralement, les écarts types des lignes régulières sont très faibles. En fait, si l'on compare ces écarts types à ceux des lignes express, la variabilité des lignes régulières est beaucoup plus stable.

¹² Vous pouvez aussi voir à l'annexe B.2 l'analyse des deux lignes interzones (94 et 98).

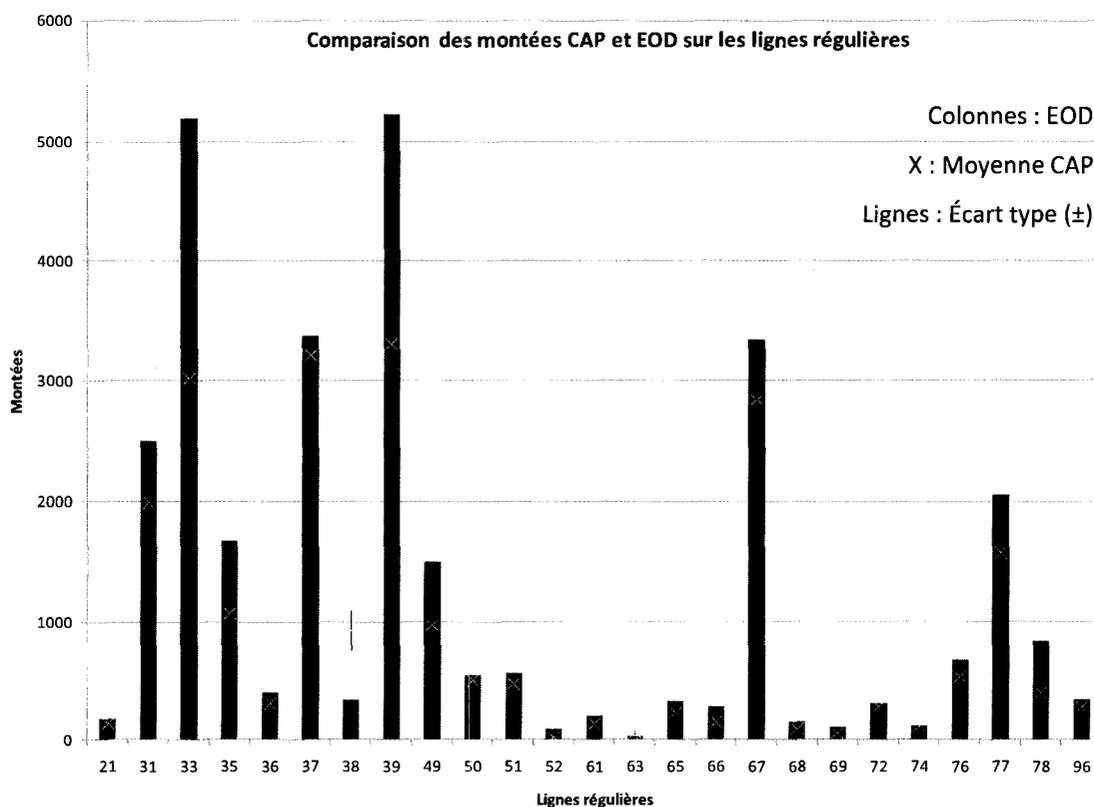


Figure 5.14: Analyse des lignes régulières (CAP et EOD)

Les lignes offertes uniquement aux heures de pointe, illustrées sur la Figure 5.15, sont par contre plus nombreuses à être sous-estimées par l'enquête origine-destination. En fait, cinq lignes (20, 25, 29, 46 et 54) ont une moyenne, selon les données des CAP, supérieure au jour moyen EOD pour cette ligne. Ces lignes ont cependant moins de montées que les lignes régulières précédentes et le nombre de montées CAP par journée varie beaucoup. Concrètement, cette variation pourrait signifier que les usagers ont plus d'une possibilité pour se déplacer et que, certains jours, ils utilisent une ligne et d'autres jours ils en prennent une autre, dépendamment de l'autobus qui arrive en premier. Ceci est d'autant plus plausible puisque plusieurs de ces lignes croisent d'autres lignes sur le réseau ou partagent les mêmes arrêts. Autrement dit, il est fort possible que les usagers utilisant les lignes lors des heures de pointe ne soient pas fidèles à une ligne en particulier, mais qu'ils prêtent plutôt allégeance à une

destination unique et à l'autobus pouvant les amener à l'endroit voulu le plus rapidement.

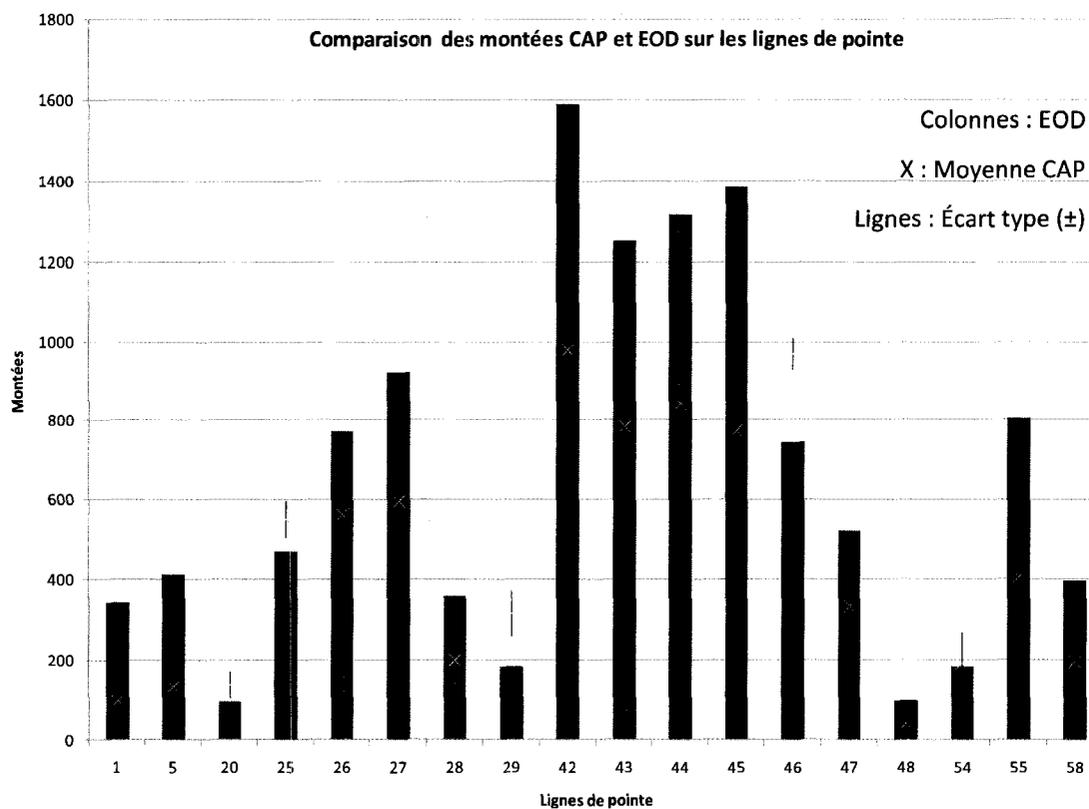


Figure 5.15: Analyse des lignes de pointe (CAP et EOD)

Ces analyses contiennent donc énormément d'informations et peuvent du même coup être segmentées davantage afin d'identifier les divergences et les similitudes dans les estimations. La combinaison des analyses par ligne et par titre de transport, qui suit, permet donc d'obtenir une segmentation des données encore plus désagrégée et ainsi comprendre davantage les facteurs pouvant influencer les données. Cependant, il serait fastidieux d'analyser tous les trajets d'autobus selon tous les titres de transport puisqu'on répertorie plus de 50 trajets d'autobus et environ 10 titres de transport différents. Un choix de ligne a donc été fait afin de simplifier les analyses et d'obtenir une représentation équitable des différentes estimations du réseau. Ces lignes, 46, 50

et 67, ont donc été choisies afin d'étudier les habitudes de déplacements des usagers selon les titres de transport utilisés. Ce choix a été basé sur les critères suivants :

- que les lignes aient un grand achalandage;
- que des titulaires de carte « adulte régulier » et « étudiant régulier » empruntent les lignes;
- qu'au moins une ligne soit sous-estimée globalement par l'EOD comparativement aux données CAP (46);
- qu'au moins une ligne soit surestimée globalement par l'EOD comparativement aux données CAP (67);
- qu'au moins une ligne ait une estimation globale de l'EOD similaire aux données CAP (50).

De plus, seulement les titres de transport « adulte régulier » et « étudiant régulier » seront analysés. Cette décision est fondée sur le nombre important de déplacements fait par ces titulaires et les raisons mentionnées lors de l'analyse par titre de transport.

Le Tableau 5.8 résume les statistiques globales de ces trois lignes tandis que la Figure 5.16 met en évidence la variabilité quotidienne sur les lignes ainsi que l'écart entre les deux estimations. Le comportement des usagers sur ces lignes ne déroge pas des comportements préalablement étudiés lors des analyses segmentées. En fait, les baisses d'achalandage ont lieu principalement lors des journées avant et après les congés du 10 octobre 2005 et du 11 novembre 2005 et lors de certains lundi et vendredi. Ceci est important puisqu'un comportement différent signifierait qu'une ligne peut être sujette à des facteurs différents de ceux qui affectent l'ensemble du réseau. On constate aussi que la ligne ayant le plus faible écart avec les données de l'EOD est aussi la ligne ayant le plus grand coefficient de variation et la moyenne la plus petite.

Tableau 5.8: Analyse des lignes 46, 50 et 67 (CAP et EOD)

Lignes	Moyenne ¹ (montées)	Écart type ¹ (montées)	Coefficient de variation (%)	Moyenne EOD ¹ (montées)
46	974	84	8,62	745
50	509	65	12,6	551
67	2849	170	5,95	3346

¹ Les nombres sont arrondis à l'entier supérieur

En analysant la variation quotidienne des montées, on remarque pour certaines journées l'achalandage baisse sur les trois lignes, mais que d'autres journées on assiste à des hausses de l'achalandage sur une ligne et une diminution sur les autres. En fait, si on regarde la journée du 7 octobre, on constate bien que les trois lignes ont eu une baisse d'achalandage. Par contre, lors de la journée du 17 novembre, l'achalandage sur la ligne 67 augmente considérablement tandis qu'elle baisse sur la ligne 50 et elle se maintient sur la ligne 46.

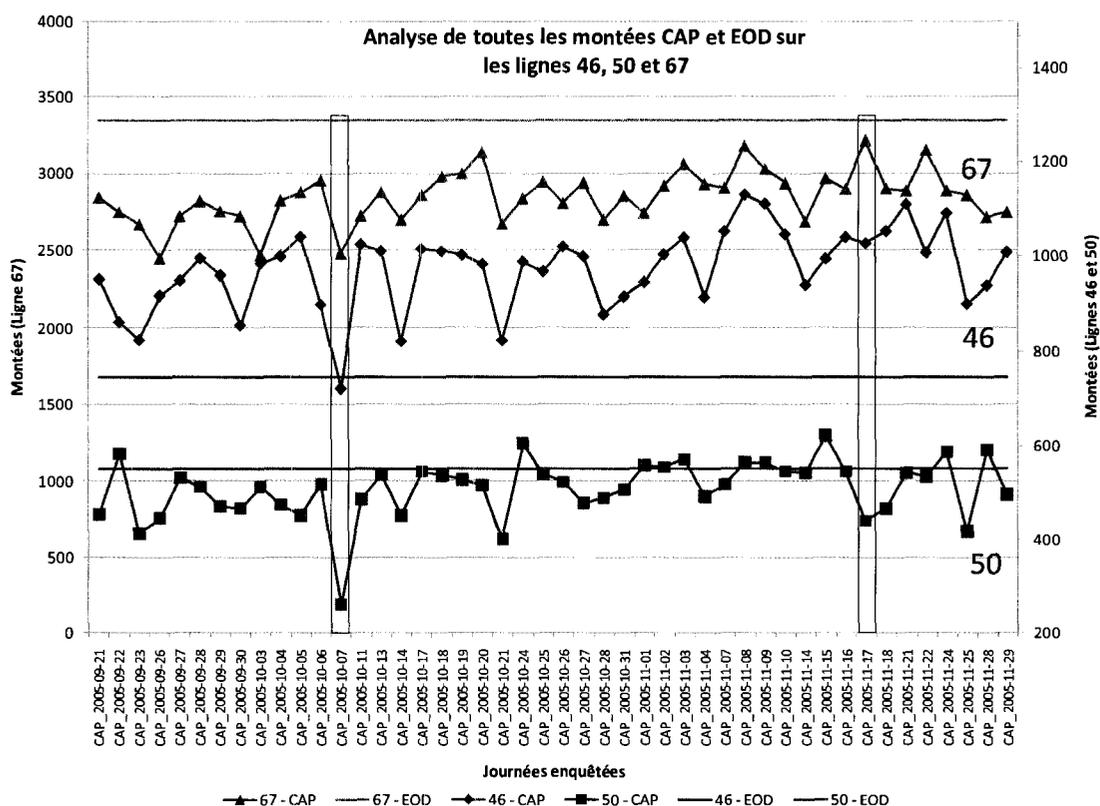


Figure 5.16: Analyse des montées sur les lignes 46, 50 et 67 (CAP et EOD)

Ceci prouve donc que les habitudes de déplacements des usagers diffèrent d'une ligne à l'autre, mais qu'il est aussi possible d'avoir des comportements similaires lors de contextes particuliers.

La segmentation par titre de transport (adulte régulier et étudiant régulier seulement) de la ligne 46 est présentée au Tableau 5.9. La majorité des déplacements sur cette ligne est faite par les adultes et les étudiants sont beaucoup moins représentés. Ceci est vrai autant pour les données provenant de l'enquête origine-destination que des cartes à puce. On constate aussi un coefficient de variation plus élevé pour les étudiants. Cependant, l'écart relatif entre l'estimation de l'EOD et des CAP pour ce groupe est beaucoup plus important que pour les adultes.

Tableau 5.9: Analyse par titre de transport de la ligne 46 (CAP et EOD)

Titre de transport	Moyenne ¹ (montées)	Écart type ¹ (montées)	Coefficient de variation (%)	Moyenne EOD ¹ (montées)
Tous	974	84	8,62	745
Adultes (réguliers)	802	80	9,89	571
Étudiants (réguliers)	157	22	13,5	90

¹ Les nombres sont arrondis à l'entier supérieur

La Figure 5.17 montre l'utilisation quotidienne de la ligne 46 pour les adultes et les étudiants. L'estimation des déplacements des étudiants selon l'EOD ne concorde pas avec les données CAP. Ceci peut probablement être en partie expliqué par une très grande variabilité quotidienne des données des systèmes de CAP sur les étudiants. On peut aussi observer une légère diminution du nombre de montées vers les derniers jours de novembre pour les étudiants. Ce phénomène est totalement inversé pour les adultes puisque la moyenne du nombre de montées augmente plus on s'approche du 29 novembre.

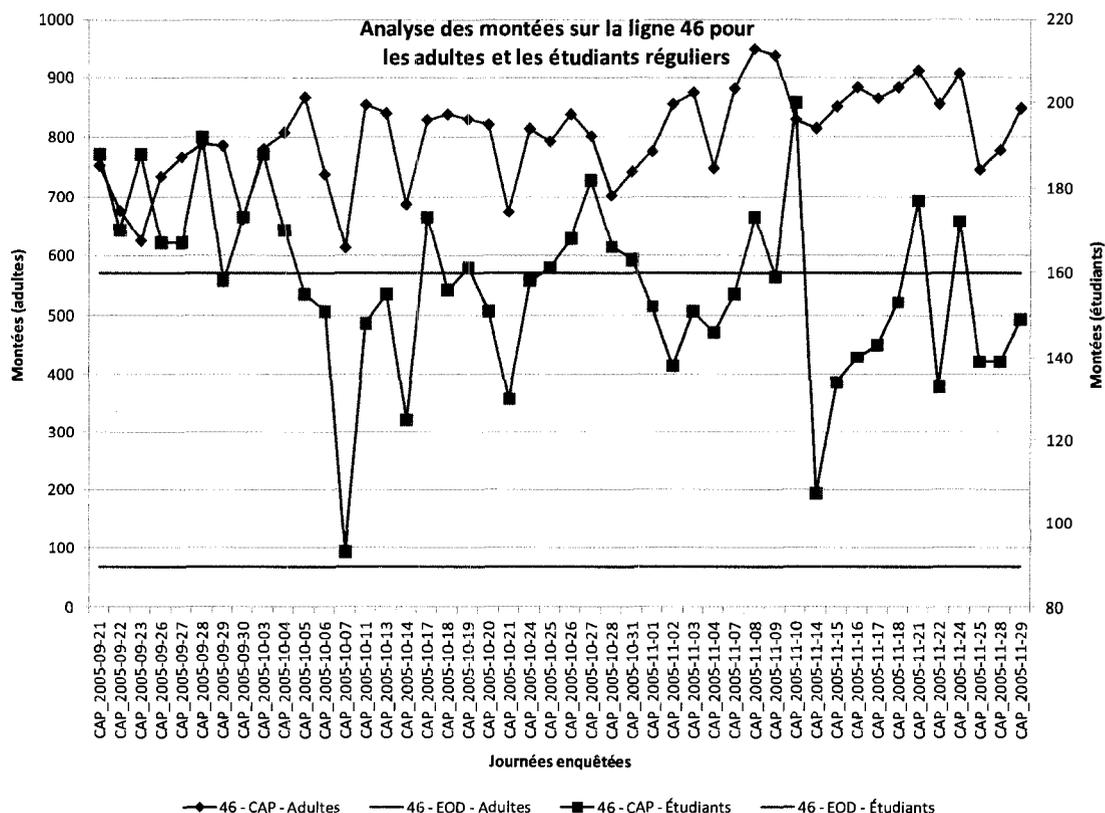


Figure 5.17: Analyse des montées sur la ligne 46 pour les adultes et les étudiants réguliers (CAP et EOD)

De plus, en aucun moment l'estimation des données des CAP pour les adultes et les étudiants n'est égale ou inférieure à l'estimation selon l'EOD. Si l'on considère que les données CAP sont fiables, alors il est clair que ce groupe particulier d'utilisateurs a été mal pondéré puisque l'écart est, en moyenne, de 230 montées de moins selon l'EOD pour les adultes et de 70 montées pour les étudiants. Bref, la comparaison globale de la ligne indique une sous-estimation et lorsqu'on analyse les deux principaux titres de transport utilisés sur cette ligne, on constate aussi une sous-estimation pour chacun de ces titres.

Cette analyse est répétée pour la ligne 50, une ligne où, initialement, les estimations des CAP et de l'EOD correspondaient. Similairement à la ligne 46, les montées faites par les adultes réguliers sont plus nombreuses que celles des étudiants (Tableau 5.10). On observe aussi des coefficients de variation plus élevés pour les étudiants que pour les adultes. Si l'on compare ce tableau à celui vu précédemment pour la ligne 46, on

constate que les étudiants sont encore sous-estimés, mais que les adultes sont très bien représentés.

Tableau 5.10: Analyse par titre de transport de la ligne 50 (CAP et EOD)

Titre de transport	Moyenne¹ (montées)	Écart type¹ (montées)	Coefficient de variation (%)	Moyenne EOD¹ (montées)
Tous	509	65	12,6	551
Adultes (réguliers)	391	52	13,17	391
Étudiants (réguliers)	115	26	22,33	73

¹ Les nombres sont arrondis à l'entier supérieur

En fait, les estimations CAP et EOD pour les adultes réguliers sont équivalentes dans la mesure où leurs moyennes sont similaires, tandis que les estimations des étudiants réguliers sont légèrement distantes. Les coefficients de variation sont très élevés et ceci est représenté sur la Figure 5.18 par une distribution en dents de scie. Il faut cependant noter qu'il y a moins de montées sur ces lignes et que la moindre variation dans les habitudes de déplacement de certains titulaires peut avoir des répercussions importantes sur ces indicateurs.

Bien que la moyenne de l'EOD soit identique à la moyenne des déplacements des adultes comptabilisés par les CAP, on remarque toutefois qu'une droite constante n'est pas représentative du portrait des déplacements selon les CAP. En fait, lors des premiers jours de l'enquête, on remarque que les données provenant des CAP sont inférieures à la moyenne de l'EOD et plus on s'approche des derniers jours enquêtés, plus les données des CAP sont supérieures au jour moyen de l'EOD.

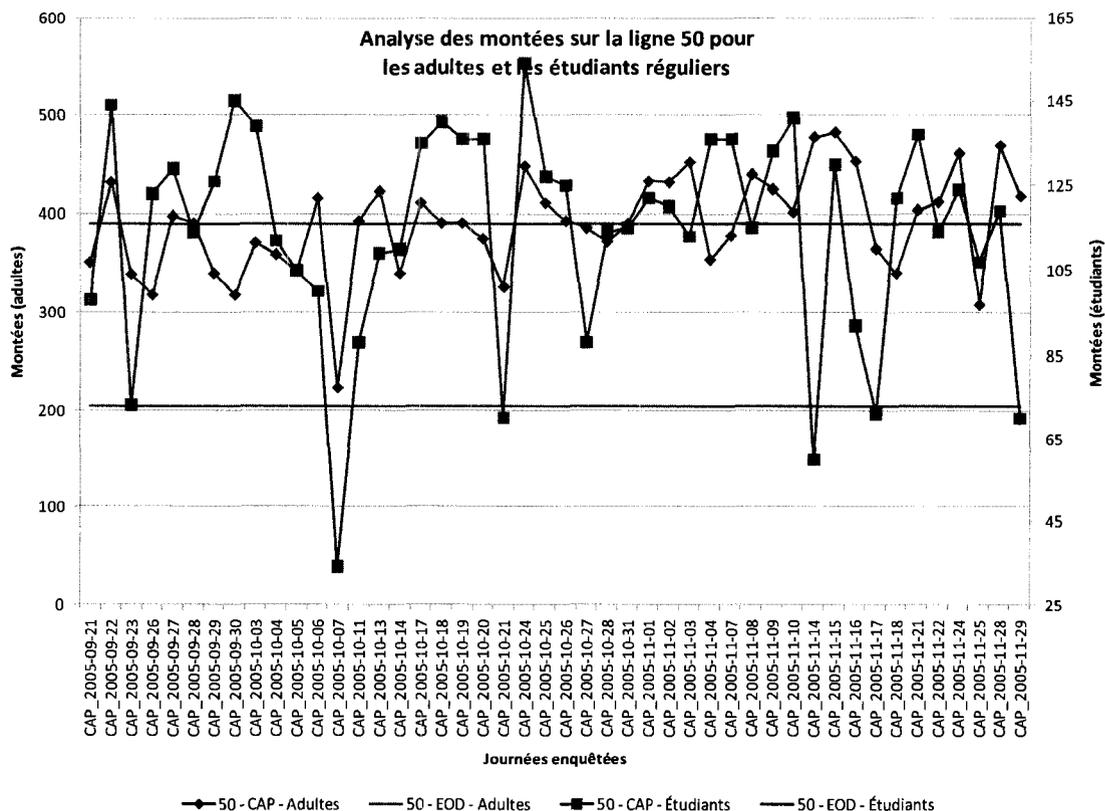


Figure 5.18: Analyse des montées sur la ligne 50 pour les adultes et les étudiants réguliers (CAP et EOD)

L'analyse se poursuit avec la ligne 67 qui a un achalandage au-dessus de la moyenne de la STO et nettement supérieur aux deux lignes étudiées précédemment. De plus, l'estimation globale extraite de l'enquête origine-destination est légèrement supérieure à la moyenne de montées calculée grâce aux données de carte à puce. Lorsque l'on analyse le tableau 5.11 et la Figure 5.19, on remarque que cet écart est principalement imputable à l'estimation des montées faites par les étudiants réguliers. En fait, l'écart entre les données de l'EOD et des CAP pour les adultes (141 montées) est inférieur à l'écart pour les étudiants (234 montées), et ce, malgré le fait que les adultes sont plus représentés (1701 montées) que les étudiants (733 montées). On remarque aussi que le coefficient de variation des adultes (7,39 %) est inférieur, bien que très similaire, à celui des étudiants (6,56 %). De plus, ces coefficients de variation sont beaucoup moins élevés que ceux des deux lignes précédentes.

Tableau 5.11: Analyse par titre de transport de la ligne 67 (CAP et EOD)

Titre de transport	Moyenne ¹ (montées)	Écart type ¹ (montées)	Coefficient de variation (%)	Moyenne EOD ¹ (montées)
Tous	2849	170	5,95	3346
Adultes (réguliers)	1701	126	7,39	1842
Étudiants (réguliers)	733	48	6,56	967

¹ Les nombres sont arrondis à l'entier supérieur

Similairement aux deux lignes précédentes, on constate une augmentation progressive du nombre de montées pour les adultes. On remarque aussi que les baisses et les hausses d'achalandage ont pratiquement toujours lieu en même temps sur les deux courbes des données CAP. Ce phénomène n'a cependant pas été observé sur les lignes précédentes.

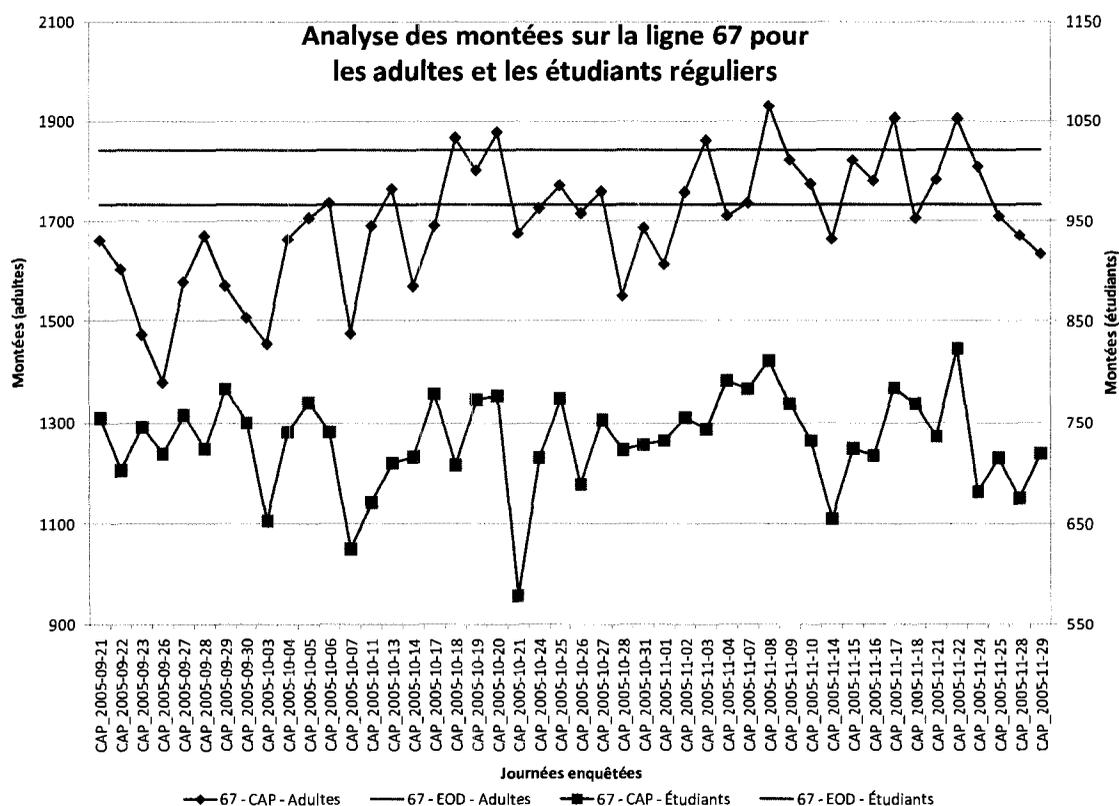


Figure 5.19: Analyse des montées sur la ligne 67 pour les adultes et les étudiants réguliers (CAP et EOD)

En analysant les déplacements par ligne et de certains titres de transport, on constate des différences importantes entre les coefficients de variation. Il est donc intéressant de regarder plus en détail les habitudes de déplacements de la population qui emprunte la ligne 50 et la ligne 67 afin de mieux comprendre l'écart entre ces coefficients de variation.

La Figure 5.20 découpe les usagers des lignes 50 et 67 selon leurs habitudes de déplacement. Par exemple, 39 % de toutes les personnes ayant emprunté la ligne 67 au cours des 48 jours enquêtés ont utilisé moins d'une fois par mois cette ligne comparativement à 46,5 % sur la ligne 50.

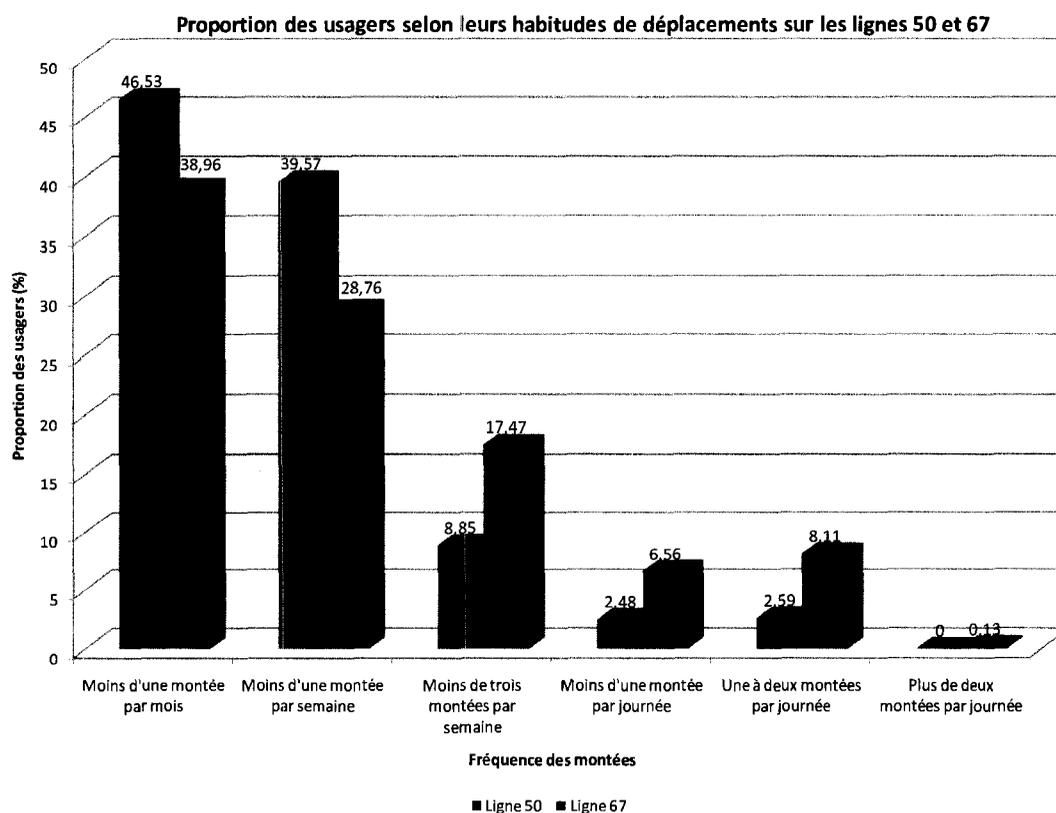


Figure 5.20: Proportion des usagers selon les fréquences de montées sur les lignes 50 et 67 (CAP)

On remarque donc que la majorité des usagers n'empruntent qu'occasionnellement, soit moins d'une fois par semaine, ces trajets. Par contre, cette proportion d'individus

empruntant peu fréquemment ces lignes est beaucoup plus importante pour la ligne 50 que la ligne 67 (86 % pour la ligne 50 et 68 % pour la ligne 67).

Ces usagers ne sont responsables que de très peu de déplacements sur la ligne 67 contrairement à la ligne 50 comme l'indique le Tableau 5.12. En fait, chaque jour, environ 17 % des déplacements faits sur la ligne 67 le sont par des usagers empruntant cette ligne qu'occasionnellement (moins d'une fois par semaine), soit environ de deux fois moins que pour la ligne 50 (près de 39 %). Autrement dit, les usagers n'empruntant la ligne 67 ont des habitudes de déplacements sur cette ligne plus fréquentes que les usagers de la ligne 50.

Tableau 5.12: Nombre de montées quotidiennes sur les lignes 50 et 67 en fonction des habitudes de déplacements

Habitudes de déplacements des usagers	Ligne 50		Ligne 67	
	Moyenne ¹ (montées)	Proportion (%)	Moyenne ¹ (montées)	Proportion (%)
Moins d'une montée par mois	44	8	109	4
Moins d'une montée par semaine	160	31	352	13
Moins de trois montées par semaine	119	23	750	26
Moins d'une montée par journée	70	14	555	19
Une à deux montées par journée	118	24	1047	36
Plus de deux montées par journée	0	0	34	2

¹ Les nombres sont arrondis à l'entier supérieur

À la lumière de ces résultats, il est certain qu'il existe sur cette ligne deux groupes distincts d'usagers, dont l'un est composé de personnes utilisant peu les lignes et de manière très variable et ponctuelle et l'autre est composé de personnes plus constantes dans leur utilisation et qui se déplacent beaucoup plus fréquemment. L'analyse suivante tente de déterminer si l'un de ces groupes d'individus peut majoritairement être imputé aux « adultes réguliers » ou aux « étudiants réguliers ».

La Figure 5.21 montre les proportions des montées effectuées selon la fréquence d'utilisation des lignes 50 et 67 par les usagers « adultes réguliers » et « étudiants réguliers ». Si l'on étudie uniquement la ligne 50, on peut constater que la proportion

des déplacements, faits par des usagers dont les habitudes de déplacements sont de moins de trois fois par semaine (colonnes bleu foncé, rouges et vertes), est sensiblement la même pour les adultes réguliers que pour les étudiants. Autrement dit, les adultes et les étudiants utilisent cette ligne quotidiennement selon des fréquences similaires. On observe d'ailleurs le même phénomène pour la ligne 67, sauf que les étudiants ont une tendance légèrement plus élevée que les adultes à utiliser sporadiquement, soit moins de trois fois par semaine, cette ligne.

Lorsque l'on compare, les deux lignes, on constate que la proportion des déplacements peu fréquents est beaucoup plus élevée pour la ligne 50 que pour la ligne 67, et ce, autant pour les adultes et les étudiants.

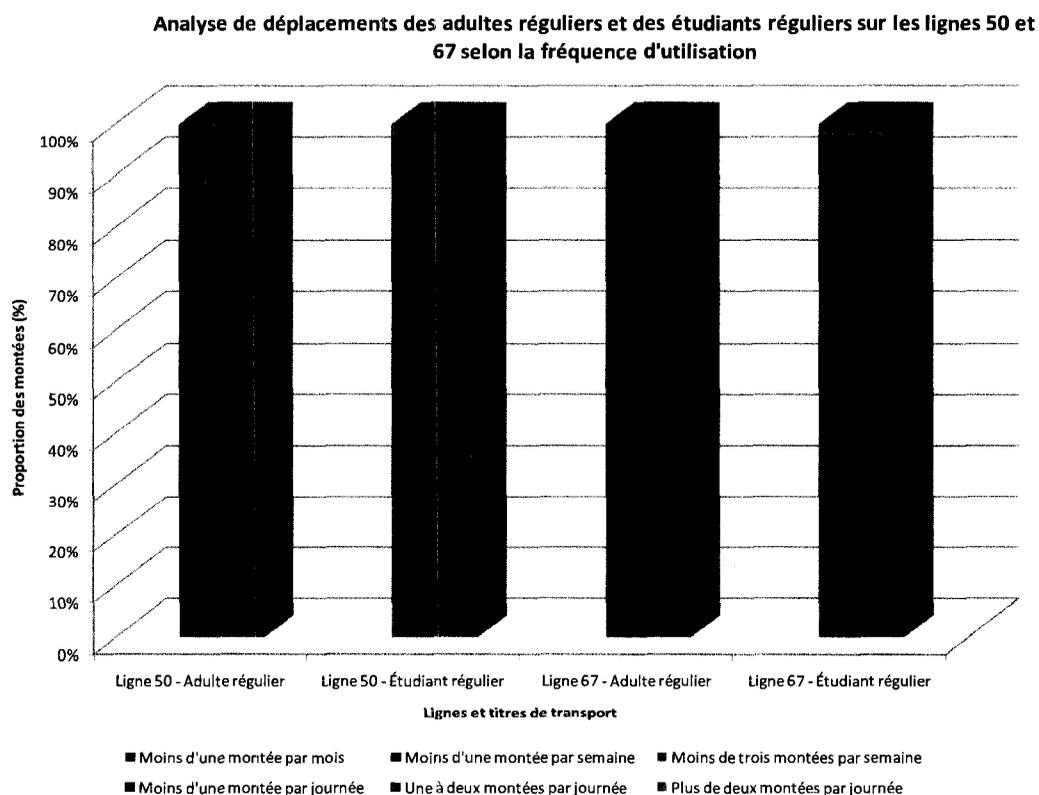


Figure 5.21: Distribution de montées sur les lignes 50 et 67 selon les habitudes de déplacements pour les « adultes réguliers » et les « étudiants réguliers » (CAP)

Par conséquent, la ligne 50 est caractérisée par une plus grande proportion de déplacements faits par des usagers utilisant peu souvent cette ligne, ce qui peut expliquer les grands coefficients de variation obtenus (Tableau 5.10). La ligne 67 est par contre beaucoup moins soumise à des déplacements faits par des usagers plus irréguliers. De plus, on constate que les étudiants et les adultes ont des comportements très similaires pour ces deux lignes. Par conséquent, on ne peut pas conclure et imputer entièrement la variation des données aux étudiants ou aux adultes puisque ceux-ci montrent aussi d'importantes utilisations sporadiques des lignes.

En analysant par ligne les déplacements, on a constaté que parfois les estimations des systèmes CAP correspondaient aux estimations de l'EOD et parfois non. Donc, afin de comprendre davantage les facteurs influençant ces estimations, on a premièrement analysé les lignes en fonction du service qu'elles offrent. Ces analyses n'ont cependant pas permis de trouver un biais systématique pouvant expliquer les différences des estimations. Les analyses ont donc été approfondies en regardant concrètement les déplacements sur trois lignes, les lignes 46, 50 et 67. Le choix de ces lignes a été justifié par le fait qu'elles représentent bien le réseau et qu'elles correspondent soit à une surévaluation, une sous-estimation ou encore une estimation juste de l'EOD par rapport aux systèmes de cartes à puce. Lors des analyses des déplacements quotidiens sur ces lignes, il a été démontré que pour certaines journées l'achalandage suit des schémas similaires sur les trois lignes, mais qu'il est aussi fréquent que le nombre de déplacements varie différemment. Ensuite, une analyse segmentée de ces lignes selon les titres de transport « adulte régulier » et « étudiant régulier » a démontré qu'il y avait des variations particulières pour chaque groupe. En fait, le nombre de déplacements des adultes tend à augmenter lorsque l'on s'approche de la fin novembre et les déplacements des étudiants sont souvent plus en dents de scie, donc plus sporadiques, que ceux des adultes. Ces résultats ont d'ailleurs été corroborés par une étude plus approfondie des fréquences d'utilisation des lignes 67 et 50. Malgré ces

comparaisons et analyses, on observe toujours un biais non systématique entre l'estimation de l'EOD et des systèmes par CAP.

5.3.2 Par origine et destination

Les analyses par origine et destination font partie intégrante des outils de planification dans le secteur du transport en commun. Une analyse comparative de données sur les déplacements de la population ne peut donc pas être complète sans une étude des matrices origine-destination (« OD »). La première étape consiste à construire deux matrices, l'une avec les données de l'EOD et l'autre avec les données CAP, ne prenant en compte que les déplacements CAP.

5.3.2.1 Élaboration de la matrice origine-destination de l'EOD

La base de données de l'EOD comporte deux champs, soit « srorig07 » et « srdest07 », qui indique le secteur de recensement de l'origine et de la destination du déplacement. Ces champs permettront d'établir l'origine et la destination, selon l'EOD, des déplacements, seulement si ces conditions sont vérifiées :

- les correspondances ne doivent pas être prises en compte;
- seulement les déplacements débutant sur le réseau de la STO peuvent avoir l'origine du champ srorig07;
- seulement les déplacements finissant sur le réseau de la STO peuvent avoir la destination du champ srdest07.

L'exclusion des correspondances s'explique simplement par la définition même de la matrice OD, soit une représentation des points de départ et d'arrivée de la chaîne de déplacements. Par exemple, les correspondances « 2 » et « 3 » de la Figure 5.22 ne devraient pas être incluses dans l'analyse par matrice OD puisque celles-ci ne modifient en rien l'origine (« 1 ») et la destination (« 4 ») de la chaîne de déplacements.

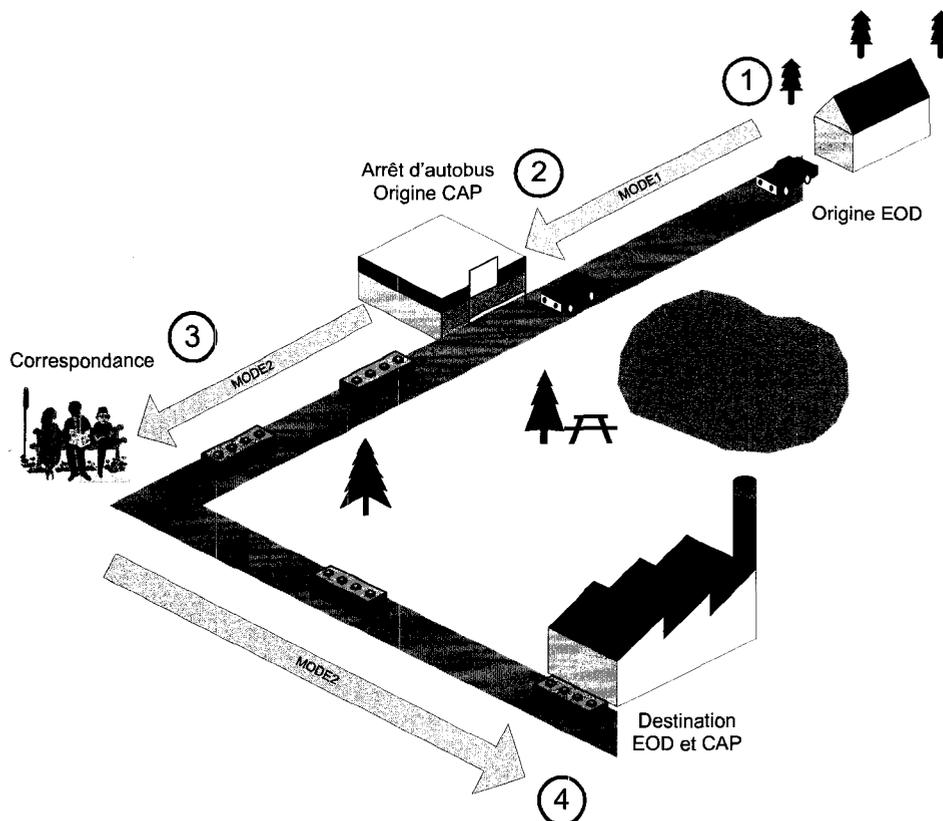


Figure 5.22: Analyse d'une chaîne de déplacements provenant de l'EOD

Les deux autres conditions sont particulières à la comparaison des données de l'EOD et des CAP. En fait, la base de données des CAP détaille un seul et unique mode de transport, soit le transport en commun. Qui plus est, seulement les déplacements CAP sur le territoire de la STO sont enregistrés. Par conséquent, on ne dispose pas actuellement de données sur la chaîne de déplacements complète, mais on peut facilement reconstruire les voyages faits sur le réseau de transport en commun. Il va donc de soi que les matrices OD doivent tenir compte uniquement des déplacements sur le réseau, et ce, même pour les chaînes de déplacements de l'EOD qui sont habituellement complètes. Autrement dit, les déplacements de l'EOD devront être limités aux déplacements TC sur le réseau de la STO où l'origine est la zone initiale d'embarquement et la destination est la zone finale de débarquement. Si on se réfère à l'exemple de la Figure 5.22, l'origine serait le point « 2 » et la destination demeurerait

le point « 4 ». Cependant, la base de données contient que l'origine et la destination de la chaîne complète, soit dans l'exemple les points « 1 » et « 4 ». Donc, les champs « srorig07 » et « srdest07 » ne sont utilisés que lorsque la chaîne de déplacements débute ou finit sur le réseau de la STO.

Pour tous les autres déplacements, deux situations peuvent survenir. Premièrement, le déplacement peut commencer sur un autre mode de transport que les autobus de la STO et se finir sur une ligne de la société québécoise. Dans ce cas, la destination indiquée par le champ « srdest07 » est valide, tandis que l'origine ne l'est pas. Deuxièmement, la chaîne de déplacement débute sur le réseau de la STO, mais elle se termine sur un autre mode de transport. Dans ce cas, l'origine déterminée par le champ « srorig07 » est valide et la destination ne l'est pas. Dans ces deux situations, il faut donc déterminer soit l'origine ou la destination sans utiliser les champs mentionnés précédemment. Afin de trouver le bon point d'embarquement sur le réseau ou de débarquement, deux approches sont possibles, soit :

- La recherche du lieu de correspondance grâce à la table « JONCTION »;
- La recherche du lieu de correspondance grâce aux lignes empruntées.

La première méthode est utilisée uniquement si le type de jonction n'est pas une correspondance entre un autobus de la STO et de l'OC Transpo, et elle consiste à utiliser un champ similaire à « srorig07 », soit « srj107 », qui contient le secteur où a lieu la correspondance. Par contre, si la jonction est de type TC à TC, alors aucune information sur la localisation de la correspondance ne se retrouve dans la base de données. Il faut donc étudier les lignes de la STO et de l'OC Transpo qui ont été utilisées lors du déplacement, et ce, afin de trouver le point de rencontre le plus probable. Heureusement, la géographie de la région ainsi que la localisation des terminus aident grandement à faire une estimation très précise des lieux de correspondances.

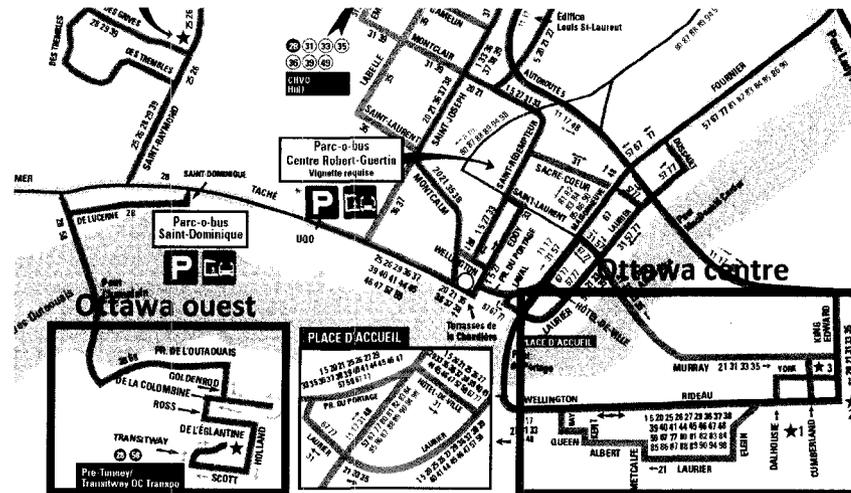


Figure 5.23: Régions ontariennes desservies par la STO (Source: STO)

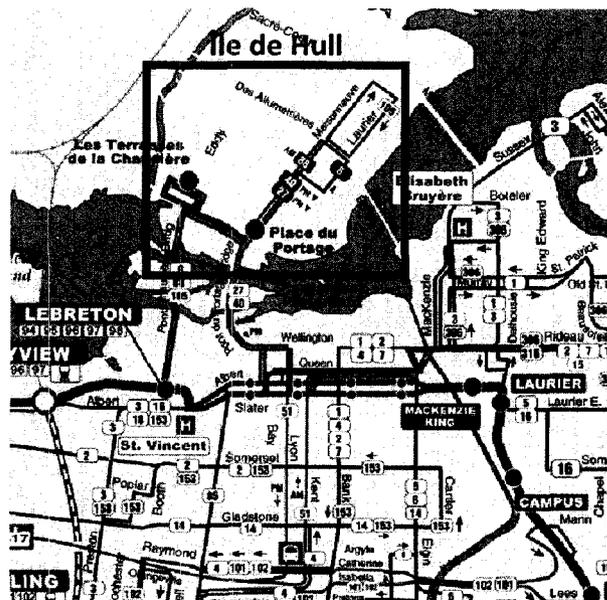


Figure 5.24: Régions québécoises desservies par l'OC Transpo (Source: OC Transpo)

En fait, on peut diviser les lignes en trois catégories, soit :

- Les lignes Ottawa Ouest : STO 28, 58;
- Les lignes Île de Hull : OC Transpo 8, 27, 40, 88, 105;
- Les lignes Ottawa Centre : toutes les autres lignes permettant un transfert STO-OC Transpo.

Les lignes de type « Ottawa Ouest » représentent des lignes où le terminus se situe dans la zone municipale d'Ottawa ouest (Figure 5.23). Lorsque ces lignes font partie d'une correspondance avec le réseau d'OC Transpo, il est certain que celle-ci a lieu dans la zone municipale d'Ottawa ouest puisque c'est le seul endroit où ces autobus peuvent se rencontrer. Similairement, les lignes de l'Île de Hull sont accessibles aux usagers de la STO seulement dans la région de l'Île de Hull (Figure 5.24). Les déplacements qui ne comportent pas les deux types de ligne précédente se font invariablement dans la région municipale d'Ottawa centre où sont situés trois terminus importants de la STO. En agissant de la sorte, il est possible de créer une matrice qui ne comporte comme origines et destinations que les zones desservies par la STO (Tableau 5.13).

Tableau 5.13: Régions municipales des origines et des destinations

Région municipale	Abréviation
Aylmer	AY
Gatineau Est	G-E
Hull Périphérie	H-P
Île de Hull	I-H
Masson-Angers	M-A
Ottawa Centre	O-C
Ottawa Ouest	O-O
Plateau	PL
Gatineau Centre	G-C
Rural Nord-est	R-NE
Rural Nord-ouest	R-NO

Cette matrice, représentée par le tableau 5.14, permet donc d'étudier les origines et les destinations, sur le réseau de la STO uniquement, des titulaires d'une carte à puce. Parmi les grandes lignes de désir, présentées en gras dans ce tableau, on constate que les régions d'Aylmer, de Gatineau Est, de l'Île de Hull, d'Hull périphérie, du Plateau, d'Ottawa centre et de Gatineau centre sont les régions qui génèrent le plus de déplacement. En faisant la somme des lignes et des colonnes, les régions de Hull périphérie et d'Ottawa centre sont les origines et les destinations les plus achalandées.

En fait, plus de 40 % des déplacements débutent ou se destinent à l'une ou l'autre de ces régions.

Tableau 5.14: Matrice OD des déplacements de l'EOD

O	D	AY	G-E	H-P	I-H	M-A	O-C	O-O	PL	G-C	R-NE	R-NO	Total
AY		1841	85	896	522	15	1103	149	13	135	17	17	
G-E		69	94	616	408	18	1280	65	0	478	21	0	
H-P		783	515	3746	1703	54	1415	74	725	1449	163	86	
I-H		536	389	1668	84	173	203	0	450	460	88	31	
M-A		15	0	54	173	52	220	14	0	67	0	0	
O-C		1133	1151	1232	153	208	19	0	512	1603	312	80	
O-O		165	75	110	52	30	0	0	108	86	19	0	
PL		49	0	648	539	0	604	108	132	165	0	0	
G-C		125	400	1437	437	87	1657	73	201	954	94	65	
R-NE		17	0	163	104	0	327	19	0	77	0	13	
R-NO		17	0	86	51	0	90	0	0	65	13	0	
Total		5730	3502	11675	4694	775	10700	556	2530	6114	810	335	47 400

Les déplacements internes, soit la diagonale de la matrice, sont importants pour les régions de Aylmer, de Gatineau centre et de Hull périphérie, mais ils sont pratiquement inexistantes pour les deux régions ontariennes desservies par la STO, soit Ottawa centre et Ottawa ouest. L'absence de déplacements internes pour les régions ontariennes est tout fait normale puisque la STO n'assure pas de service dans ces régions outre les terminus permettant les correspondances sur le réseau de l'OC Transpo.

La prochaine étape est donc de créer la matrice OD des données CAP afin de pouvoir comparer les deux matrices.

5.3.2.2 *Élaboration de la matrice origine-destination des données CAP*

La procédure d'élaboration de la matrice des données CAP diffère grandement de celle vue précédemment. En fait, les données CAP représentent uniquement les déplacements CAP sur le réseau STO et, par conséquent, elles ne nécessitent pas de

traitement afin d'éliminer les autres modes de déplacements. Cependant, les données CAP sont totalement désagrégées et cela implique que la chaîne de déplacements est enregistrée dans la base de données comme plusieurs déplacements uniques. En fait, si on reprend l'exemple de la Figure 5.22, la base de données des CAP comporterait deux déplacements pour cette chaîne, soit le déplacement « 2 » vers « 3 » étiqueté comme première montée et le déplacement « 3 » vers « 4 » étiqueté comme correspondance. Il suffit donc de retracer l'origine du premier déplacement et la destination de la correspondance.

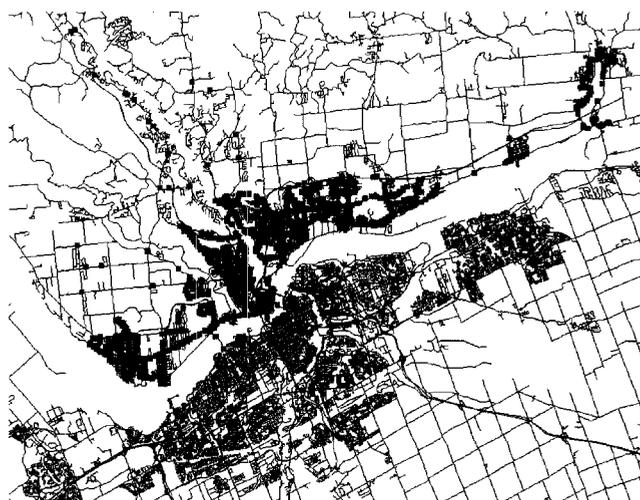


Figure 5.25: Répartition géographique des arrêts

Malheureusement, l'origine du déplacement n'est pas inscrite dans la base de données comme elle l'est pour l'EOD. Cependant, le numéro d'arrêt de la montée est toujours recueilli et il servira donc à établir la zone municipale du déplacement. De plus, tous les arrêts sont déjà géocodés dans la base de données, il suffit donc d'exporter ces coordonnées géographiques dans un logiciel de cartographie, tel que FGis, et d'affecter chaque arrêt à une région. La Figure 5.25 montre donc le géocodage des arrêts dans le logiciel FGis, avec comme arrière-plan la carte des routes de la région de la capitale nationale. Chacun des points rouges correspond à un arrêt, mais aucune région municipale n'est encore attribué. Pour ce faire, la carte a été divisée selon les secteurs

municipaux (Figure 5.26) et les coordonnées géographiques de ces zones et de ces arrêts ont été exportées dans un fichier Excel. De là, l'affectation des arrêts aux régions municipales a été faite grâce à une macro faite sur mesure.

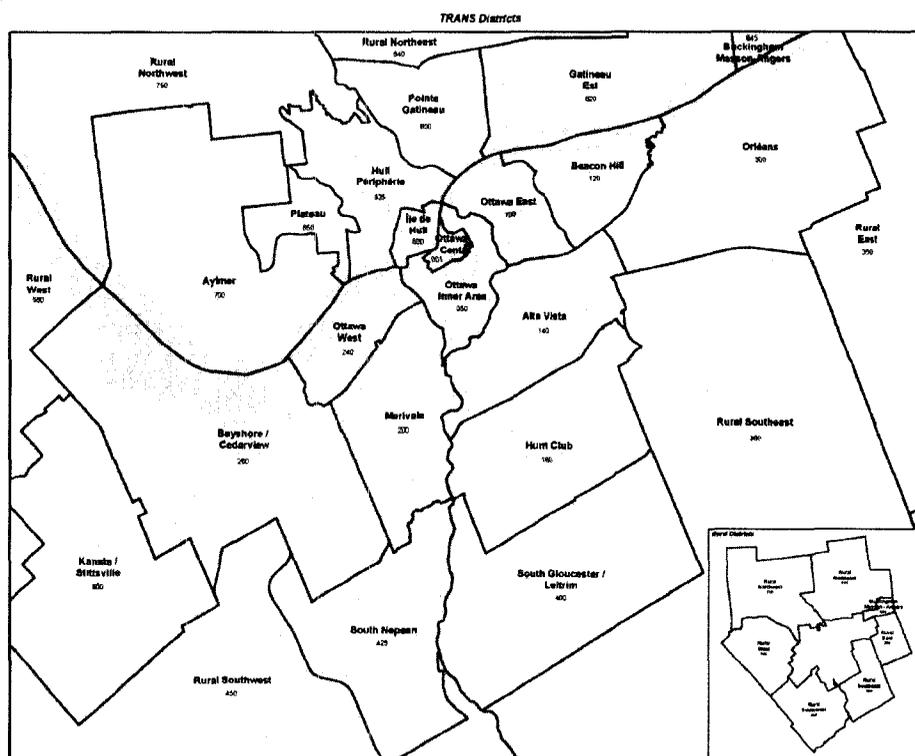


Figure 5.26: Secteurs municipaux de la région de la capitale nationale (Source : iTRANS Consulting inc., (2006))

Avec les zones municipales des arrêts, il est possible de déterminer les origines de tous les déplacements, y compris les correspondances même si celles-ci ne sont pas nécessaires à l'élaboration de la matrice OD.

Afin de compléter la matrice, il faut estimer la destination sur le réseau STO de la chaîne de déplacement puisque celle-ci n'est pas disponible dans les bases de données. Autrement dit, il faut trouver le dernier arrêt où l'utilisateur a débarqué ou, dans l'exemple de la Figure 5.22, l'arrêt correspondant au point « 4 ».

Un algorithme, élaboré par Trépanier, Chapleau, & Tranchant (2007), permettant de retracer la destination finale des déplacements a déjà été testé et validé sur les données de la STO. En fait, cet algorithme a un taux de succès de l'ordre de 66 % pour l'estimation des destinations et, lors des heures de pointe, ce taux augmente jusqu'à 80 %. C'est donc à partir des résultats obtenus par cet algorithme que les destinations, utilisées dans la matrice représentée par le Tableau 5.15, ont été estimées. Cette matrice représente l'équivalent CAP du jour moyen de l'EOD.

Tableau 5.15: Matrice OD des déplacements des CAP

<i>O</i> \ <i>D</i>	AY	G-E	H-P	I-H	M-A	O-C	O-O	PL	G-C	R-NE	R-NO	Total
AY	1510	5	436	463	2	1417	72	22	34	0	1	
G-E	4	324	157	274	8	1046	0	2	246	0	0	
H-P	407	134	3450	1116	35	1773	9	346	582	7	3	
I-H	418	239	920	620	114	399	0	257	504	8	3	
M-A	2	6	36	115	68	383	0	1	28	0	0	
O-C	1338	970	1850	520	387	844	1	513	1894	15	11	
O-O	96	0	13	1	0	1	15	50	1	0	0	
PL	36	2	338	264	1	515	43	272	49	1	0	
G-C	30	223	572	576	29	2077	1	55	1027	2	1	
R-NE	0	1	10	12	0	19	0	0	4	1	0	
R-NO	1	0	5	5	0	14	0	0	1	0	1	
Total	3842	1904	7787	3966	644	8488	141	1518	4370	34	20	32 700

L'analyse de cette matrice démontre que les régions d'Aylmer, de Gatineau est, d'Hull périphérie, de l'Île de Hull, d'Ottawa centre et de Gatineau centre sont les plus grands générateurs de déplacements. De ces régions, Hull périphérie et Ottawa centre sont l'origine et la destination de tout près de 50 % des déplacements. Les déplacements internes aux zones, soit la diagonale de la matrice, représentent 25 % de l'ensemble des déplacements. Les régions ayant le plus de déplacements internes sont les régions d'Aylmer, d'Hull périphérie, de Gatineau centre et d'Ottawa centre. Il est en fait surprenant que la région d'Ottawa centre ait un nombre si important de déplacements internes étant donné que cette région sert principalement comme lieu de

correspondance avec le réseau de l'OC Transpo. La matrice montre aussi que le réseau est très peu utilisé dans les régions Rural nord-est, Rural nord-ouest et Ottawa ouest.

Les deux matrices sont donc formatées identiquement et, surtout, elles ont été construites afin de représenter le même contexte. La prochaine section a pour but de comparer les deux matrices OD.

5.3.2.3 Comparaison des matrices OD

Une première comparaison possible est l'évaluation de la différence entre les deux estimations. Cette comparaison, présentée à la Figure 5.27, met en évidence la surévaluation des estimations de l'EOD par rapport aux CAP pour une majorité de paires OD. En fait, toutes les régions où le cône pointe vers le haut, l'estimation EOD est plus grande que celle provenant des cartes à puce. On remarque que plusieurs paires origine-destination ont un différentiel nul, ce qui signifie que les estimations EOD et CAP sont équivalentes.

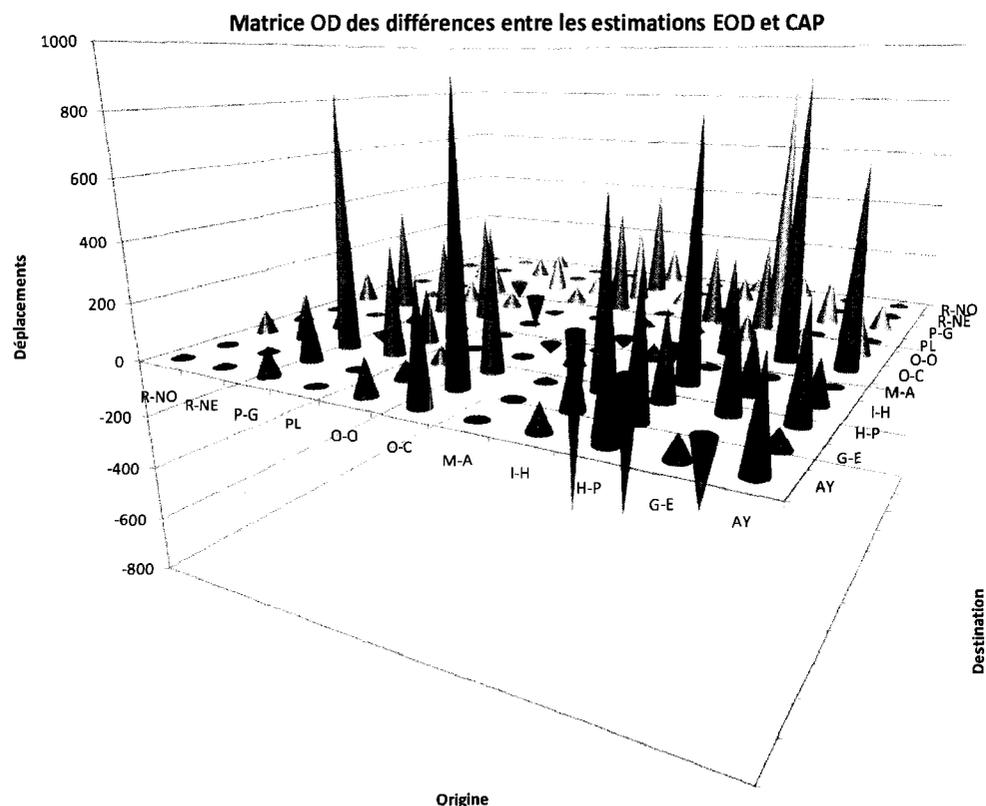


Figure 5.27: Différence entre l'estimation EOD et CAP des matrices OD

Bien que cette comparaison mette en lumière les écarts globaux entre les paires OD de l'enquête origine-destination et des cartes à puce, il est difficile de comparer les répartitions des déplacements dans chacune des matrices. Pour étudier cette répartition, il faut exprimer les paires OD de l'EOD et des CAP en pourcentage du nombre total de déplacements pour chacune des méthodes. La matrice du Tableau 5.16 montre ces ratios, les premiers étant ceux de l'EOD et les deuxièmes ceux des CAP (Ratio EOD\Ratio CAP). La proportion des déplacements internes (diagonale de la matrice en vert) est toujours plus élevée pour les données CAP que celles de l'EOD. En fait, 15 % de tous les déplacements de l'EOD sont étiquetés comme étant internes, comparativement à 25 % pour les données CAP. Le fait que les déplacements internes soient, pour toutes les régions, toujours sous-représentés par l'enquête origine-destination met en évidence un biais systématique entre les données CAP et EOD. Un

autre phénomène peut être observé grâce à cette matrice, soit la surreprésentation par l'EOD des paires OD peu achalandées (les ratios en gras). L'écart entre les ratios EOD et CAP est parfois très important ce qui laisse présager des facteurs d'expansion de l'EOD trop élevés. En fait, les régions d'Ottawa ouest, de Rural nord-est et de Rural nord-ouest sont moins sujettes à de nombreux déplacements, par conséquent, il est difficile de capturer les déplacements lors de l'enquête, ce qui peut introduire un biais. On sait aussi que le taux d'échantillonnage est directement relié à la qualité des estimations et, par conséquent, cela pourrait expliquer la différence entre les ratios CAP et EOD.

Toujours en étudiant la proportion des paires OD, on constate que la région d'Ottawa centre, qu'elle soit l'origine ou la destination du déplacement, est pratiquement toujours surreprésentée par les données CAP, sauf lorsque cette région est jumelée aux régions moins achalandées mentionnées précédemment (par exemple, Rural nord-est).

Tableau 5.16: Matrice des proportions par paires OD (EOD/CAP)

O	D										
	3,89	0,18	1,89	1,18	0,04	3,82	0,43	0,03	0,29	0,04	0,04
	\	\	\	\	\	\	\	\	\	\	\
	4,62	0,02	1,34	1,42	0,01	4,34	0,23	0,07	0,11	0	0,01
	0,15	0,2	1,3	1,03	0,04	4,17	0,14	0	1,01	0,05	0
	\	\	\	\	\	\	\	\	\	\	\
	0,02	1	0,48	0,84	0,03	3,2	0	0,01	0,76	0	0
	1,66	1,09	7,9	3,67	0,12	4,74	0,16	1,53	3,06	0,35	0,19
	\	\	\	\	\	\	\	\	\	\	\
	1,25	0,41	10,55	3,42	0,11	5,42	0,03	1,06	1,78	0,03	0,01
	1,21	0,95	3,66	0,21	0,37	0,88	0	0,95	0,98	0,19	0,07
	\	\	\	\	\	\	\	\	\	\	\
	1,28	0,74	2,82	1,9	0,35	1,22	0	0,79	1,55	0,03	0,01
	0,04	0	0,12	0,37	0,11	0,73	0,03	0	0,15	0	0
	\	\	\	\	\	\	\	\	\	\	\
	0,01	0,02	0,12	0,36	0,21	1,18	0	0,01	0,09	0	0
	4,24	3,98	4,61	0,9	0,73	0,11	0	1,91	4,6	0,84	0,26
	\	\	\	\	\	\	\	\	\	\	\
	4,09	2,97	5,66	1,59	1,19	2,58	0,01	1,57	5,79	0,05	0,04
	0,5	0,16	0,24	0,11	0,07	0	0	0,23	0,19	0,05	0
	\	\	\	\	\	\	\	\	\	\	\
	0,3	0	0,04	0,01	0	0,01	0,05	0,16	0,01	0	0
	0,11	0	1,37	1,17	0	1,91	0,23	0,28	0,35	0	0
	\	\	\	\	\	\	\	\	\	\	\
	0,12	0,01	1,04	0,81	0,01	1,58	0,14	0,84	0,15	0,01	0
	0,27	0,85	3,04	0,93	0,19	5,01	0,16	0,43	2,02	0,2	0,14
	\	\	\	\	\	\	\	\	\	\	\
	0,1	0,69	1,75	1,77	0,09	6,35	0,01	0,17	3,14	0,01	0,01
	0,04	0	0,35	0,22	0	0,91	0,05	0	0,17	0	0,03
	\	\	\	\	\	\	\	\	\	\	\
	0	0,01	0,04	0,04	0	0,06	0	0	0,02	0,01	0
	0,04	0	0,19	0,11	0	0,34	0	0	0,14	0,03	0
	\	\	\	\	\	\	\	\	\	\	\
	0,01	0	0,02	0,02	0	0,05	0	0	0,01	0	0,01
	12,2	7,4	24,7	10	1,7	22,6	1,2	5,4	13	1,8	0,7
Total	\	\	\	\	\	\	\	\	\	\	\
	11,8	5,9	23,9	12,2	2	26	0,5	4,7	13,4	0,1	0,1

Lorsque l'on regarde la distribution des déplacements selon les origines uniquement (dernière colonne) ou les destinations (dernière ligne), on remarque que, outre les régions qui ont très peu de déplacements, les proportions CAP et EOD sont sensiblement les mêmes. Ceci est d'autant plus étonnant que le nombre global de déplacements est de 47 373 pour l'EOD et de 32 689 pour les CAP. Autrement dit, la

répartition des déplacements sur l'ensemble du territoire est globalement la même pour l'EOD et CAP, sauf lorsque les déplacements sont internes ou qu'ils se font sur des territoires moins achalandés.

Cette analyse permet donc d'établir que les deux méthodes répartissent réciproquement les déplacements sur le territoire, mais qu'en termes de nombre de déplacements il y a toujours un fossé entre les deux estimations. Cependant, on peut croire que ce biais vient en partie de l'évaluation des facteurs d'expansion de l'enquête. En fait, il serait impensable que le système de collecte de données CAP n'ait pas capté environ 15 000 déplacements et, par conséquent, il est beaucoup plus probable que le biais vienne de l'enquête et principalement de la manière dont les facteurs d'expansion ont été quantifiés.

CHAPITRE 6

DISCUSSION RELATIVE À L'EOD ET AUX RÉSULTATS OBTENUS

Suite aux différents résultats obtenus lors de la comparaison des données provenant de l'EOD et des CAP, un questionnement sur l'écart entre les données s'impose. Les facteurs d'expansion de l'EOD peuvent, en partie, expliquer cette différence et c'est pourquoi la première partie de cette section leur sera consacrée.

La deuxième partie de cette section porte sur la fusion des données. En fait, la comparaison et l'analyse des données sont un premier pas vers la fusion des données et, par conséquent, une discussion sur les méthodes de fusion est abordée.

Finalement, la dernière discussion porte sur les améliorations, relatives aux techniques de cueillette d'information, possibles afin de faciliter l'échange d'informations entre les différentes sources.

6.1 Les facteurs d'expansion

Les facteurs d'expansion de l'enquête origine-destination ont été utilisés lors des calculs des analyses précédentes et on est à même de se questionner sur l'impact de ces facteurs concernant l'écart entre les estimations de l'EOD et des CAP.

Lors de l'analyse des matrices origine-destination de la section 5.3.2.3, on a été en mesure de constater que la distribution des déplacements selon les origines et les destinations est très similaire pour les deux méthodes de collecte d'informations. Cependant, l'écart entre le nombre de déplacements est grand et cela remet en cause la pertinence de la méthode d'évaluation des facteurs d'expansion. Bien qu'il existe une incertitude sur les données CAP reliée aux erreurs de collecte de données ainsi qu'aux algorithmes appliqués, il en demeure pas moins qu'un écart d'environ 15 000

déplacements doit vraisemblablement être dû à un biais dans l'estimation de l'achalandage issue des données de l'enquête origine-destination.

Mais la validation des facteurs d'expansion de l'EOD est ardue puisqu'il n'existe pas d'autres sources d'informations, autre que les bases de données des CAP, concernant les usagers de la STO. Il serait aussi possible de vérifier les calculs initiaux qui ont permis d'établir ces facteurs d'expansion, cependant, l'objectif de ce mémoire n'est pas d'étudier la méthodologie employée lors de l'EOD, mais bien de comparer les résultats de l'EOD et des CAP et ainsi essayer de mieux comprendre leurs similitudes et leurs différences. Par conséquent, ce qui est proposé dans cette section est une expérimentation afin d'étudier les écarts entre les facteurs d'expansion de l'EOD et les données CAP. Le but de cette étude est donc d'établir la faisabilité d'une telle démarche et de mieux comprendre les causes de l'écart entre les données

La normalisation des facteurs d'expansion est basée sur les résultats obtenus lors de la comparaison des matrices OD. Il a été mentionné que la proportion des déplacements par paires OD est sensiblement la même pour les données de l'EOD et des CAP. Cependant, deux cas particuliers ont été signalés, soit la sous-représentation des déplacements internes par l'EOD et la surévaluation des déplacements de l'EOD pour les régions à faible achalandage. Ces paires OD doivent nécessairement avoir des mesures de l'erreur différentes des autres paires OD. Il est donc nécessaire de déterminer les trois mesures d'erreur pour les trois cas, soit les déplacements internes, à faible achalandage et globaux (autres déplacements). Pour ce faire, on a tenté de minimiser la somme des erreurs aux carrés des déplacements selon les paires OD. En fait, les équations suivantes ont été utilisées :

- Erreur au carré des déplacements globaux :

$$\{(Déplacements_{EOD-Moyen} \cdot Mesure\ Erreur_{global}) - Déplacements_{CAP-Moyen}\}^2$$

- Erreur au carré des déplacements internes :

$$\{(Déplacements_{EOD-Moyen} \cdot Mesure Erreur_{interne}) - Déplacements_{CAP-Moyen}\}^2$$

- Erreur au carré des déplacements à faible achalandage :

$$\{(Déplacements_{EOD-Moyen} \cdot Mesure Erreur_{faible achalandage}) - Déplacements_{CAP-Moyen}\}^2$$

Les mesures de l'erreur ont été initialisées à 1 et, ensuite, le solveur Excel a été utilisé afin de minimiser la somme de toutes les erreurs aux carrés des paires OD. Les trois mesures de l'erreur obtenues sont :

- mesure de l'erreur globale : 0,6823,
- mesure de l'erreur interne : 0,9174,
- mesure de l'erreur faible achalandage : 0,1116.

On observe immédiatement que la mesure de l'erreur interne est beaucoup plus grande que la mesure de l'erreur des zones à faible achalandage. Ceci permet aussi de valider le choix des variables utilisées puisqu'il est évident que les estimations de ces paires OD sont sujettes à un biais de nature différente. On constate aussi que les trois mesures de l'erreur sont en bas de un. En fait, si les mesures avaient été supérieures à un, alors cela aurait signifié que les estimations des CAP auraient été plus élevées que celles de l'EOD. Étant donné que la situation observée précédemment révèle le contraire, il est donc tout à fait normal et souhaitable que les mesures de l'erreur soient inférieurs à un. Bref, à première vue ces mesures semblent cohérentes avec la situation analysée. Cependant, il faut davantage étudier le comportement de ces nouveaux facteurs d'expansion lors d'analyses plus désagrégées.

Lorsque l'on applique ces mesures de l'erreur aux facteurs d'expansion de l'EOD, on obtient une nouvelle matrice OD (Tableau 6.1) qui ressemble davantage à la matrice CAP (Tableau 5.15). En fait, les déplacements pour les paires OD internes et à faible achalandage sont maintenant beaucoup plus près de l'estimation des CAP et la

proportion de ces déplacements est aussi plus similaire à ce que l'on a trouvé originellement.

Tableau 6.1: Matrice OD des déplacements EOD normalisés

<i>O</i> \ <i>D</i>	AY	G-E	H-P	I-H	M-A	O-C	O-O	PL	G-C	R-NE	R-NO	Total
AY	1689	58	612	380	11	1236	23	9	93	2	2	
G-E	48	87	421	331	13	1347	8	0	327	3	0	
H-P	535	352	3437	1185	37	1534	9	495	989	19	10	
I-H	389	307	1184	91	119	282	0	308	314	10	4	
M-A	11	0	37	119	48	235	2	0	46	0	0	
O-C	1370	1286	1491	307	237	45	0	615	1486	45	14	
O-O	27	9	13	6	4	0	0	13	10	3	0	
PL	34	0	443	377	0	617	13	122	113	0	0	
G-C	86	273	981	299	60	1620	9	138	876	11	8	
R-NE	2	0	19	12	0	48	3	0	9	0	2	
R-NO	2	0	10	6	0	18	0	0	8	2	0	
Total	4193	2372	8648	3113	529	6982	67	1700	4271	95	40	32010

L'application des mesures de l'erreur donne des résultats très similaires aux CAP lorsque l'on regarde uniquement les paires OD. Cependant, on peut se questionner sur la validité et la pertinence de ces facteurs lors d'études plus précises comme celles réalisées précédemment. En fait, il serait possible de refaire toutes les analyses de la section 5 avec les nouveaux facteurs d'expansion afin de les valider. Mais, ces analyses sont longues et, par conséquent, il est préférable de se concentrer uniquement sur quelques analyses touchant à des secteurs d'étude différents, comme les heures, les paires OD et les lignes.

La Figure 6.1 montre les estimations EOD originales, les estimations EOD modifiées par les mesures de l'erreur et les estimations CAP. Globalement, on constate que la courbe des estimations de l'EOD corrigées colle beaucoup mieux aux estimations des CAP. Lors des analyses précédentes, il a été souligné que les estimations EOD et CAP des groupes d'heure 3, 4 et 7 étaient compatibles et cela est encore vrai avec les facteurs corrigés. De plus, les groupes 2, 5 et 6, qui originellement n'étaient pas inclus entre le maximum

et le minimum des estimations CAP, sont maintenant inclus dans ces limites. Seule l'estimation pour le groupe d'heure 1 est à l'extérieur du minimum et du maximum des données CAP, bien que l'écart soit minime. On remarque aussi que la pente des droites est moins abrupte et, par le fait même, beaucoup plus similaire à celle des données CAP. Autrement dit, la variation du nombre de déplacements entre les groupes d'heure fluctue maintenant beaucoup moins brusquement et ceci est plus représentatif de la situation observée par les données CAP.

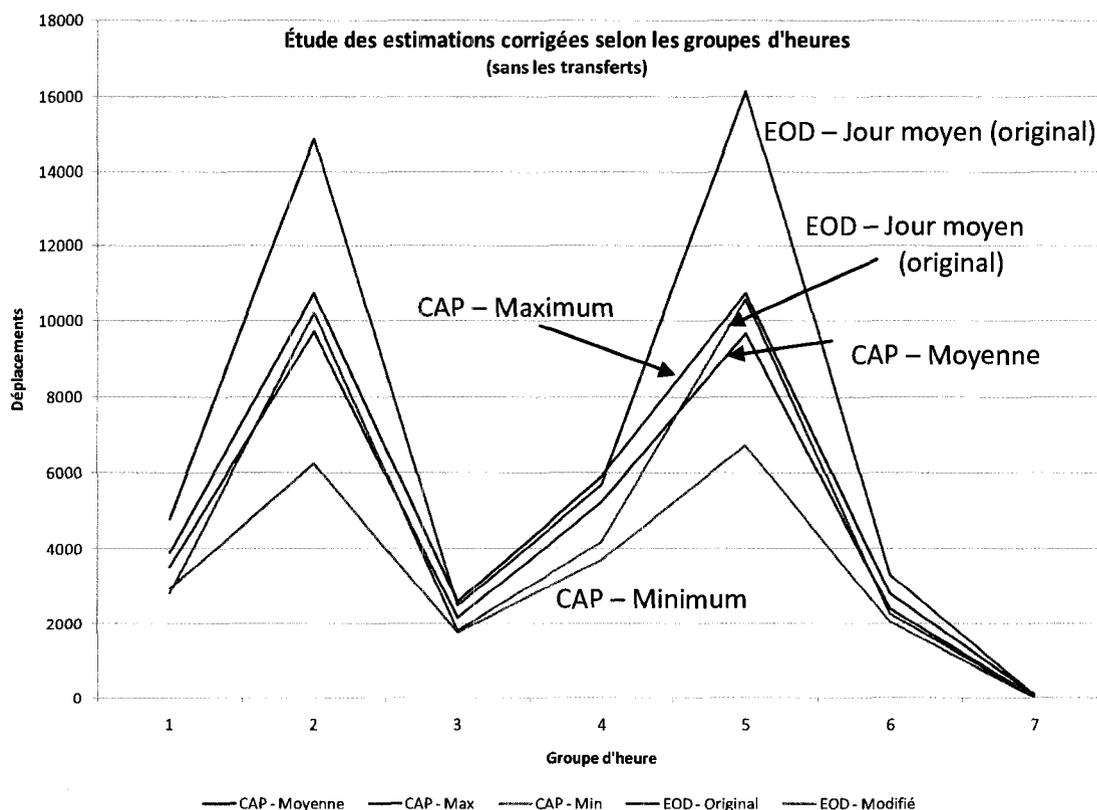


Figure 6.1: Analyse des mesures de l'erreur par groupe d'heure (CAP et EOD)

Cette analyse montre donc que les facteurs d'expansion corrigés se comportent très similairement aux estimations CAP, et ce, même s'ils sont analysés dans un contexte différent des matrices OD. Mais cette analyse est somme toute assez agrégée et il serait donc intéressant de voir comment ces nouveaux facteurs réagissent à une

analyse plus désagrégée. La Figure 6.2 montre donc une analyse par demi-heure des facteurs EOD originaux, EOD corrigés et CAP. On remarque que les mesures de l'erreur ont principalement eu un impact sur les périodes de pointe du matin et du soir. En fait, les sommets atteints lors des heures de pointe avec les facteurs d'expansion corrigés sont maintenant du même ordre de grandeur que les estimations CAP. On remarque aussi que les périodes hors pointes n'ont pas diminué énormément suite à l'application des mesures de l'erreur. En fait, cela est très encourageant puisqu'initialement les estimations de l'EOD et des CAP étaient très similaires et, avec l'application des mesures de l'erreur, on constate toujours cette compatibilité.

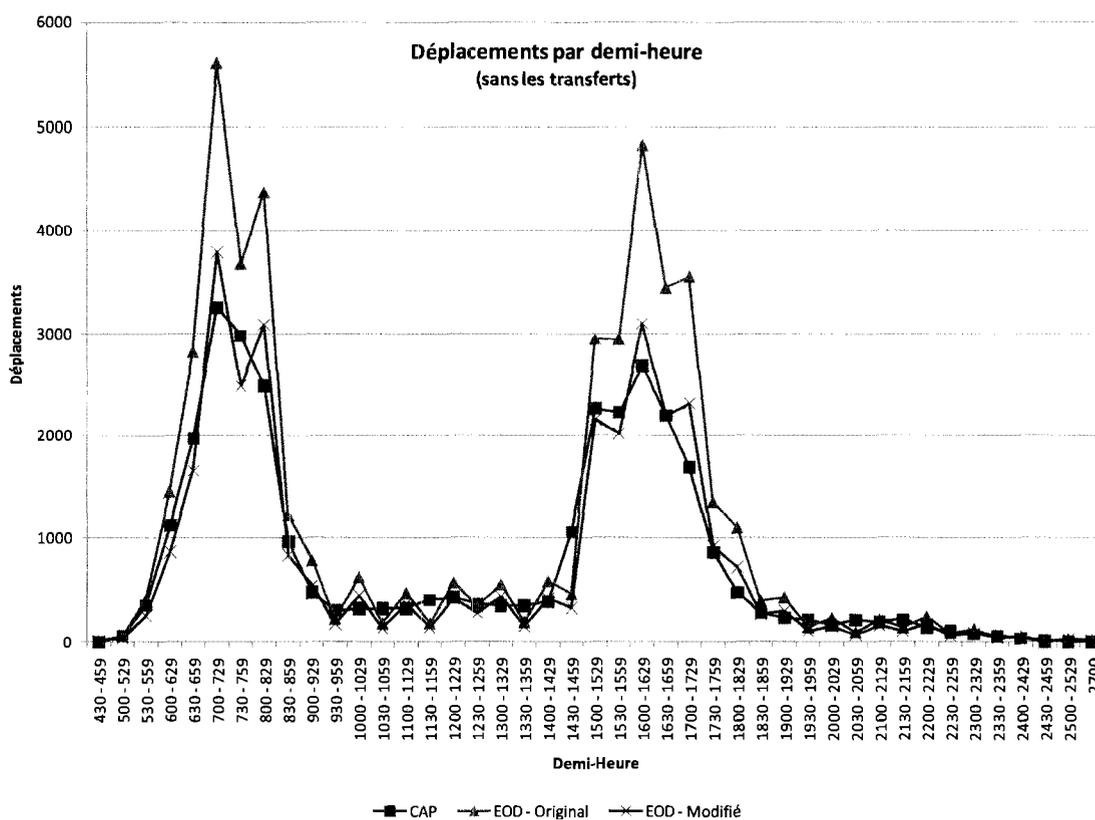


Figure 6.2: Analyse des mesures de l'erreur par demi-heure (CAP et EOD)

Cette analyse montre une fois de plus des facteurs corrigés beaucoup plus près des estimations des CAP. La dernière analyse est présentée à la Figure 6.3 et concerne les

montées par lignes. Sur cette figure, les carrés noirs représentent les estimations originales de l'EOD, les losanges mauves les estimations corrigées et les autres points de couleurs les données quotidiennes des CAP. On constate que les nouveaux facteurs d'expansion sont majoritairement plus près des données CAP, mais il existe encore quelques différences, dont certaines étaient présentes avec les facteurs originaux et d'autres sont nouvelles. Par exemple, la ligne 33, qui était surestimée, et la ligne 38, qui était sous-estimée, sont toujours distantes des estimations des CAP. Dans le cas de la ligne 33, l'estimation est plus près, mais elle est toujours surestimée. La ligne 38, par contre, est davantage sous-estimée par les facteurs corrigés que les facteurs originaux. De nouveaux écarts entre les données CAP et les données de l'EOD corrigées sont apparus, comme pour les lignes 37 et 67.

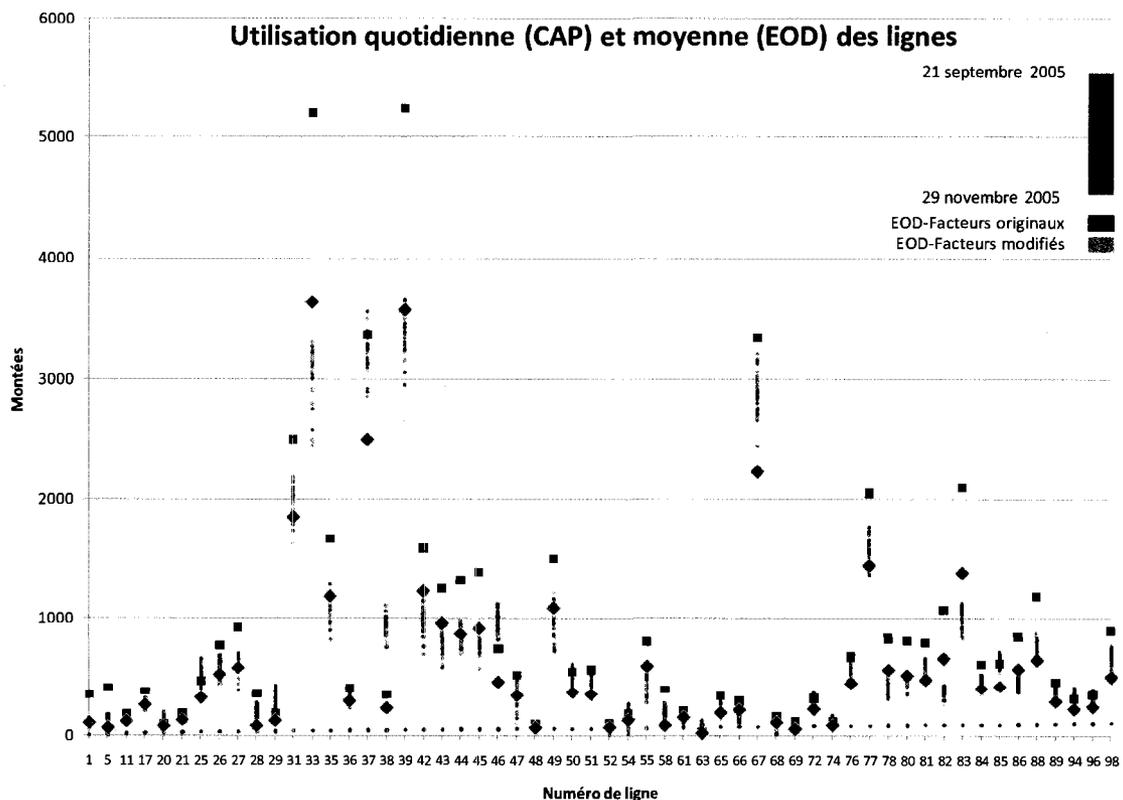


Figure 6.3: Analyse des mesures de l'erreur par lignes (CAP et EOD)

À la lumière de ces analyses, on peut constater que globalement les facteurs d'expansion corrigés sont beaucoup plus cohérents et représentatifs de la réalité illustrée par les données des CAP. Même lorsque l'on analyse ces nouvelles estimations de manière plus désagrégée, on constate que l'écart est moins important et que, surtout, elles s'harmonisent mieux à la réalité présentée par les données des CAP. Donc, en considérant que l'estimation des données CAP est, à quelques erreurs près, très représentative de la réalité, on peut affirmer que les estimations de l'EOD sont biaisées selon que le déplacement est interne ou fait dans une zone peu achalandée ou encore dans une zone non-interne et achalandée.

Naturellement, cette analyse s'applique uniquement dans le cadre des données sur les titulaires de cartes à puce. Il serait par contre pertinent d'analyser l'impact de ces nouveaux facteurs d'expansion sur les déplacements de l'EOD fait sur d'autres modes, comme l'automobile ou encore sur l'estimation du nombre de détenteurs de CAP.

Mais, cette démarche démontre, d'abord et avant tout, une partie du potentiel des données CAP. De plus, cela démontre que les données peuvent être utilisées conjointement afin d'aboutir à un résultat intéressant. C'est donc un pas de plus vers la fusion des sources de données afin d'obtenir une information plus riche et détaillée.

6.2 Le potentiel de fusion

La fusion des données provenant de sources distinctes est considérée par plusieurs acteurs du milieu des transports en commun comme étant une solution potentiellement efficace qui permettrait de répondre aux différentes difficultés actuelles d'échantillonnage et de qualité des estimations. Cependant, bien souvent les différentes sources de données sont indépendantes l'une de l'autre et il est très ardu de créer des liens afin de pouvoir jumeler l'information.

Cependant, la fusion des données n'implique pas uniquement une intégration totale et unilatérale des données. En fait, en analysant et en comparant les données de la STO, on constate que cette fusion peut se faire en partie ou en totalité selon les besoins informationnels des planificateurs en TC ou de tous autres utilisateurs des données. De plus, il est évident que les données peuvent s'intégrer l'une aux autres selon différentes méthodes. En prenant exemple sur la base de données de l'EOD et celle des CAP, on constate qu'il peut y avoir trois méthodes d'intégration de l'information. Ces trois méthodes sont :

1. l'intégration totale;
2. l'intégration partielle;
3. l'intégration analytique.

L'intégration totale est une substitution d'une partie des données d'une source par la totalité de l'information contenue dans une ou plusieurs autres sources. La Figure 6.4 met en contexte l'intégration totale en prenant l'exemple des sources d'informations analysées précédemment. Dans cet exemple, la source de données CAP contient des informations quotidiennes, telles que le nombre de déplacements, le titre de transport utilisé, l'heure du déplacement, qui peuvent être intégrées aux données de l'EOD. En fait, l'intégration des données CAP viendrait remplacer les estimations des déplacements ou, plus généralement, la dimension déplacements dans l'EOD. Cependant, cette intégration conserverait les dimensions personne et ménage déjà présentes dans l'enquête origine-destination.

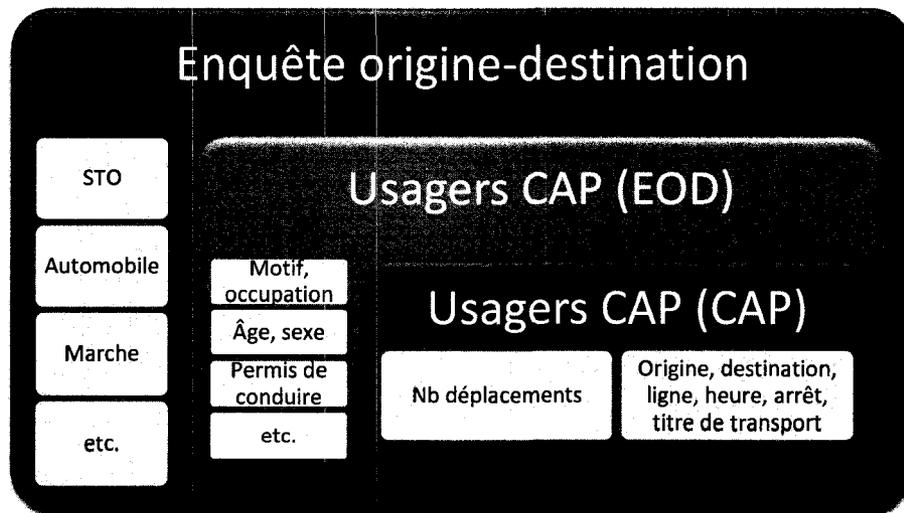


Figure 6.4: Intégration totale

Il est certain que cette méthode permettrait un regroupement de l'information et une précision des données unique. Mais l'aboutissement à ce résultat demande une étude approfondie des sources de données afin d'être en mesure d'affecter les différentes variables d'une source à celles d'une autre et ainsi aboutir à la fusion des données. Plus il y a de variables en jeu, plus la complexité de la méthode augmente. Il est clair que cette méthode doit être appliquée en plusieurs étapes, dont la première est l'analyse des variables. Il faut donc être en mesure d'identifier les variables conjointes aux deux bases de données et les variables uniques. Les variables conjointes permettent donc de lier les deux bases de données entre elles tandis que les variables uniques apportent un élément nouveau et recherché. Dans le cas des bases de données de l'EOD et des CAP, les variables conjointes seraient entre autres :

- le titre de transport;
- la ligne utilisée;
- la nature du déplacement (correspondance ou première montée);
- l'origine et la destination.

Ces variables, présentes à la fois dans les différentes bases de données, permettent donc d'établir des relations qui constitueront la base même de la fusion des données. Ces relations peuvent être quantifiables, par exemple grâce à des calculs de distances.

Les variables uniques ne jouent pas un rôle majeur dans l'établissement des liaisons entre les sources de données, mais elles en sont souvent à l'origine puisque ce sont principalement ces données que l'on cherche à intégrer à une autre source. En fait, les variables uniques viennent compléter, raffiner et améliorer l'information présente afin d'obtenir un portrait encore plus précis de la population à l'étude. Ces variables, dans le cas étudié précédemment, sont principalement :

- enquête origine-destination
 - le motif du déplacement, l'occupation, l'âge, le sexe, la possession de permis de conduire, etc.
- cartes à puce
 - le numéro d'arrêt de la montée et de la descente (obtenu par des algorithmes), l'heure des déplacements en continue, etc.

Lorsque toutes les variables sont définies et analysées, la deuxième étape vers l'intégration totale des données est naturellement la création de relations ou de ponts entre les différentes sources. Ces relations sont le cœur même de la fusion puisqu'elles sont basées sur des hypothèses et des calculs qui dicteront la précision et la compatibilité des données. C'est donc à cette étape qu'il faut particulièrement être vigilant puisqu'un biais, dû entre autres aux estimations et à l'incompatibilité de certaines variables, peut apparaître. Il faut donc choisir judicieusement la technique ou le modèle mathématique permettant d'établir une corrélation entre les différentes variables conjointes.

Lorsque ces relations sont établies, il faut distribuer ou répartir les objets d'études, tels que les déplacements, selon le nouveau modèle relationnel. Cette répartition doit se

faire en respectant les hypothèses posées précédemment. C'est aussi à ce moment qu'il faut s'assurer de corriger toutes anomalies présentes.

L'intégration partielle, contrairement à la méthode précédente, n'implique qu'une partie d'une ou plusieurs sources d'informations. En fait, cette intégration peut venir combler un besoin spécifique en information. Par exemple, ce type d'intégration pourrait être utilisé pour compléter des données provenant de faibles taux d'échantillonnage ou encore afin de corriger des anomalies. En fait, il a été mentionné précédemment qu'il existe un écart considérable entre les estimations CAP et de l'EOD pour les étudiants réguliers. Si, dans les faits, les cartes à puce s'avèrent être une estimation plus réaliste, alors il serait envisageable de modifier les données liées à ce groupe d'individus dans la base de données de l'EOD (Figure 6.5).

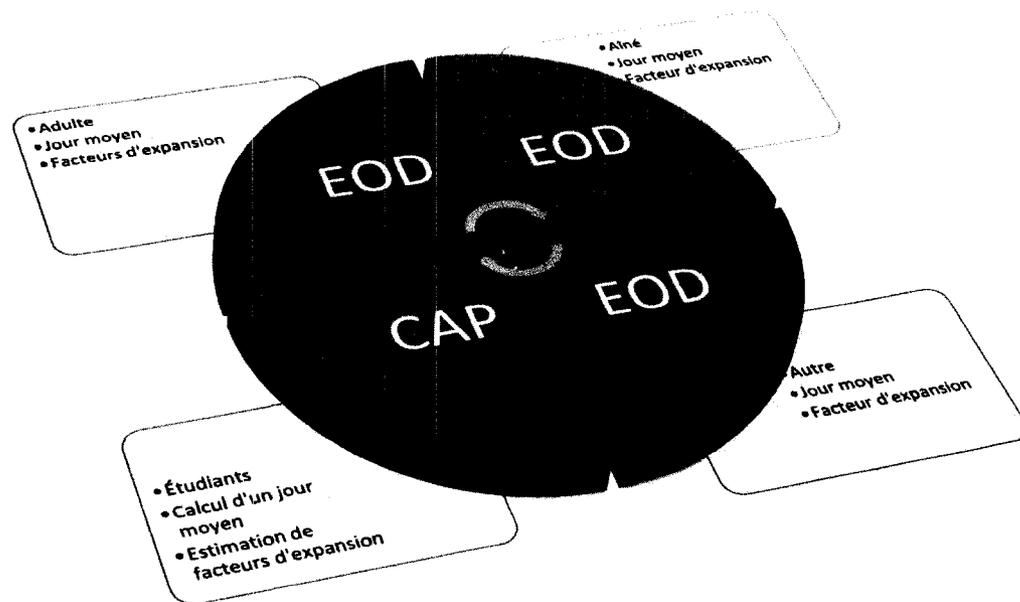


Figure 6.5: Exemple d'intégration partielle

En fait, le principe de ce type de fusion est d'incorporer uniquement les données souhaitées tout en gardant l'intégrité de la source qui va les accueillir. Donc, comme

pour l'intégration totale, la première phase consiste en une analyse des différentes variables. Cependant, la deuxième phase n'est pas une recherche de liens entre les sources d'informations, mais bien en un remodelage des données afin de se conformer à la source de donnée d'accueil. Par exemple, au lieu d'intégrer des données quotidiennes à une base de données contenant des jours moyens, cette intégration transformerait les données quotidiennes en journées moyennes afin de les fusionner.

Naturellement, ce type d'intégration n'est pas toujours applicable et il faut tout de même s'assurer de la compatibilité du remodelage des données. Mais, cette solution a comme avantage d'être moins complexe que l'intégration totale.

L'intégration analytique peut se définir comme la fusion de données afin de faire une analyse d'un contexte particulier qui ne peut être déduit d'une seule source. Autrement dit, il est question d'extraction d'informations spécifiques provenant de plusieurs sources d'informations afin de créer une nouvelle source totalement indépendante et dont l'utilité est limitée à un contexte défini.



Figure 6.6: Exemple d'intégration analytique

Par exemple, si l'on veut étudier, à partir des bases de données de l'EOD et des CAP, les déplacements quotidiens selon l'âge des personnes (Figure 6.6), il faut nécessairement avoir accès à l'âge, présent dans l'EOD, et aux déplacements journaliers, contenus dans les données CAP. La mise en commun de ces informations permet la création d'une nouvelle source de données qui servira aux analyses subséquentes. Donc, selon

l'analyse souhaitée, des hypothèses et des calculs différents seront utilisés pour parvenir à jumeler les données. Dans l'exemple ci-dessus, on pourrait calculer, avec les données de l'EOD, la proportion de personnes pour chaque âge selon le titre de transport et poser comme hypothèse que cette proportion est valide pour les données CAP. Il suffirait donc d'affecter à ces proportions aux données CAP et ainsi créer une nouvelle source de données qui permettrait d'obtenir des informations relatives à l'âge et aux déplacements quotidiens.

Cependant, il est important de mentionner que ce type d'intégration doit être limité au contexte et aux hypothèses définis dès l'extraction des données. En fait, toujours avec le même exemple, cette nouvelle source de données ne pourrait tout simplement pas être utilisée pour analyser les déplacements quotidiens sur une ligne en fonction de l'âge des usagers sans, au préalable, faire de nouveaux calculs et définir de nouvelles hypothèses. Il faut donc être très prudent avec l'analyse de ces données puisqu'un biais supplémentaire peut se créer selon le contexte et le niveau de précision des analyses.

La fusion des données est donc sans contredit un sujet qui se doit d'être expérimenté davantage afin de mieux en connaître la portée. Cependant, il est certain que cette fusion peut s'effectuer différemment selon les besoins informationnels en cause. On a vu trois méthodes d'intégration des données possibles qui permettent des analyses complètes ou spécifiques. L'utilisation de ces méthodes passe par quatre étapes, soit l'étude des variables présentes dans les bases, la création de liens entre les sources, l'affectation des objets d'étude et la correction d'anomalie. De plus, tout au long du processus de fusion, des hypothèses sont utilisées afin de rendre compatibles les différentes données. Ces hypothèses sont fondamentales pour le succès de la fusion puisqu'elles sont directement reliées à la qualité et la précision des données fusionnées. On peut d'ailleurs se questionner sur les mesures qui pourraient aider à diminuer l'impact de ces hypothèses sur la qualité de données. La prochaine section soulève donc quelques améliorations intéressantes à apporter aux différentes sources

d'informations afin d'établir des relations plus fiables entre les différentes sources et, par le fait même, d'augmenter la confiance envers les sources fusionnées.

6.3 Les améliorations

Suite aux difficultés rencontrées afin de comparer les données de l'EOD et des CAP, il serait intéressant de faciliter ces comparaisons en modifiant les techniques de collecte d'informations ou en recherchant d'autres informations. Entre autres, les données provenant des cartes à puce bénéficieraient d'informations supplémentaires sur l'usager. Bien que ce sujet soit délicat à cause des lois sur la protection de la vie privée, il serait néanmoins pertinent d'avoir accès à ces données afin d'être en mesure de mieux comparer les objets personne et ménage des différentes sources. Non seulement ces comparaisons permettraient une meilleure compréhension des données, mais elles seraient particulièrement utiles pour valider les hypothèses lors de la fusion des données. De plus, certaines sociétés de transport en commun collectent déjà des informations personnelles sur leurs usagers, entre autres la *Chicago Transit Authority* (Utsunmiya, Attanucci, & Wilson, 2006). En fait, ces données sont enregistrées lors d'une inscription et elles permettent de déterminer, entre autres, l'âge et le sexe du titulaire. Bref, l'ajout de ces données enrichirait grandement les bases de données des CAP, et ce, à peu de coûts supplémentaires.

Un autre ajout, plus dispendieux, serait la validation du titre de transport lors de la descente. Bien que les algorithmes d'estimation de la destination fait par Trépanier, Chapleau, & Tranchant (2007), qui ont été utilisés dans ce mémoire, soient performants, il n'en demeure pas moins que la validation lors de la descente permettrait d'obtenir des destinations plus précises. Aussi, ce type d'informations peut être pertinent pour analyser le temps des trajets ou encore les arrêts les plus fréquents. Il est cependant évident que cet ajout demanderait du matériel supplémentaire et exigerait de modifier le processus de validation, mais les données

recueillies seraient plus complètes et des analyses temporelles et géographiques encore plus fines pourraient utiliser cette information à bon escient.

Ces informations supplémentaires permettraient donc d'enrichir les données CAP et ouvriraient la porte à de nouvelles analyses. Mais ce ne sont pas seulement les données CAP qui pourraient bénéficier d'améliorations. En fait, le questionnaire des enquêtes origine-destination pourrait comporter une section spécifique aux cartes à puce afin d'établir les fréquences d'utilisation de la carte et le lieu géographique de la correspondance, si applicable. Cela permettrait de tracer un portrait plus fidèle des habitudes des usagers et, surtout, de faire des liens plus facilement entre les bases de données de l'EOD et des CAP. Il est par contre évident que ce type de modifications engendrerait des coûts importants et, avant d'aller de l'avant avec ces mesures, il faut des analyses et des expérimentations plus approfondies.

CHAPITRE 7

CONCLUSION

L'objectif de ce mémoire était d'établir une méthodologie de comparaison de données provenant d'une enquête ménage origine-destination et d'un système de paiement par cartes à puce. Le sujet a porté sur les forces, les faiblesses et les limites liées à l'utilisation de ces données.

Il a été démontré que les enquêtes origine-destination tracent un vaste portrait des caractéristiques sociodémographiques de la population et de leurs habitudes de déplacements. Cependant, réussir à obtenir des grands taux d'échantillonnage est de plus en plus ardu entre autres à cause de nouvelles technologies, telles que les téléphones cellulaires, et de l'apparition de liste *Do not call*. Parallèlement à ces enquêtes, plusieurs sociétés de transport à travers le monde possèdent ou implantent des systèmes de paiement par cartes à puce. Ces systèmes ont la particularité de recueillir des quantités impressionnantes de données en continu et leur précision ainsi que leur détail sont remarquables. Les lois sur la protection de la vie privée ainsi que des questions d'éthiques limitent la collecte d'informations personnelles sur les usagers.

La fusion des données s'impose donc comme une solution ayant le potentiel de satisfaire les besoins en information des planificateurs. Des techniques de fusion doivent d'ailleurs être plus documentées et expérimentées pour valider la pertinence, la précision et la qualité des données fusionnées.

La fusion, pour être sensée, doit se baser sur des données ayant des caractéristiques temporelles, géographiques et sociodémographiques similaires. La STO dispose justement de sources d'informations correspondant à ces exigences. En fait, en 2005 a eu lieu une grande enquête ménage origine-destination de la région de la capital

nationale (Ottawa et Gatineau) fait par le comité TRANS. Pour la même période, des données ont été archivées sur les transactions effectuées à l'aide de cartes à puce sur le réseau de la Société de transport de l'Outaouais. Ces deux bases de données contiennent donc des informations sur la même population, la même région géographique et elles ont été collectées sur la même période de temps, soit du 21 septembre 2005 au 29 novembre 2005.

Avant de comparer ces données, une explication du contexte particulier de la STO a été exposée. Un historique des grands événements ayant marqué l'histoire de cette société, constamment à la recherche d'innovation, permet de comprendre les nombreuses étapes menant à l'amélioration du service et à l'implantation de technologies de pointe, comme les systèmes de cartes à puce. La présentation de certaines caractéristiques démographiques et des principales lignes de désirs permet une meilleure compréhension des habitudes de déplacement de la population. C'est donc en connaissant la technologie utilisée, la population étudiée et les services offerts qu'il est possible d'aller de l'avant avec des analyses plus fines et complexes

Une démarche d'exportation des données a ensuite été élaborée pour unifier et comparer les informations contenues dans les bases de données de l'EOD et des systèmes par cartes à puce. Cette démarche propose un premier formatage des données directement dans la base de données. Ce formatage utilise l'approche orientée objet afin de cibler l'information pertinente et de la traiter. Un deuxième formatage des données a été nécessaire afin d'utiliser une nomenclature unique pour les deux bases de données. De longs processus de validation des données et l'élaboration d'une terminologie commune aux deux bases de données ont été nécessaires pour aboutir à des requêtes comparables en termes d'informations et de notations. Ensuite les requêtes des deux bases de données ont été unies et extraites vers un fichier de comparaison. Cet outil, fait sous Excel 2007, a permis de conserver l'intégrité des données en plus de faciliter grandement les calculs. De plus, l'interface

graphique créée permet de visualiser et de comparer les données en quelques clics de souris. C'est grâce à ce format très uniforme et standardisé des données que les analyses et les comparaisons plus approfondies ont pu être réalisées.

Une première comparaison, à l'échelle macroscopique, a montré que de grandes différences existent entre les indicateurs de l'EOD et des CAP qui ont été calculés. Ce sont d'ailleurs ces indicateurs qui ont tracé la ligne directrice des analyses subséquentes.

Deux catégories d'analyses ont été faites sur les données, soit les analyses segmentées et les analyses du réseau. Les analyses segmentées servent à étudier les déplacements selon les journées, les heures des déplacements et les titres de transport utilisés. Elles ont révélé des habitudes comportementales de la population, telles que des baisses d'achalandages les lundis, les vendredis ainsi que lors des journées avant et après des congés fériés. Ces analyses ont aussi démontré que l'écart entre les données de l'EOD et des CAP varie en fonction de l'heure de la journée, du type de journée et des usagers qui se déplacent. Bien que ces analyses ont permis une meilleure compréhension des données, elles n'ont pas été en mesure d'identifier un biais systématique pouvant expliquer l'important écart entre les données de l'EOD et des cartes à puce.

Une première analyse du réseau a été effectuée sur les trajets d'autobus. Cette analyse a permis de comparer les données de l'EOD et des systèmes de CAP pour toutes les lignes du réseau de la STO ainsi que d'étudier la variabilité quotidienne des déplacements enregistrés par les systèmes de CAP. On a constaté lors de l'analyse des données des systèmes de CAP que certaines lignes étaient soumises à des coefficients de variation plus élevés tandis que d'autres lignes avaient des usagers beaucoup plus réguliers. Cependant, aucune corrélation entre la variation des déplacements sur une ligne et l'écart entre les données de l'EOD et des systèmes de CAP n'a été décelée.

On a donc procédé à une analyse plus fine des coefficients de variation sur deux lignes où les usagers ont des comportements différents. Cette analyse a démontré que la fréquence d'utilisation diffère d'une ligne à l'autre et même pour des titres de transport différents.

La dernière analyse porte sur les déplacements selon les paires origine-destination. Une méthodologie est détaillée afin de construire une matrice OD à partir des deux bases de données. Lors de la comparaison des matrices OD, on a été en mesure de clairement identifier trois groupes de déplacements ayant un biais systématique. Ces groupes sont les déplacements internes (sous-représentation), les déplacements faits dans des régions à faible achalandage (surreprésentation) et les déplacements globaux (tous les autres déplacements). Une tentative de d'ajustement des facteurs d'expansion a été réalisée et les résultats obtenus sont très concluants. Ces ajustements ont permis d'identifier l'écart entre certains facteurs d'expansion utilisés dans les données de l'EOD et l'estimation des données CAP.

Ce mémoire se veut donc un premier pas vers la fusion des données et il contribue à améliorer notre compréhension de l'information provenant des systèmes de cartes à puce et des enquêtes origine-destination. Il exprime aussi une méthodologie et un savoir-faire québécois afin d'intégrer différentes sources d'informations et ainsi obtenir des données plus riches et facilement utilisables pour établir une offre de services juste et adaptée aux besoins de la population.

Il reste encore beaucoup de travail afin de s'approprier complètement ces données. C'est entre autres le cas avec le traitement des codes d'erreur présents dans les bases de données. En fait, des méthodes de répartition des codes d'erreurs devront être étudiées et validées afin d'obtenir des données cohérentes. Par exemple, l'analyse des déplacements sur les trajets d'autobus, provenant des systèmes de CAP, a montré qu'environ mille déplacements par journée ont lieu sur la ligne 900, soit une ligne

représentant un code d'erreur. Ces déplacements appartiennent à une ou plusieurs lignes et ils peuvent faire une différence importante lors des comparaisons avec les données de l'EOD. Une approche possible aurait été de faire correspondre l'arrêt d'origine et l'heure du déplacement à une ligne et ainsi répartir les codes d'erreurs. Cependant, pour une très forte majorité de ces déplacements l'arrêt est inconnu et il est donc impossible de procéder de cette manière.

Une autre possibilité serait d'analyser la variabilité quotidienne sur chacune des lignes et de détecter des baisses d'achalandage à des heures précises. Si ces baisses d'achalandage correspondent aux déplacements sur la ligne 900 et qu'aucun facteur ou habitude de déplacement, tel que les journées fériées, n'intervient, alors ils peuvent être associés à cette ligne. Plusieurs questions demeurent nébuleuses, telle que dans quelle mesure est-il possible de réaliser ces répartitions d'erreurs ou quel est l'indice de confiance de ces méthodes. Il serait donc souhaitable que des approches soient élaborées et testées afin de minimiser l'impact des codes d'erreurs sur la précision des données et ainsi connaître davantage les limites et les incertitudes associées à la répartition des codes d'erreurs.

De plus, il est nécessaire d'expérimenter davantage la fusion de données dans le milieu du transport. En fait, les bénéfices d'une telle approche sont nombreux et il faut investir davantage de temps et d'argent afin d'aboutir à des résultats concrets et applicables. Ce mémoire a exposé certains types de fusion et il serait intéressant de les détailler davantage afin de connaître leurs forces, leurs faiblesses ainsi que leurs limites.

Finalement, ce mémoire contribue à la résolution d'une problématique (Trépanier, Morency, & Blanchette, Les systèmes de paiement par cartes à puces: un complément aux enquêtes origine-destination?, 2008) (Trépanier, Morency, Blanchette, Enhancing household travel surveys using smart card data, 2008) multidisciplinaire qui est de plus

en plus préoccupante dans le milieu des transports en commun, soit la collecte d'informations de qualité sur les habitudes de déplacements de la population. Il est certain que la fusion des données s'imposera comme une solution viable et deviendra sous peu essentielle aux planificateurs en transport urbain et, en ce sens, ce mémoire contribue à ses premiers pas.

RÉFÉRENCES

Allison, C. Y., Iseki, H., Taylor, B. D., & King, D. A. (2006). Interoperable transit smart card systems: Are we moving too slowly or too quickly? *Transportation Research Board, 1986* , 69-77.

Axhausen, K. W., Zimmerman, A., Schönfelder, S., Rindsfuser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation, 29* , 95–124.

Bagchi, M., & White, P. (2005). The potential of public transport smart card data. *Transport Policy, 12* , 464-474.

Balaban, D. (2007, Novembre 1). The industry's new look. *Card Technology* .

Baladan, D. (2002, Janvier 1). Wave of the future? *Card technology* .

Bayart, C., Bonnel, P., & Morency, C. (2008). SURVEY MODE INTEGRATION AND DATA FUSION: METHODS AND CHALLENGES. *8th INTERNATIONAL CONFERENCE ON SURVEY METHODS IN TRANSPORT*. Annecy, France: ISCTSC.

Bergeron, D., & Grondines, J. (1998). Une nouvelle enquête origine-destination pour la grande région de Montréal en 1998: pourquoi et comment? *AQTR(Présentation)* , , 1-31, www.amt.qc.ca/docs/expose/AQTR-980421OD.ppt.

Blobel, B., Pharow, P., Spiegel, V., Engel, K., & Engelbrecht, R. (2001). Securing interoperability between chip card based medical information systems and health networks. *International Journal of Medical Informatics, 64 (2)* , 401-415.

Blythe, P. (2004). Improving public transportation ticketing through smart cards. *Proceedings of the institution of Civil Engineers, Municipal Engineer, 157 (ME 1)*, 47-54.

Blythe, P. (1998). Integrated Ticketing - Smart Cards in Transport. *IEEE seminar, Using ITS in public transport and in emergency services* (p. 20). Londres: IEEE.

Bonnel, P., & Le Nir, M. (1998). The quality of survey data: Telephone versus face-to-face interviews. *Transportation*, 25 , 147-167.

Bryan, H., & Blythe, P. (2007). Understanding behaviour through smartcard data analysis. *Proceedings of the Institution of Civil Engineers*, 160 (TR4) , 173-177.

Business Wire. (1999, Avril 22). *Bibliography for "Singapore Transit Authority Awards Cubic Transportation Systems \$23.6 Million Contract for Smart Card Ticketing Machines"*. Consulté le Septembre 30, 2008, sur Business Wire: http://findarticles.com/p/articles/mi_m0EIN/is_1999_April_22?pnun=2&opg=54448138&tag=artBody;col1

Card Technology Today. (2002, Juin). The role of smart cards in mass transit systems. *Card Technology Today*, 14(6) , p. 10.

Chapleau, R., Allard, B., Trépanier, M., & Morency, C. (2001, Janvier-Mars). Les logiciels d'enquête transport comme instruments incontournables de la planification analytique. *Recherche transport sécurité* , pp. 59-77.

Chapleau, R., Trépanier, M., & Allard, B. (1998). Practical implementations of object-oriented GIS-T. *Proceedings of the World Conference on Transportation Research*. Anvers, Belgique: World Conference on Transport Research Society.

Cheung, F. (2006). Implementation of nationwide public transport smart card in the netherlands: Cost-benefit analysis. *Transportation Research Record*, 1971 , 127-132.

Clarke, R. (2001). Person location and person tracking: Technologies, risks and policy implications. *Information Technology & People*, 14(2) , 206-231.

de Leeuw, E. D. (1992). *Data quality in mail, telephone and face to face surveys*. Amsterdam: T.T.Publikaties.

de Leeuw, E., & Van der Zouwen, J. (1992). Data quality and mode of data collection: methodology and explanatory model. *Dunod (ed) La qualité de l'information dans les enquêtes*, 11–31.

Electronic product news. (2005, Octobre 1). *The Hapless Tale of Geoffrey Dummer*. Consulté le Septembre 14, 2008, sur Electronic product news: <http://www.epn-online.com/page/22909/the-hapless-tale-of-geoffrey-dummer.html>

Guillou, L. (2004). Histoire de la carte à puce du point de vue d'un cryptologue. *Colloques sur l'histoire de l'informatique, des réseaux et des télécommunications* (pp. 126-154). Rennes-Cesson: ACONIT.

Iseki, H., Allison, C. Y., & Taylor, B. D. (2007). Are smart cards the smart way to go? *Transportation Research Board, 1992*, 45-53.

iTRANS Consulting inc. (2006). *Enquête origine-destination 2005: sommaire des résultats*. Ottawa.

McDonald, N. (2000). Multipurpose smart cards in transportation: Benefits and barriers to use. *UCTC Research Papers: 27*, 27 p.

Morency, C., & Chapleau, R. (2002). Implications of settlement patterns on travel behaviours: A totally disaggregate empirical study in the greater Montreal area. *30e conférence annuelle de la Société canadienne de génie civil*, 227-236.

Morency, C., & Chapleau, R. (2002). Implications of Settlement Patterns on Travel Behaviours: A totally disaggregate empirical study in the greater Montreal area. *30e conférence annuelle de la Société canadienne de génie civil*, 263-272.

Morency, C., Trépanier, M., & Agard, B. (2007). Measuring transit use variability with smart-card data. *Transport policy*, 14, 193-203.

Multisystems Inc.; Mundle & Associates Inc.; Simon & Simon Research and Associates Inc. (2003). *Fare policies, structures and technologies: update*. Washington D.C.: Transportation Research Board of National Academies.

Portalis, C. (2007, juin 1). *Fonctionne une carte à puce ?* Consulté le Septembre 1, 2008, sur L'Internaute: <http://www.linternaute.com/science/technologie/comment/07/carte-puce/carte-puce.shtml>

Secrétariat aux enquêtes origine-destination métropolitaines. (2003). *Enquête origine-destination 2003: La mobilité des personnes dans la région de Montréal*. Montréal.

Shelfer, K. M., & Procaccino, J. D. (2002, Juillet 1). Smart card evolution. *Communication of the ACM*, 45 (7) , pp. 83-88.

Shiibashi, A., Yajima, T., Xiaodong, L., & Mori, K. (2006, Juin 16-18). The evaluation of high reliability in an autonomous decentralized IC card ticket system. *Computer Networks, 2006 International Symposium on* , 209-213.

Statistique Canada. (2008, Octobre 14). *Recensement du Canada*. Consulté le Août 18, 2008, sur Statistique Canada: <http://www12.statcan.ca/census-recensement/2006/rt-td/index-fra.cfm>

Stopher, P. R. (1998). A review of separate and joint strategies for the use of data on revealed and stated choices. *Transportation*, 25 , 187–205.

Stopher, P. R., & Greaves, S. P. (2007). Household travel surveys: Where are we going? *Transportation Research Part A* 41 , 367-381.

Transport Québec. (2007). *Méthodologie*. Consulté le 10 02, 2008, sur Transport Québec:

http://www.mtq.gouv.qc.ca/portal/page/portal/ministere/ministere/recherche_innovation/modelisation_systemes_transport/enquetes_origine_destination/methodologie

Trépanier, M., & Chapleau, R. (2001a). Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination. *Revue canadienne de génie civil*, 28(1), 48-58.

Trépanier, M., & Chapleau, R. (2001b). Linking transit operational data to road network with a transportation object-oriented GIS. *URISA*, 13(2), 23-30.

Trépanier, M., Chapleau, R., & Morency, C. (2006, Septembre 1). *Tools and methods for a transportation household survey*. Consulté le 10 04, 2008, sur Urban and regional information systems association: http://www.urisa.org/publications/journal/articles/tools_and_methods_for_transportation

Trépanier, M., Chapleau, R., & Tranchant, N. (2007). Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System. *Journal of Intelligent Transportation Systems*, 11(1), 1-14.

Trépanier, M., Morency, C., & Blanchette, C. (2008). Les systèmes de paiement par cartes à puces: un complément aux enquêtes origine-destination? *43e congrès de l'Association québécoise du transport et des routes*. Québec: Association québécoise du transport et des routes.

Turban, E., & Brahm, J. (2000). Smart card-based electronic card payment systems in the transportation industry. *Journal of Organizational Computing and Electronic Commerce*, 10(4), 281-293.

Utsunmiya, M., Attanucci, J., & Wilson, N. (2006). Potential uses of transit smart card registration and transaction data to improve transit planning. *Transportation research record, 1971* , 119-126.

Venigalla, M. (2004). Household travel survey data fusion issues. *National Household Travel Survey Conference: Understanding Our Nation's Travel* (pp. 1-11). Washington D.C.: Transportation research board.

ANNEXE A :

LES ÉQUIVALENCES DES NOMENCLATURES DE L'EOD ET DES SYSTÈMES DE CAP

La nomenclature utilisée dans les bases de données de l'EOD et des systèmes de CAP est différente et elle nécessite des ajustements pour établir une compatibilité des terminologies. Cette section a pour but de décrire la méthodologie employée et les hypothèses utilisées pour aboutir à cette compatibilité de l'information.

A.1 : Les titres de transport

La base de données provenant des systèmes de cartes à puce contient une table, « sivt_titre », qui contient la signification textuelle de la clé associée à chacun des titres de transport utilisés lors des déplacements et enregistrés dans la table « sivt_transactionbus ». Cependant, la description des titres de transport présents dans cette table ne correspond pas avec les titres de transport de l'EOD. Une équivalence entre ces différents titres est nécessaire puisque la comparaison par titre de transport serait impossible. Cette équivalence est montrée au tableau A.1 ci-dessous.

La base de données contient des déplacements faits par des usagers détenant un titre de transport pour les étudiants émis en 2004 ou en 2003. Bien que ces cartes semblent désuètes, les déplacements associés à ces titres ont été faits sur la période du 21 septembre 2005 au 29 novembre 2005 et ont été validés par le système de cartes à puce. Par conséquent, ces titres de transport sont traités comme des titres valides et une équivalence a été adoptée pour eux.

Deux titres de transport présents dans la base de données des systèmes de cartes à puce, « archive » et « employé STO », ont nécessité la création d'un code spécifique « autre ». Il est cependant impossible de confirmer que ce code « autre » est réellement équivalent à celui présent dans la base de données de l'EOD.

Tableau A.1: Équivalence des titres de transport (CAP et EOD)

Numéro du titre de transport	Cartes à puce		CAP et EOD
		Description	Équivalence
1		ADULTE REGULIER	Adulte Régulier
2		ADULTE EXPRESS	Adulte Express
3		ADULTE INTERZONE	Adulte Interzone
8		AINE	Aine
10		EMPLOYE STO	Autre
15		CF REGULIER	Adulte Régulier
16		CF EXPRESS	Adulte Express
17		CF INTERZONE	Adulte Interzone
18		CF AINE	Aine
23		ETUDIANT ETE REGULIER	Étudiant Régulier
24		ETUDIANT ETE EXPRESS	Étudiant Express
25		ETUDIANT ETE INTERZONE	Étudiant Interzone
30		ETUDIANT INTEGRE 2003	Étudiant Régulier
31		ETUDIANT REGULIER 2003	Étudiant Régulier
34		CAMPUS REGULIER ARG	Étudiant Régulier
35		CAMPUS EXPRESS ARG	Étudiant Express
36		CAMPUS INTERZONE ARG	Étudiant Interzone
37		ETUDIANT INTEGRE 2004 (7)	Étudiant Régulier
38		ETUDIANT REGULIER 2004	Étudiant Régulier
39		ETUDIANT EXPRESS 2004	Étudiant Express
40		ETUDIANT INTERZONE 2004	Étudiant Interzone
41		INTEGRE SCOLAIRE 2004 (5)	Étudiant Régulier
42		ARCHIVE	Autre
44		ETUDIANT INTEGRE 2005 (7)	Étudiant Régulier
45		ETUDIANT REGULIER 2005	Étudiant Régulier
46		ETUDIANT EXPRESS 2005	Étudiant Express
47		ETUDIANT INTERZONE 2005	Étudiant Interzone
48		INTEGRE SCOLAIRE 2005 (5)	Étudiant Régulier

A.2 : Les lignes du réseau

Similairement aux titres de transport, il existe des différences, parfois subtiles, entre les lignes décrites dans la base de données de l'EOD et des systèmes de CAP. Les équivalences adoptées sont présentées ci-dessous.

Tableau A.2 : Équivalence des lignes du réseau (CAP)

Lignes CAP	Equivalence CAP et EOD						
1	1	51	51	98	98	639	39
3	3	52	52	201	1	642	42
5	5	54	54	221	21	643	43
11	11	55	55	233	33	647	47
17	17	58	58	235	35	649	49
20	20	61	61	243	43	654	54
21	21	63	63	274	74	655	55
25	25	65	65	277	77	666	66
26	26	66	66	296	96	676	76
27	27	67	67	327	27	696	96
28	28	68	68	333	33	731	31
29	29	69	69	338	38	733	33
31	31	72	72	339	39	735	35
33	33	74	74	431	31	737	37
35	35	76	76	433	33	739	39
36	36	77	77	439	39	742	42
37	37	78	78	467	67	743	43
38	38	80	80	477	77	747	47
39	39	81	81	500	500	749	49
42	42	82	82	533	33	750	50
43	43	83	83	537	37	754	54
44	44	84	84	539	39	767	67
45	45	85	85	542	42	771	71
46	46	86	86	547	47	776	76
47	47	88	88	549	49	839	39
48	48	89	89	627	27	849	49
49	49	94	94	633	33	900	900
50	50	96	96	637	37		

Plusieurs numéros de trajet inscrit dans la base de données des cartes à puce ne représentent qu'un segment d'une ligne, un changement de trajet ou encore un code utilisé par la STO. Cependant, lors des interviews de l'enquête ménage, il est évident que les répondants connaissent uniquement le numéro de la ligne entière. L'équivalence est donc basée sur les trajets présents dans la base de données de l'EOD. À noter, la ligne 900 correspond à un code d'erreur de la base de données des cartes à

puce. Le nombre de montées étant associé à ce code d'erreur est important, de l'ordre de 47 000 montées pour l'ensemble des journées enquêtées. Ceci représente quotidiennement environ 1 000 montées qui n'ont pas été comptabilisées correctement sur les lignes du réseau. Le manque d'information sur ces montées a obligé une répartition de ce code d'erreur selon la proportion des montées pour chaque ligne. En fait, plus de 46 000 montées ont comme arrêt d'embarquement l'arrêt numéro « 1 » ou « 2 », ce qui correspond à un autre code d'erreur de la base de données. Dans ces conditions, il est impossible de retracer géographiquement le vrai numéro de trajet de ces montées, puisque l'information n'est pas disponible. Il reste cependant une possibilité, soit l'analyse temporelle. Cette possibilité est discutée à la section 7 de ce mémoire. Bref, cette façon de répartir le code d'erreur sur l'ensemble des lignes crée un biais dans les données dont il faut tenir compte lors des analyses. De plus, les données de l'EOD contiennent aussi un code d'erreur, mais celui-ci a été négligé étant donné le faible nombre de montées associées.

A.3 : Les modes de transport

La base de données de l'EOD contient les informations sur l'ensemble des déplacements faits sur le territoire de la capitale nationale. Ces déplacements sont segmentés selon le mode de transport utilisé et un code est inscrit dans la table « DEPLACEMENT » afin de les identifier au bon mode de transport. C'est l'utilisation de ces modes qui permet d'établir la chaîne complète des déplacements pour un usager.

Le tableau A.3 ci-dessous résume donc les principaux modes utilisés. On constate qu'il y a trois modes pour le transport en commun, soit le mode « Autobus STO », le mode « Autobus OC Transpo » et le mode « Autre transport en commun ». Dans le cadre des analyses faites lors de ce mémoire, c'est le mode « Autobus STO » qui est utilisé puisque seulement les déplacements faits grâce au réseau de la Société de transport de l'Outaouais sont étudiés.

Tableau A.3: Modes de transport enquêtés (EOD)

Numéro du mode	Mode utilisé
1	Conducteur auto/fourgonnette/camionnette
2	Passager d'auto
3	Autobus OC Transpo
4	Autobus STO
5	Autre transport en commun
6	Autobus scolaire
7	O-Train
8	Traversier
9	Taxi
10	Transport adapté
11	Vélo
12	À pied
15	Moto
16	Autre mode

ANNEXE B :

LES FICHIERS DE COMPARAISONS

Les fichiers de comparaisons permettent d'étudier graphiquement et statistiquement les données contenues dans les deux bases de données. Ces fichiers ont comme fonctions d'unir les informations provenant des bases de données des systèmes de CAP et de l'EOD, de les synthétiser et de les traiter. Il est impératif que ces fichiers gardent l'intégrité des données afin de ne pas créer de biais supplémentaire. Une démarche rigoureuse d'extraction des données a donc été créée afin d'assurer de la justesse de l'information et de la répétitivité des opérations effectuées.

B.1: L'exportation des données vers le fichier de comparaison

Les requêtes sur les bases de données des systèmes de CAP et de l'enquête origine-destination sont jointes (« UNION ») afin de créer un seul fichier de comparaison. Ce fichier, de format « .xlsx » (Excel 2007), offre une interface intéressante afin de synthétiser et de comparer graphiquement l'information ce qui facilite grandement les analyses. Mais arriver à ce résultat nécessite préalablement des requêtes dans les deux bases de données. Une méthodologie stricte d'élaboration des requêtes a été mise en place afin de s'assurer de la pertinence, de la justesse et de la nécessité de l'information. Cette méthodologie ainsi que les requêtes elles-mêmes sont présentées dans les prochaines sections.

Travailler avec deux bases de données différentes peut engendrer des problèmes de compatibilité de données allant d'une simple notation différente jusqu'à l'inexistence de certains champs. Ces problèmes de compatibilité des données sont d'autant plus préoccupants lorsque les requêtes doivent être unies (« UNION »). Il est donc obligatoire d'adopter des structures et des mises en page identiques.

énormément l'analyse des données. Voici les 13 champs utilisés pour les créer les requêtes.

1. *Provenance* :
 - a. [CAP] Ce champ début toujours par « CAP » suivi de la date du déplacement.
 - b. [EOD] Ce champ a comme unique valeur « OD ».
2. *Demi-Heure* : Représente la demi-heure des déplacements.
3. *Gr Hre* : Représente le groupe d'heure, selon la segmentation faite lors de l'enquête origine-destination, des déplacements.
4. *CB_Libél_Titre*: Titre de transport utilisé lors des déplacements.
5. *NUM_LI* : Numéro de l'autobus emprunté.
6. *SommeDep* :
 - a. [CAP] Comptabilisation des déplacements ayant les mêmes caractéristiques (date, ligne, titre de transport, etc.),
 - b. [EOD] Estimation du déplacement pour une journée moyenne (sur 48 jours).
7. *Jour* : Journée du déplacement
 - a. [CAP] Lundi, Mardi, Mercredi, Jeudi, Vendredi,
 - b. [EOD] OD_Lundi, OD_Mardi, OD_Mercredi, OD_Jeudi, OD_Vendredi.
8. *SommeJournée*: Sur les 48 journées enquêtées, il y a neuf lundi, neuf vendredi, dix mardi, dix mercredi et dix jeudi. Pour faire la moyenne des déplacements pour un type de jour (par exemple, la moyenne des déplacements un lundi), il faut donc additionner tous les déplacements pour ce jour et diviser ce total par le nombre de fois que ce type de jour a été enquêté.
 - a. [CAP] Moyenne des déplacements par journées (lundi, mardi, mercredi, jeudi, vendredi),

- b. [EOD] Estimation du déplacement pour une journée moyenne (sur 48 jours).
- 9. *Jour_OD* : Le but de ce champ est de pouvoir comparer les journées moyennes CAP au jour moyen EOD (pour 48 jours).
 - a. [CAP] Lundi, Mardi, Mercredi, Jeudi, Vendredi,
 - b. [EOD] « Jour OD ».
- 10. *Origine* : Détermine si le déplacement provient de la base de données CAP ou EOD.
- 11. *typeTran* : Détermine, pour les déplacements de la base de données CAP seulement, si le déplacement est une correspondance (« 3 ») ou si c'est le premier déplacement du voyage (« 1 »). Tous les déplacements provenant de l'EOD ont le code « OD ».
- 12. *Ligne_Equi* : Les lignes d'autobus du champ NUM_LI (5) sont directement importées des bases de données. Certaines de ses lignes ne sont cependant pas présentes dans les deux bases de données, ce qui cause une perte de compatibilité entre les données. Pour y remédier, une étude sur les lignes a été faite. Cette étude a permis de constater que certaines lignes représentent des segments d'une plus grande ligne. Par exemple, la ligne 233 existe uniquement dans la base de données CAP, mais elle ne représente qu'une partie du segment de la ligne 33. Par conséquent, ce champ « ligne_Equi » permet d'attirer le bon numéro de ligne afin de s'assurer que les deux bases de données aient exactement le même nombre de circuits, et ce, sans perte d'information ().
- 13. *SommeDep_Div48* : Ce champ a pour but d'évaluer la journée moyenne des données CAP. Autrement dit, tous les déplacements de la base de données CAP sont divisés par 48 (soit le nombre de journées enquêtées) afin d'obtenir l'équivalent de la journée moyenne EOD, mais pour les données des cartes à puce.

- a. [CAP] Déplacement divisé par 48,
- b. [EOD] Somme des facteurs d'expansion.

Bien que ce format de requête couvre en grande partie les données des bases et permet de multiples comparaisons, certaines informations doivent être extraites « manuellement », par exemple les données concernant l'origine et la destination des déplacements. Les requêtes permettant l'exportation de ces données n'est pas détaillée puisqu'elles sont très similaires aux requêtes vues ci-dessus.

L'importation des requêtes faites sous Access 2007 dans le chiffrier Excel 2007 se fait directement et sans intermédiaire. Cette section a donc pour but d'expliquer la méthodologie utilisé afin de conserver la mise en page et la cohérence des données.

Dans l'onglet « Données » ou « Data », choisir l'option « Provenant d'Access » ou « From Access » de la section « Données externes » ou « Get External Data »().

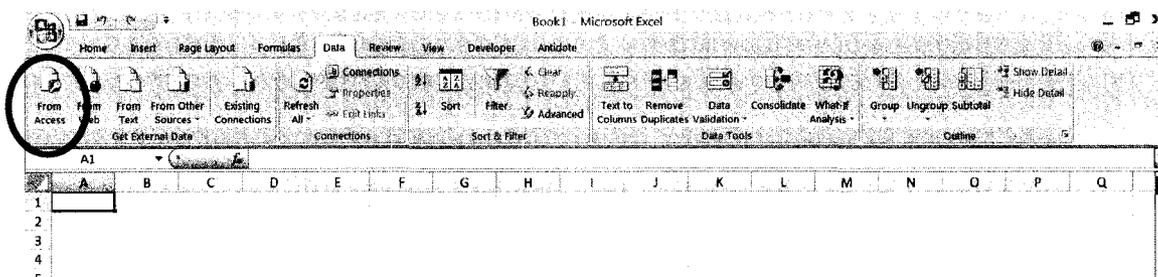


Figure B.2: Première étape de l'importation dans le fichier de comparaison

La fenêtre de dialogue suivante apparait et il suffit de chercher l'emplacement où est située la base de données et l'ouvrir.

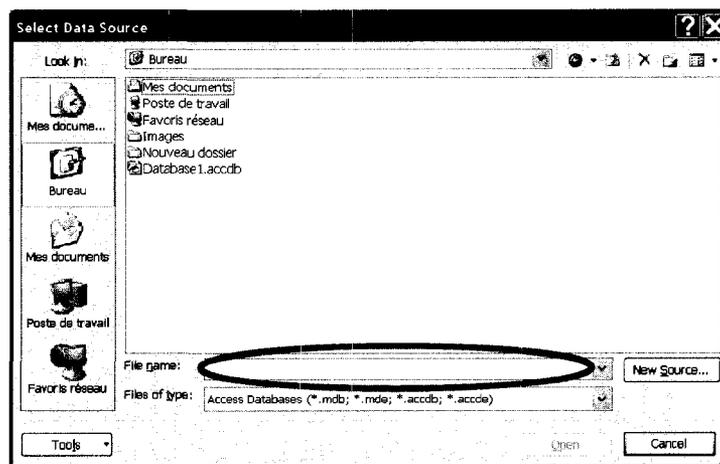


Figure B.3: Deuxième étape de l'importation dans le fichier de comparaison

Lorsque la base de données est correctement ouverte, il suffit de choisir la table ou la requête à importer.

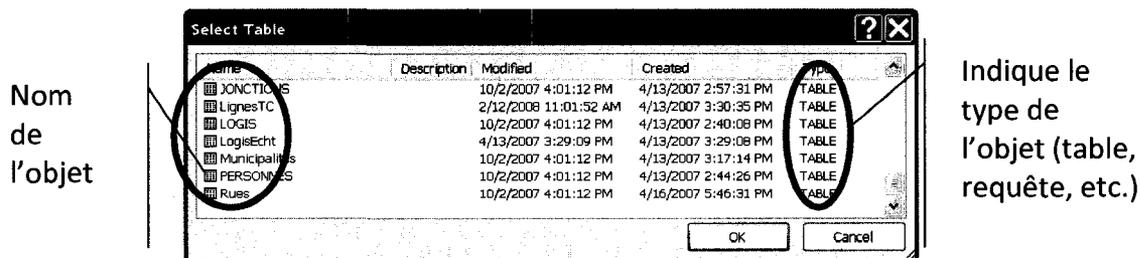


Figure B.4: Troisième étape de l'importation dans le fichier de comparaison

Avant de procéder à l'importation, un dernier choix s'impose, soit l'emplacement des données. La fenêtre ci-dessous permet de choisir sur quelle feuille et à partir de quelle cellule les données seront écrites. De plus, l'option « Table » devrait toujours être cochée, à moins de vouloir sortir des rapports de tableaux dynamiques croisés. Le fait que les données soient sous forme de « Tables » permet de modifier plus aisément les données et donne un fini esthétique beaucoup plus agréable.

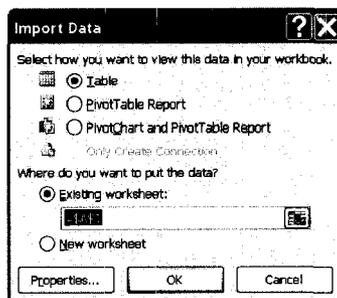


Figure B.5: Dernière étape de l'importation dans le fichier de comparaison

Les données sont maintenant dans un fichier Excel et prête à être utilisées.

B.2 : L'analyse des lignes interzones

Les lignes « interzones » comprennent seulement deux lignes, les lignes 94 et 98. Ces lignes desservent les mêmes régions, c'est-à-dire Buckingham, Masson-Angers, Gatineau, Hull et Ottawa. L'achalandage sur la ligne 98 est, selon l'EOD, environ le triple de celui sur la ligne 94. Cependant, les systèmes de cartes à puce estiment le nombre de montées sur la ligne 98 à tout près du double de la ligne 94.

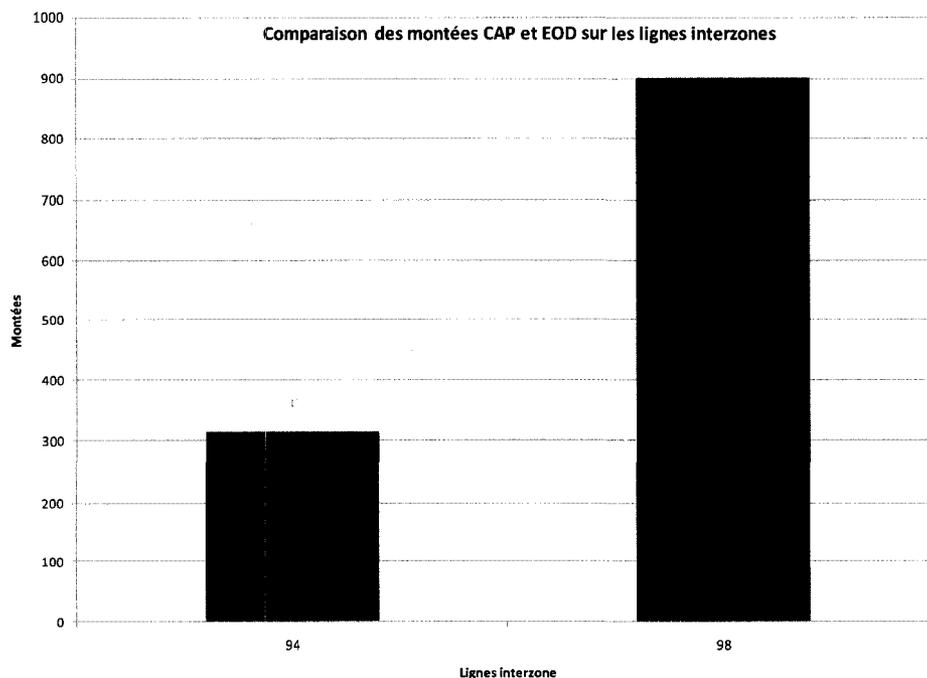


Figure B.6 : Analyse des lignes interzones (CAP et EOD)

L'écart type sur les données de la ligne 98 est d'environ 50 montées comparativement à moins de 10 montées pour la ligne 94. La variation des données de la ligne 98 est donc plus grande que celle de la ligne 94. Cependant, les estimations de l'EOD et des systèmes de cartes à puce sont différentes pour les deux lignes. En fait, la ligne 94 est légèrement sous-estimée par l'estimation de l'EOD tandis que la ligne 98 est surestimée.