

UNIVERSITÉ DE MONTRÉAL

VALIDATION PAR LOGIQUES ET TRAITEMENT DE DONNÉES CARTES
À PUCE DANS UN CONTEXTE DE TRANSPORT EN COMMUN

SÉBASTIEN GAGNÉ

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

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

AOÛT 2006



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-19301-3

Our file Notre référence

ISBN: 978-0-494-19301-3

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 :

VALIDATION PAR LOGIQUES ET TRAITEMENT DE DONNÉES CARTES
À PUCE DANS UN CONTEXTE DE TRANSPORT EN COMMUN

Présenté par : GAGNÉ Sébastien

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

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

M. BAASS Karsten, Ph.D., président

M. CHAPLEAU Robert, Ph.D., membre et directeur de recherche

M. TRÉPANIÉ Martin, Ph.D., membre

REMERCIEMENTS

L'auteur aimerait principalement remercier son directeur de recherche, M Robert Chapleau notamment pour sa remarquable disponibilité, son inextinguibilité et son entêtement bienveillant à nous faire progresser. Ce progrès se caractérise certainement par l'acquisition de compétences techniques et professionnelles mais aussi par une perception davantage critique et rigoureuse des multiples réalités associées au domaine des transports.

Des remerciements s'adressent également à la Société de Transport de l'Outaouais, à son directeur général M Georges Gratton ainsi qu'au professeur M Martin Trépanier qui ont rendu possible l'accès aux données étudiées dans cette recherche.

Merci à tous les collègues du département pour leur agréable camaraderie, leur soutien ainsi que leur complicité face aux épreuves communes.

Finalement, merci à Lise, pour son support et sans qui cette aventure n'aurait jamais existé.

RÉSUMÉ

Un système de transport en commun renferme certaines règles et régularités faisant en sorte qu'il est organisé et structuré. Par contre, à l'intérieur de bases de données censées représenter un tel système, rien ne garantit que ces règles soient respectées. C'est le cas notamment d'ensembles de données opérationnelles de type carte à puce, où chaque embarquement est traduit en un enregistrement, comportant de l'information sur les objets impliqués (usager/carte, véhicule, chauffeur). Dans ces circonstances, il est possible d'isoler et de suivre ces objets réels afin de valider leurs comportements face à des contraintes physiques et opérationnelles.

À la manière d'une enquête de transport où l'on évalue ce qui est déclaré, on désire ainsi dresser une série de mécanismes de validation pouvant être appliqués aux enregistrements afin de valider leur cohérence. Tirant profit du large échantillon disponible et du lien des données avec la carte utilisée, on souhaite proposer des traitements permettant de caractériser certains aspects de la demande et de l'offre de service. Un objectif à long terme étant d'extraire des éléments qui puissent servir comme outils d'aide à la planification.

Cette démarche de recherche entend tirer profit du fin niveau de résolution que procurent des données géolocalisées de type carte à puce. À cette fin, une approche de traitement totalement désagrégée est favorisée, couplée à diverses technologies informationnelles. On y présente une revue de littérature sur les systèmes d'information et de collection automatique de données où des traitements et imputations de données sont décrits. S'en suit une description des données opérationnelles utilisées, ainsi que l'application de logiques spatio-temporelles venant repérer les enregistrements suspects, incomplets ou manquants. Des concepts relatifs au transport en commun sont définis et explicités à travers ces données.

Également à partir des embarquements réalisés, une méthode de déduction du service planifié est exposée afin de corriger les transactions comportant une assignation erronée, repérées à l'aide de règles de validation. Une évaluation des correspondances réalisées est également faite afin d'établir les bases d'une modélisation réseau. Enfin, une caractérisation de la demande est entreprise à travers les profils de la clientèle. Des indicateurs comportementaux sont dégagés et explorés selon plusieurs perspectives et types de variabilité.

ABSTRACT

A transit system contains defined rules and regularities making it organized and structured. However, in databases supposedly representing such systems, nothing guarantees that these rules are being acknowledged. It is particularly the case with operational datasets coming from smart card media, where each boarding is being converted into a data record, which features information about the transportation objects implied (user/card, vehicle, driver). In these circumstances, it is possible to isolate and track these real objects in order to validate their behavior against physical and operational constraints.

As in the case of a travel demand survey where declared information is evaluated, a series of validation mechanisms could be applied on records in order to validate their coherence. Taking advantage of the large sample available and the connection between the data and the card used, methods characterizing certain aspects of the transportation demand and the planned transit service could be proposed. A long-term objective is to extract information elements that could be used as planning tools.

This research intends to take advantage of the fine resolution level coming from geographically-referenced smart card data. For this reason, a completely disaggregate approach is favored, in association with various informational technologies. A review of literature on information systems and automatic information gathering is presented, where various processes and imputation techniques are described. A presentation of the operational data used in this research follows, as well as the application of space-time logics in order to locate suspect, incomplete or missing records. Transit concepts are also defined and presented through the data.

Also using boarding data, a transit service derivation method is exposed in order to correct records with erroneous trip assignation, revealed by validation rules.

An evaluation of performed transit transfers is also made in order to build the basis of a network model. Finally, a travel demand characterization is undertaken through customer profiles. Behavioral benchmarks are appointed and explored according to several perspectives and variability types.

TABLE DES MATIÈRES

REMERCIEMENTS.....	iv
RÉSUMÉ.....	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xiii
LISTE DES FIGURES.....	xv
LISTE DES ACRONYMES	xxii
LISTE DES ANNEXES	xxiii
INTRODUCTION.....	1
CHAPITRE 1. SYSTÈMES D'INFORMATION ET COLLECTION	
AUTOMATISÉE DE DONNÉES.....	4
1.1 Outils d'aide à la planification	4
1.1.1 Instruments classiques.....	4
1.1.2 Automatisation des outils	9
1.2 Validité de l'information	19
1.2.1 Création d'une donnée.....	19
1.2.2 Notion de validation.....	21
1.2.3 Assistance par ordinateur	23
1.3 Traitement et imputation.....	24
1.3.1 Segmentation de traces GPS.....	25
1.3.2 Caractérisation des déplacements	25
1.3.3 Taille de l'échantillon.....	26
1.3.4 Imputation du mode	26
1.3.5 Imputation du motif.....	27

1.3.6	Qualité du service rendu	30
1.3.7	Comportement de la clientèle	30
1.3.8	Développement d'une matrice OD	31
1.4	Cadre et contraintes spatio-temporelles	32
1.4.1	Géographie temporelle.....	32
1.4.2	Théorie des activités	35
1.4.3	Enrichissement par données STI	36
1.4.4	Application au transport collectif.....	37
1.4.5	Cartes d'accessibilité	39
1.5	Régularité et données carte à puce.....	41
1.5.1	État des recherches	42
CHAPITRE 2. ÉVALUATION DES DONNÉES, VALIDATION ET LOGIQUES		44
2.1	Société de Transport de l'Outaouais	44
2.1.1	Présentation du réseau	44
2.1.2	Modes de paiement et échantillon.....	46
2.1.3	Abonnements et tarification.....	46
2.2	Structure des données	47
2.2.1	Processus de validation d'un embarquement	47
2.2.2	Processus d'assignation du véhicule	48
2.2.3	Dérivation de l'arrêt d'embarquement	49
2.2.4	Création et organisation des données.....	49
2.2.5	Données disponibles.....	51
2.2.6	Description des champs	51
2.2.7	Autres tables de données	55
2.2.8	Confidentialité des données.....	55
2.3	Concepts du transport collectif	55
2.3.1	Arrêts et arrêts-ligne.....	55

2.3.2	Déplacements et chaînes de déplacement	57
2.3.3	Véhicules, tournées et départs.....	59
2.3.4	Occupation du réseau et pointe	62
2.4	Préparation des données	63
2.4.1	Identificateur de récoltes	63
2.4.2	Séquence des transactions d'un usager-jour	63
2.5	Application de logiques de validation	65
2.5.1	Usager	65
2.5.2	Véhicule	69
2.5.3	Chauffeur	75
2.5.4	Assignation	78
CHAPITRE 3.	CARACTÉRISATION DE L'OFFRE DE SERVICE	79
3.1	Évaluation des correspondances réalisées	79
3.1.1	Méthodologie.....	80
3.1.2	Résultats	83
3.1.3	Simplification du réseau	85
3.2	Déduction du service réalisé	87
3.2.1	Méthodologie.....	87
3.2.2	Résultats	91
3.2.3	Vers une correction de l'assignation	94
CHAPITRE 4.	CARACTÉRISATION DE LA DEMANDE	96
4.1	Détermination des profils d'utilisateur	96
4.1.1	Attributs et méthodologie	97
4.1.2	Numéro de carte	97
4.1.3	Apparition mensuelle.....	98
4.1.4	Titre de transport.....	98
4.1.5	Arrêt-domicile	99
4.1.6	Facteur d'éloignement.....	100

4.1.7	Indicateurs de mobilité	101
4.2	Analyse des résultats	102
4.2.1	Répartition mensuelle des cartes	103
4.2.2	Démographie des usagers	103
4.2.3	Taux de déplacement	107
4.2.4	Taux de correspondances/déplacement	110
4.2.5	Taux d'embarquements/déplacement lié	111
4.2.6	Jours d'activité	111
4.3	Variabilité sur échantillon constant	115
4.3.1	Représentativité de l'échantillon	116
4.3.2	Variabilité du titre et de l'arrêt-domicile	117
4.3.3	Taux de déplacement	119
4.3.4	Taux de correspondances/déplacement	120
4.3.5	Analyse transversale vs longitudinale	121
CONCLUSION		123
RÉFÉRENCES		127
ANNEXES		134

LISTE DES TABLEAUX

Tableau 1.1 – Motifs de déplacement possibles selon utilisation du sol (Wolf et al, 2001). Le code « 99 » équivaut à une utilisation mixte, ambiguë ou inconnue.....	28
Tableau 2.1 – Statistiques du réseau de la STO, juillet 2006 (Source : www.sto.ca).....	45
Tableau 2.2 – Tarification par catégorie d'usager (Source : Société de Transport de l'Outaouais, 2006)	47
Tableau 2.3 – Description des champs rattachés à chaque embarquement.....	52
Tableau 2.4 – Titres utilisés au mois de février 2005, triés selon nombre de transactions	52
Tableau 2.5 – Messages d'erreur relatifs aux transactions refusées (octobre 2003 et février 2005)	54
Tableau 2.6 – Récoltes où l'on constate un décalage important de l'horloge interne	68
Tableau 2.7 – Indicateurs relatifs à la qualité des données étudiées : juillet 2003, octobre 2003 et février 2005.....	73
Tableau 2.8 – Éléments informationnels de deux transactions successives où une vitesse de 358 km/h est dérivée	75
Tableau 3.1 – Légende des codes d'anomalie et de leur répartition	83
Tableau 3.2 – Clichés d'enregistrement et description des champs de la table des tracés-sens-départ (TSD)	90
Tableau 4.1 – Clichés d'enregistrement et description de la table des profils, caractérisant les cartes en service	97

Tableau 4.2 – Catégorisation des titres de transport en groupe démographique	98
Tableau 4.3 – Comparaison entre taux de déplacement (non-liés) de Bagchi (2003) et STO (juillet, octobre 2003 et février 2005).....	110
Tableau 4.4 – Comparaison entre taux de transactions par déplacement (lié) de Bagchi (2003) et STO (juillet, octobre 2003 et février 2005).....	111
Tableau 4.5 – Distribution ordonnée et cumulée des combinaisons de jours d'activité pour une semaine (lundi 21 au dimanche 27 février 2005).....	113
Tableau 4.6 – « Top 10 » des titres les plus utilisés pour les cartes avec un seul jour d'activité mensuelle (juillet 2003, octobre 2003 et février 2005)	114
Tableau 4.7 – Répartition des messages d'erreur pour les transactions uniques des mois d'octobre 2003 et février 2005	115
Tableau 4.8 – Taux de déplacement de l'échantillon constant selon mois et groupe démographique	119
Tableau 4.9 – Taux de correspondances/déplacement de l'échantillon constant selon mois et groupe démographique.....	120

LISTE DES FIGURES

Figure 1.1 – Opposition schématique des méthodes analytiques agrégées et totalement désagrégées pour des déplacements entre deux zones	7
Figure 1.2 – Itinéraire de déplacement totalement désagrégé (Source: Chapleau et al, 2002)	8
Figure 1.3 – Localisation d'un récepteur à la surface de la Terre (Source : home.arcor.de/m.panitzki/html/navigation/gps.htm)	12
Figure 1.4 – Composantes d'une carte à puce sans contact (Source : www.calypsonet-asso.org)	14
Figure 1.5 – Borne de validation et carte à puce (Source : www.ratp.fr)	14
Figure 1.6 – Taxonomie des technologies STI répertoriées par le Advanced Public Transportation Systems (Source : http://web.mitretrek.org/its/aptsmatrix.nsf/taxonomy)	17
Figure 1.7 – Hiérarchie des niveaux de résolution spatiale des applications visant l'acquisition automatique de données (Furth et al, 2004).....	20
Figure 1.8 – Cheminements spatio-temporels sur 24 heures des éléments d'un ménage (tiré de présentation de Kwan, 2002).....	34
Figure 1.9 – Prismes spatio-temporels selon mode, entre une origine et une destination	34
Figure 1.10 – Cheminements dans l'espace-temps des femmes afro-américaines, provenant des données de l'enquête « Activity-Travel » de Portland de 1994-95 (Source : Kwan, http://geog-www.sbs.ohio-state.edu/faculty/mkwan/Gallery/STPaths.htm).....	37

Figure 1.11 – Cheminements espace-temps possibles afin de se rendre d'un point A au point B entre les temps i et j, en utilisant un service de transport collectif	38
Figure 1.12 – Prisme spatio-temporel délimitant l'arbre de potentialités discret associé à un réseau de transport collectif.....	39
Figure 1.13 – Prismes spatio-temporels tridimensionnels (Source: O'Sullivan, 2000)	40
Figure 1.14 – Carte d'accessibilité TC du campus de l'UCSB (à gauche) et du centre-ville (au centre) afin d'arriver à 8h00 (Source : Church et al, 2005)	41
Figure 1.15 – Représentation géographique (à gauche) du réseau de métro de Londres, puis réorganisée selon les temps de parcours depuis la station Wembley Park (à droite) (Source interactive : http://www.tom-carden.co.uk/p5/tube_map_travel_times/applet).....	41
Figure 2.1 – Réseau de la STO superposé aux secteurs de la ville de Gatineau (Source : www.ville.gatineau.qc.ca et www.sto.ca).....	45
Figure 2.2 – Caractérisation des objets engendrant une transaction ainsi que la transaction engendrée, qui complète l'information sur un embarquement	48
Figure 2.3 – Processus d'acquisition et de sauvegarde des transactions pour une chaîne de départs.....	50
Figure 2.4 – Décomposition des séquences de transactions en récoltes, jours et départs.....	50
Figure 2.5 – Diagramme fonctionnel du paiement par carte à puce à la STO (Source : Trépanier et al, 2005).....	51
Figure 2.6 – Distinction entre arrêts et arrêts-ligne	56

Figure 2.7 – Parcours et arrêts associés couvrant une partie du secteur d'Aylmer	57
Figure 2.8 – Représentation d'une chaîne de déplacement composée de deux déplacements, où trois transactions ont été observées pour un usager	58
Figure 2.9 – Nombre de transactions de validation et de correspondance par jour (juillet 2003, octobre 2003 et février 2005)	59
Figure 2.10 – Nombre de bus, tournées, départs et ratio moyens de départs par bus pour les jours de février 2005	60
Figure 2.11 – Intervalles temporels des tournées pour un jour moyen de semaine (haut : une ligne par tournée, bas : une ligne par bus).....	61
Figure 2.12 – Estimation de l'occupation du réseau (ventilé par catégorie de titre), des bus en service et du taux d'occupation dérivé par quart d'heure (2 juillet 2003)	62
Figure 2.13 – Distinction entre numéros de transaction originaux et numéros de séquence d'un usager-jour	64
Figure 2.14 – Dates d'embarquement présentes dans les enregistrements des 3837 récoltes et répartition du nombre de jours par récolte (juillet 2003)	66
Figure 2.15 – Transactions de la carte 1923469362, du 14 au 18 juillet 2003...	68
Figure 2.16 – Répartition standard et cumulée des intervalles de temps entre une correspondance et sa validation simple associée (juillet 2003)	69
Figure 2.17 – Répercussion de l'omission d'un changement de départ sur l'information des transactions	71

Figure 2.18 – Distribution des vitesses dérivées entre deux transactions consécutives selon seuil temporel maximal entre les deux transactions (2 juillet 2003).....	74
Figure 2.19 – Étalement temporel des transactions des bus 9403 et 9425, le 31 juillet 2003 (même chauffeur)	76
Figure 2.20 – Distribution (aux quinze minutes) des intervalles entre premier et dernier embarquement, par chauffeur-jour-bus (pièces de travail) et chauffeur-jour (amplitudes) (juillet 2003)	77
Figure 2.21 – Étalement temporel des transactions pour les cinq pièces de travail dérivées les plus importantes (juillet 2003)	77
Figure 3.1 – Simplification des arrêts du réseau	80
Figure 3.2 – Détermination de l'arrêt de descente le plus probable lors d'une correspondance.....	82
Figure 3.3 – Répartition des correspondances par code d'anomalie, ainsi que distribution de la distance entre paires d'arrêts issues de correspondances valides.....	83
Figure 3.4 – Segments de transfert entre paires d'arrêts avec leurs intensités respectives (juillet2003).....	85
Figure 3.5 – Distribution des distances (< 500 m) entre tous les arrêts	86
Figure 3.6 – Détail de l'agrégation à 75 m des arrêts.....	86
Figure 3.7 – Identification du départ précédent et suivant le plus fréquent au départ i, sur les jours où il apparaît dans les transactions.....	88
Figure 3.8 – Section d'un bus-tournée type	88
Figure 3.9 – Distribution des départs selon nombre de jours apparus (juillet 2003)	91

Figure 3.10 – Distribution des départs selon régularité et régularité inversement cumulée (I.C.) du départ précédent, du départ suivant, du chauffeur et de la durée (juillet 2003)	93
Figure 3.11 – Distribution des débuts des tournées selon l'heure du 1 ^{er} embarquement, ventilé par le nombre de départs par tournée.....	94
Figure 3.12 – Schéma illustrant la correction des transactions (possédant une assignation erronée) par le service déduit.....	95
Figure 4.1 – Processus de synthèse de l'information d'une carte/usager	96
Figure 4.2 – Représentation spatiale et schématique des transactions d'une carte sur trois jours. Pour les trois jours, l'arrêt A est le premier arrêt de la journée.....	99
Figure 4.3 – Distribution statistique par boîtes à moustache du champ RatioArrêt (juillet 2003, octobre 2003 et février 2005)	100
Figure 4.4 – Intensité (axe z) des embarquements aux arrêts en pointe PM (juillet 2003).....	101
Figure 4.5 – Répartition des cartes selon appartenance aux ensembles de données.....	103
Figure 4.6 – Répartition des arrêts-domicile selon leur nombre de cartes d'utilisateur rattachées (février 2005).....	104
Figure 4.7 – Agrégation spatiale du nombre de cartes d'utilisateur par zone de 0,25 km ² (février 2005)	105
Figure 4.8 – Variation du nombre de cartes de type étudiant entre le mois de juillet 2003 et octobre 2003, par zone de 0,25km ² (Représentation surfacique et par grille)	106

Figure 4.9 - Évolution du taux moyen de déplacements/jour de semaine, selon groupe démographique (juillet 2003, octobre 2003 et février 2005)	108
Figure 4.10 – Régression linéaire entre les taux de déplacement et le nombre de cartes en service (selon type de carte et mois)	108
Figure 4.11 – Taux de déplacement et nombre de cartes (moyen) selon éloignement au centre, par ensemble de données et par groupe démographique	109
Figure 4.12 – Évolution du taux de correspondances/déplacement en semaine, selon groupe démographique (juillet 2003, octobre 2003 et février 2005)	110
Figure 4.13 – Distribution des pourcentages de cartes en activité selon nombre de jours dans le mois (juillet, octobre 2003 et février 2005)	112
Figure 4.14 – Distribution du pourcentage de cartes selon nombre de jours d'activité en semaine (juillet, octobre 2003 et février 2005)	113
Figure 4.15 – Distribution des pourcentages de cartes selon le jour où a eu lieu l'unique jour d'activité dans le mois (juillet, octobre 2003 et février 2005)	114
Figure 4.16 – Échantillon comprenant les mêmes éléments dans les trois ensembles de données	116
Figure 4.17 – Répartition des groupes démographiques des cartes selon l'ensemble étudié (juillet, octobre 2003 et février 2005)	117
Figure 4.18 – Constance de l'arrêt-domicile et distribution selon distance maximum pour les arrêts-domicile non identiques (3981 cartes apparaissant dans mois de juillet 2003, octobre 2003 et février 2005)..	118

- Figure 4.19 – Variation du ratio de validations simples par jour de semaine, distribué selon l'éloignement radial au centre. Arrière-plan : distribution radiale du nombre de cartes..... 120
- Figure 4.20 – Variation du taux de correspondance/déplacement en semaine distribué selon l'éloignement radial au centre. Arrière-plan : distribution du nombre de cartes 121
- Figure 4.21 – Méthode des sondages répétés vs analyse longitudinale. Application aux taux de déplacement des 3981 cartes de l'échantillon. Distribution T à 95 % (degrés de liberté ∞) 122

LISTE DES ACRONYMES

AFC	Automatic Fare Collection
AMT	Agence métropolitaine de transport
APC	Automatic Passenger Count
APTS	Advanced Public Transportation Systems
AVL	Automatic Vehicle Location
CAP	Carte à puce
FTA	Transit Administration
GIS	Geographic Information System
GPS	Global Positionning System
ITS	Intelligent Transportation System
OD	Origine-destination
RFID	RadioFrequency IDentification
SIG	Système d'information géographique
SIVT	Système informatisé de validation des titres
STI	Système de transport intelligent
STM	Société de transport de Montréal
STO	Société de transport de l'Outaouais
TC	Transport en commun

LISTE DES ANNEXES

ANNEXE A – Échantillons des tables de données.....	134
ANNEXE B – Requêtes SQL : Application des logiques	135
ANNEXE C – Script évaluant les numéros de séquence des transactions d'un usager-jour	141
ANNEXE D – Script de caractérisation et de validation des correspondances.....	143
ANNEXE E – Script de caractérisation des tracés-sens-départ depuis les embarquements réalisés	146
ANNEXE F – Script de caractérisation des profils d'usager.....	151

INTRODUCTION

Qu'est-ce qui peut être dérivé de données opérationnelles de type carte à puce? Quel est le degré de fiabilité de cette information rapportée? Ces questions constituent le cœur de cette démarche de recherche. Elles acquièrent leur pertinence en regard de l'implantation croissante de ces systèmes à l'intérieur de sociétés de transport et de l'impact que leurs réponses pourraient avoir sur les méthodes de planification actuelles.

Dans un système carte à puce, chaque embarquement de la part d'un usager se traduit en un enregistrement, comportant de l'information détaillée sur chaque objet de transport concerné (usager/carte, véhicule, chauffeur). Ces enregistrements, emmagasinés par dizaines de milliers quotidiennement, fournissent une source de données inédite quant au volume d'information ainsi qu'au lien avec chaque carte/usager. C'est dans l'optique d'une utilisation pertinente de cette information nouvelle que cette recherche s'inscrit.

À la différence des enquêtes origine-destination, où un échantillon (typiquement 5 % pour Montréal) décrivant les déplacements d'une journée type s'avère assez significatif pour extrapoler la mobilité d'une population, les données opérationnelles de types carte à puce expriment directement, et sur de vastes périodes, la mobilité de tous les détenteurs.

Avec de larges ensembles de données, comportant une traçabilité des objets de transport et où la perspective d'observation est variable, on retrouve tous les éléments (composantes, complexité, problématique) habituels des résultats qu'on obtient avec les modes opérationnels actuels et qu'on examine pour fins de planification. Les données carte à puce nous fournissent d'ailleurs une opportunité de mettre en contraste ces résultats.

Validation par logiques

On est cependant dans une problématique où il existe un modèle en matière de transport en commun. C'est un modèle qui est contraint et qui comporte certaines exigences sur la validité de l'information qui le décrit. L'objectif initial de cette démarche de recherche est ainsi de présenter certains mécanismes de validation de l'information brute. À la manière d'une enquête origine-destination où l'information déclarée est évaluée, les données carte à puce peuvent faire ressortir les données par objet de transport et ainsi les confronter aux contraintes physiques et opérationnelles auxquelles ces objets sont soumis. On recourt ainsi à des filtres informationnels qui viennent homologuer des situations réelles.

Traitement de données

L'autre objectif escompté est qu'à partir de ces données représentant des embarquements, on puisse y dériver de l'information nouvelle permettant de caractériser la variabilité de la demande, de l'offre, la régularité, la ponctualité, etc. afin de pouvoir en tirer des éléments qui puissent servir comme outils d'aide à la planification. L'approche de dérivation de l'information se transpose ainsi par des techniques et des traitements, effectués dans l'optique d'une approche totalement désagrégée, qui désirent parvenir à une bonification des processus de connaissance.

Articulation de la recherche

Ce mémoire est organisé en quatre chapitres, résumés dans les prochains paragraphes.

Revue de littérature

Après une brève présentation d'outils et technologies visant l'acquisition de données opérationnelles de transport collectif, le chapitre aborde la notion de validité de l'information, acquise via les systèmes d'information en vigueur. On décrit certaines analyses, traitements et imputations effectués sur ces ensembles de données. Les contraintes spatio-temporelles sont abordées à travers les concepts de la géographie temporelle et de la théorie des activités.

Présentation des données et validation

Le chapitre présente brièvement la Société de Transport de l'Outaouais, d'où proviennent les données étudiées, ainsi que son contexte géographique, tarifaire, etc. Le processus d'acquisition des données via les montées à bord est expliqué ainsi que la structure du système informationnel. Suite à l'élaboration et l'application de mécanismes de validation à l'aide de logiques spatio-temporelles, diverses problématiques et données suspectes sont ainsi dégagées.

Caractérisation de l'offre de service

Deux méthodologies susceptibles de bonifier notre connaissance de l'offre sont décrites. Avec comme objectif à long terme une modélisation réseau, la première s'attarde sur les correspondances effectuées par la clientèle, où des logiques spatio-temporelles mettent à l'épreuve leur validité. L'autre méthode, présentée dans le but d'aboutir à une correction des enregistrements suspects, utilisera les mêmes données d'embarquement afin de déterminer l'horaire complet pour un jour de semaine, à travers la régularité des chaînes de départs rencontrées.

Caractérisation de la demande

La vraie demande c'est le client. C'est dans cette optique que le chapitre analyse diverses facettes de la demande de transport à travers les profils de consommation du réseau par la clientèle. Des indicateurs comportementaux sont dégagés et explorés selon plusieurs perspectives et types de variabilité.

CHAPITRE 1. SYSTÈMES D'INFORMATION ET COLLECTION AUTOMATISÉE DE DONNÉES

Après une brève présentation d'outils et technologies visant l'acquisition de données opérationnelles de transport collectif, ce chapitre dresse un inventaire de la littérature relative à l'application de méthodes analytiques utilisant ces données opérationnelles.

1.1 Outils d'aide à la planification

1.1.1 Instruments classiques

Le planificateur de réseau de transport doit ses interventions aux informations que lui procurent ses outils. Un historique de quelques techniques est présenté.

Comptages

Les comptages manuels représentent une des premières techniques afin de quantifier l'utilisation faite des services de transport en commun offerts et selon un sondage récent du Transportation Research Board (Boyle, 1998), ils sont encore largement utilisés. Souvent exécutés à l'embarquement des usagers par des pointeurs voyageant à l'intérieur des véhicules ou par comptages à des endroits donnés, ces procédures manuelles servent souvent de base de comparaison à d'autres technologies de recensement des embarquements.

Enquêtes à bord

L'enquête à bord consiste en un questionnaire remis à l'utilisateur, habituellement à l'entrée des véhicules. Une fois complété, l'utilisateur le rend à la sortie du véhicule ou par la poste. En plus d'incorporer des renseignements sur les déplacements, l'information recueillie peut renfermer des données sur les motifs de déplacement (travail, études, loisirs, etc.) et des renseignements individuels tels

l'âge, le sexe, la langue, le lieu de domicile, etc. Cette information permet ainsi une association entre le profil sociodémographique de la clientèle et sa consommation sur le réseau.

Enquêtes origine destination

Les enquêtes origine destination (OD), effectuées souvent par entrevues téléphoniques, recueillent des descriptions des déplacements effectués par les éléments d'un ménage, généralement au cours d'une journée. La nature de l'information recueillie est multiple. En plus d'une spécification des déplacements, elle comprend une caractérisation démographique et socio-économique du ménage et de ses individus.

Les grandes agglomérations utilisent ce type d'enquête afin de connaître les besoins de mobilité de la population et à détailler une stratégie d'action à long terme de leurs systèmes et infrastructures de transport. Stopher (1975) résume en ces termes les principaux objectifs (théoriques) des enquêtes transport :

- fournir un plan détaillé et continu du développement, de l'évaluation et de l'implantation des propositions éventuelles de planification de transport;
- mettre l'accent sur le rôle de premier plan du transport dans les activités urbaines;
- faire une évaluation détaillée de l'effet de l'alternance des politiques de transport et
- effectuer une allocation des priorités en matière de d'investissement et développement.

Dans le cas de Montréal, ces enquêtes sont effectuées environ aux cinq ans. Pour l'année 2003, le document préparé par le Secrétariat à l'enquête Origine-Destination (2005) fait état d'un territoire couvert de 5 500 km² et touchant une population de 3,6 M de personnes. L'information recueillie lors des entrevues est la suivante :

- Ménage : localisation, nombre de personnes, nombre de véhicules, adresse précédente du domicile (dans les cas où les ménages résident à cette adresse depuis moins de cinq ans), revenu global.
- Personnes du ménage : âge, sexe, possession d'un permis de conduire, occupation principale, lieu habituel de travail ou d'études.
- Déplacements : motif, origine, destination, heure de départ, modes de transport utilisés incluant lignes de transport en commun, point de transfert modal, mode de paiement du déplacement en transport en commun ou type de stationnement, type de covoiturage, utilisation des autoroutes, utilisation des ponts.

Au total, 71 400 ménages ont été interrogés afin de décrire les caractéristiques de 366 300 déplacements réalisés par les 169 900 personnes composant ces ménages. L'échantillon en résultant représente près de 5 % des ménages de la région. Le couplage des données d'enquête avec des éléments provenant de sources externes renforce la capacité à décrire les comportements de mobilité de la population. À cet égard, Chapleau et al (2001) précisent que

« L'exploitation, selon une approche totalement désagrégée, des informations issues de ces enquêtes cerne les potentialités analytiques des données de mobilité et autorise l'examen de problématiques complexes par l'intégration cohérente de multiples sources (recensements, système d'activités, territoire) et systèmes d'information (gestion opérationnelle des réseaux, transactions tarifaires, diffusion d'information aux usagers). »

Approches d'analyse

La citation précédente fait mention de l'approche totalement désagrégée. Dans cet ordre d'idées, deux approches analytiques sont essentiellement utilisées afin de traiter les données issues de ce type d'enquête : l'approche agrégée (procédure séquentielle classique) en vigueur depuis les années cinquante et l'approche totalement désagrégée, utilisée depuis les années soixante-dix, notamment dans le cas des enquêtes de la grande région de Montréal.

La modélisation agrégée, encore couramment utilisée, découpe le territoire en zones d'analyses où les caractéristiques généralisées de sa population y sont rattachées. Chaque déplacement à l'intérieur de la zone se voit attribuer les caractéristiques de la zone. Pour chaque zone d'analyse, on applique le paradigme de la procédure séquentielle classique, soit la génération des déplacements, la distribution sur les zones, la répartition modale de ces déplacements ainsi que leur affectation sur le réseau.

L'analyse totalement désagrégée prend plutôt avantage de la nature discrète des déplacements échantillonnés puisqu'elle ne perd aucune caractéristique individuelle et n'agrège que les résultats. La Figure 1.1 illustre de façon simplifiée comment les déplacements entre deux zones peuvent être perçus selon la méthode totalement désagrégée, qui les traite indépendamment, et la méthode agrégée, qui les rapporte sur les centroïdes de zones.

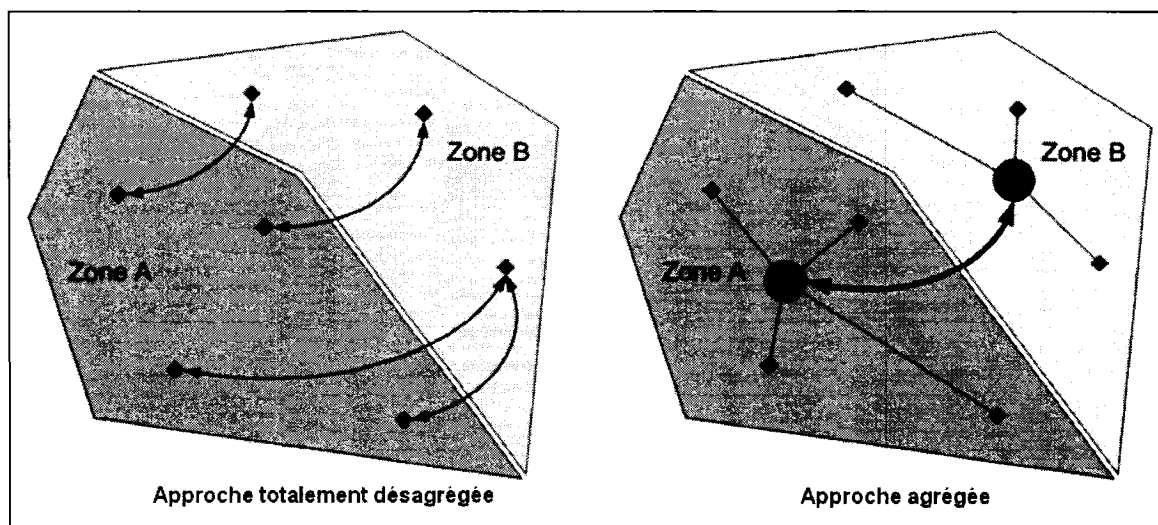


Figure 1.1 – Opposition schématique des méthodes analytiques agrégées et totalement désagrégées pour des déplacements entre deux zones

Approche totalement désagrégée

Une des spécificités des enquêtes OD de la grande région de Montréal est l'emploi de l'approche totalement désagrégée. Ici, chaque déplacement est

décomposé au plus fin niveau possible, et sa caractérisation est adaptée à chaque type de réseau emprunté. De telle sorte qu'un déplacement effectué en transport en commun sera segmenté par mode (incluant les accès à pied) et selon les nœuds et liens TC empruntés (Figure 1.2).

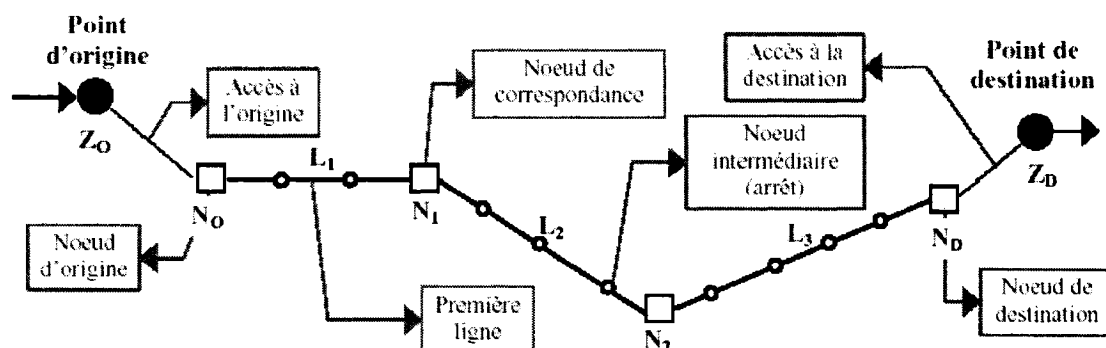


Figure 1.2 – Itinéraire de déplacement totalement désagrégé (Source: Chapleau et al, 2002)

Par la résolution très fine des éléments composant les déplacements, l'approche totalement désagrégée permet un niveau avancé de validation de l'information, notamment de validations réseau.

« Cette approche assimile plusieurs procédures de validation des données qui mettent une emphase particulière sur la cohérence informationnelle des logiques spatiales. Les données géoréférencées sont exploitées afin de déterminer les points d'accès et de correspondance les plus probables, de vérifier les logiques spatiales et temporelles des chaînes de déplacement [...] » (Chapleau, 1999)

Les données opérationnelles issues de systèmes d'information, particulièrement de données carte à puce couplées à de l'information géolocalisée, favorisent ce type d'approche. C'est d'ailleurs sous cet éclairage que les données opérationnelles contenues dans cette recherche seront étudiées et mises en contexte.

Problématiques des enquêtes

Malgré le haut niveau de résolution que permet l'approche totalement désagrégée, il reste qu'elle s'appuie sur la qualité de l'information analysée. Les problèmes avec l'information recueillie d'une manière auto-rapportée (self-reported) [par les enquêtes OD] incluent la sous représentation des déplacements de courte durée, la faible qualité des données temporelles des débuts/fins de déplacement, les durées précises des déplacements et l'emplacement des destinations (Murakami et al, 1999). En effet, Yalamanchili et al (1999) rapportent que dans l'esprit des gens sondés, les déplacements de courte durée sont souvent associés à des déplacements peu significatifs et ainsi sont facilement enclins à être oubliés. Malgré une grande variation sur le plan des résultats, cette affirmation est vérifiée lorsque des données GPS sont utilisées afin de confronter les réponses des gens interrogés (Yalamanchili et al, 1999, Wolf, 2004). À ce propos, la prochaine section présente un éventail de technologies permettant l'acquisition de données de manière automatisée.

1.1.2 Automatisation des outils

De nouvelles technologies prennent en charge de façon automatique et à faible coût l'acquisition de données de transport de sorte qu'il est maintenant plus facile d'accumuler une impressionnante quantité d'information. Les prochaines sections présentent un aperçu des technologies fréquemment recensées dans la littérature, traitant de données opérationnelles issues de transport collectif.

Comptages automatiques

On retrouve des systèmes de comptages automatisés (Automatic Passenger Counting - APC) dont la détection se fait mécaniquement (sensible à la pression) ou par détection optique ou infrarouge (Furth, 2000). Ces systèmes automatiques de recensement des embarquements (parfois débarquements) peuvent être couplés à d'autres systèmes, dont de géolocalisation. Les systèmes de comptages automatisés peuvent comporter des erreurs de comptages

pouvant être atténuées par des modèles d'ajustement logiciels (calibrage) ou, pour les sociétés disposant des comptages à la descente, par un équilibrage des entrants/sortants. À cet effet, la STM (Société de Transport de Montréal) procède à des tests de validation rigoureux sur ses données APC avant de les utiliser comme indicateurs de performance (Furth et al, 2004).

Systèmes de perception électronique

Ce type de système (Automated Fare Collection - AFC) permet de valider automatiquement les embarquements payés en argent comptant en plus d'accepter d'autres titres de transport comme le jeton, ticket ou carte magnétique. Ils permettent un recensement de la clientèle, souvent ventilé par catégorie de titre, essentiellement afin d'en tirer de l'information sur l'achalandage.

Afin de raffiner cette information, certains dispositifs existent afin de segmenter les périodes d'embarquement par jour, par tournée, par ligne, par heure ou au mieux, par départ. Une segmentation manuelle impliquant le chauffeur s'avère souvent problématique (induit des erreurs) alors qu'une segmentation quotidienne ou horaire a l'avantage de pouvoir être totalement automatisée (Furth, 2000).

Furth mentionne que la qualité des données de ces systèmes est difficile à maintenir souvent par rapport au degré d'implication des diverses catégories d'employés (chauffeurs, opérateurs de transfert de données). Par contre, beaucoup de ces problèmes tendent à disparaître avec l'apparition de nouveaux lecteurs, majoritairement liés à un ordinateur de bord et où la grande majorité des interventions manuelles sont éliminées. Cette liaison permet aussi une certaine évolution de ces systèmes alors que chaque transaction peut être combinée à d'autres types d'information (spatiale, temporelle, opérationnelle, etc.).

Géolocalisation et système GPS

Plusieurs technologies permettent un positionnement spatial des véhicules de transport en commun. Il y a les balises infrarouges, les boucles à induction magnétique et l'utilisation du système de positionnement par satellite (Global Positionning System - GPS). La suite de cette section ne portera cependant que sur le système GPS étant donné le nombre imposant d'études utilisant ce type de données.

Le système GPS permet de localiser un objet physique de manière extrêmement précise dans l'espace-temps. Mis en place par le Département de la Défense des États-Unis, le système consiste en 24 satellites positionnés sur plusieurs orbites autour de la Terre qui émettent en tout temps un signal comprenant leur position ainsi que l'heure précise de transmission¹.

Un récepteur GPS, recevant le signal d'un satellite saura calculer sa distance à celui-ci en se basant sur l'information spatio-temporelle. En répétant le procédé pour deux autres satellites, on parvient à faire coïncider trois cercles virtuels qui établiront avec précision la localisation du récepteur sur la surface de la terre à un instant précis. Notons qu'avant le 1^{er} mai 2000, la précision du système était volontairement dégradée (brouillage sélectif).

¹ Source : <http://fr.wikipedia.org/wiki/GPS>

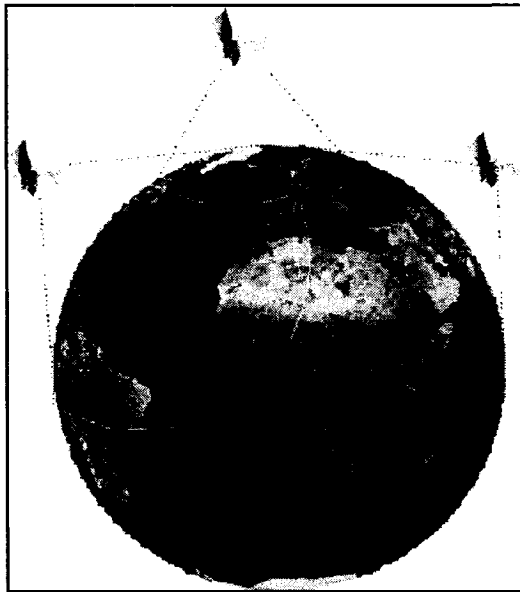


Figure 1.3 – Localisation d'un récepteur à la surface de la Terre
 (Source : home.arcor.de/m.panitzki/html/navigation/gps.htm)

Avec plus de trois satellites disponibles, il devient possible de dériver une troisième dimension spatiale (l'élévation), de majorer la précision 2D ou estimer la vitesse si le récepteur permet cette option. L'intérêt du point de vue des objets de transport en commun se trouve dans l'enregistrement de ces positions de manière systématique pour un traitement en temps réel ou en différé. Ainsi, ce type de système automatisé de localisation (Automated Vehicle Location - AVL) sera souvent adjoint à d'autres composantes visant à compléter l'information, tels des systèmes embarqués de comptage APC.

On catégorise habituellement les récepteurs GPS en deux classes : passifs ou actifs. Les récepteurs passifs emmagasinent l'information indépendamment du chauffeur ou d'un ordinateur de bord, le récepteur GPS n'alignant qu'une longue série de positions spatio-temporelles à intervalle temporel (ou spatial) fixe. Bullock, Jiang et Stopher (2005) soulignent que bien que de nombreuses et pertinentes analyses peuvent en découler, les systèmes GPS passifs, non couplés à un autre système d'information s'avèrent pauvres en données de référence lorsque l'on veut analyser de façon globale le système étudié.

D'autre part, les récepteurs actifs accompagnent leur positionnement géographique et temporel à d'autres renseignements liés aux objets de transport et à leur opération (ex : assignation du véhicule, début/fin d'une tournée/voyage, changement de chauffeur, etc.)

Les systèmes de géolocalisation par GPS peuvent être bonifiés afin de pallier certaines situations où la qualité du positionnement s'affaiblit, comme dans les canyons urbains (signal réfléchi par les édifices) ou les endroits souterrains. Une de ces bonifications est la navigation à l'estime (Dead Reckoning - DR) qui consiste à déduire la position d'un véhicule à partir de la précédente, de la trajectoire assumée et de d'autres données, comme la vitesse, l'orientation du véhicule, la distance parcourue, etc. Une autre technique est le GPS assisté (Assisted GPS - AGPS) où le récepteur GPS utilise les positions (connues) des autres éléments du réseau afin de clarifier sa propre position (Wolf, 2004).

Carte à puce

Cette technologie commercialisée dès la fin des années soixante-dix (Trépanier et al, 2005) consiste essentiellement en une carte plastifiée à laquelle on a incorporé une puce informatique capable d'emmagasiner (et dans certains cas, de traiter) de l'information. La carte à puce (CAP) interagit avec le reste du système avec ou sans contact physique. Dans ce dernier cas, c'est une antenne interne qui assure la communication avec la borne de validation par infrarouge ou radiofréquence (RadioFrequency IDentification - RFID). Selon la technologie (proximity card vs vicinity card), l'identification à la borne peut se faire à l'intérieur de 10 cm ou 70 cm de distance.

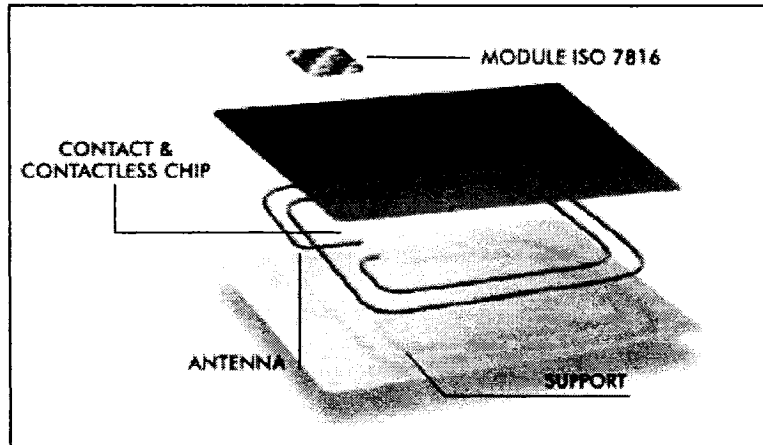


Figure 1.4 – Composantes d'une puce sans contact
(Source : www.calypsonet-asso.org)



Figure 1.5 – Borne de validation et carte à puce
(Source : www.ratp.fr)

Dans les conjonctures relatives aux transports collectifs, les cartes à puce sont notamment utilisées par les usagers comme titre de transport afin de valider leur embarquement à bord des véhicules. Une carte à puce peut servir de support à un abonnement (souvent mensuel) ou agir comme carte débitrice. Le degré de personnalisation de la carte à puce est variable selon qu'on y inclut des éléments d'information nominale sur l'utilisateur et/ou sa photo, ce qui la rend plus difficilement transférable. Pour fins de planification, la grande pertinence de ce type de technologie réside dans le fait que toutes les transactions sont traduites par des enregistrements détaillés, éventuellement stockés sur une base de données centrale.

Bagchi et White (2004) catégorisent l'information issue des transactions carte à puce en deux types : la transaction pécuniaire (monetary transaction), relative à la facturation des titres de transport, du lieu d'achat, du type de titre, etc. ainsi que la transaction de déplacement (journey transaction), qui nous intéressera plus particulièrement. Elle renferme des renseignements précis à propos des objets de transport impliqués. Dans cette optique, on peut isoler l'information selon que l'on veuille suivre les transactions d'un véhicule, d'un usager ou d'un chauffeur. Un exemple important étant qu'en pouvant suivre la carte d'un usager,

il est maintenant possible de lier deux déplacements, qui semblaient distincts, comme étant une correspondance.

Selon les mêmes auteurs (2005), les avantages des cartes à puce comportent :

- de grandes quantités de données de mobilité (échantillon)
- un lien entre données et individu (ou carte)
- un accès à des données continues sur de longs intervalles
- une connaissance des usagers les plus fréquents

Profitant de ces atouts, plusieurs travaux (Tranchant, 2005, Trépanier et al, 2005, Morency et al, 2006) provenant de l'École Polytechnique de Montréal portent sur les données opérationnelles carte à puce de la Société de Transport de l'Outaouais (STO), étudiées dans ce mémoire. Entre autres, Tranchant applique certains traitements et imputations aux données, soit :

- une reconstitution de données (voyages planifiés, séquences de transaction),
- la détermination des destinations (correction des transactions, notamment des arrêts, et dérivation des points de descente),
- une caractérisation in extenso des départs/tracés (profils de charge, diagrammes espace-temps, surface d'achalandage, matrice OD)

Cependant, la valeur des traitements effectués repose sur la validité et la cohérence de l'information amassée. En conséquence, au chapitre suivant, ces données carte à puce seront analysées de manière uni et multidimensionnelles, en faisant intervenir certaines logiques de validation afin de s'assurer que les données respectent bien leur cadre opérationnel. Ainsi, on s'assurera par exemple qu'un usager ne monte pas dans deux véhicules différents en même temps ou qu'un véhicule ne dépasse pas une certaine vitesse (dérivée) entre deux embarquements.

Avec ces règles de validations, fondées sur le fonctionnement réel d'un réseau de transport collectif, on sera à même de déceler certaines données erronées et éventuellement de les corriger. Bullock, Jiang et Stopher (2005) préviennent que des systèmes actifs (ici GPS ou carte à puce) où des interactions humaines sont nécessaires (ex. par le chauffeur), risquent d'entraîner des erreurs potentiellement importantes. Nous verrons que l'échantillon de données étudié sera grandement contaminé par ce genre de problématiques.

Systèmes de transport intelligents

L'intégration des technologies évoquées à des systèmes de communication et de traitement informationnel conduit à ce qu'on désigne couramment comme les systèmes de transport intelligents (STI, ou Intelligent Transportation Systems - ITS). Les STI présentent une tentative d'amélioration de la sécurité, de l'efficience et de la capacité des systèmes de transport de surface, à travers l'utilisation des technologies informationnelles et de communication (Miller et Shaw, 2001).

Les auteurs élaborent à l'effet que les systèmes STI reflètent un changement de paradigme à propos des systèmes de transport. En effet, au lieu d'augmenter physiquement la capacité des infrastructures, les STI ont pour objectif d'utiliser la capacité actuelle de manière plus efficace en acquérant des données détaillées de nature spatiale et temporelle concernant les systèmes de transport et d'utiliser cette information afin d'améliorer leur gestion (management). En effet, les systèmes STI permettent d'accumuler d'énormes quantités de données afin d'aider les planificateurs de transport à prendre de meilleures décisions et de manière plus éclairée (Smith et al, 2005).

D'une manière générale, la dernière taxonomie des technologies et applications STI, répertoriées par le consortium américain Advanced Public Transportation Systems (APTS), issu du Federal Transit Administration (FTA) fait état de trois catégories regroupant huit grandes orientations, illustrées à la Figure 1.6.

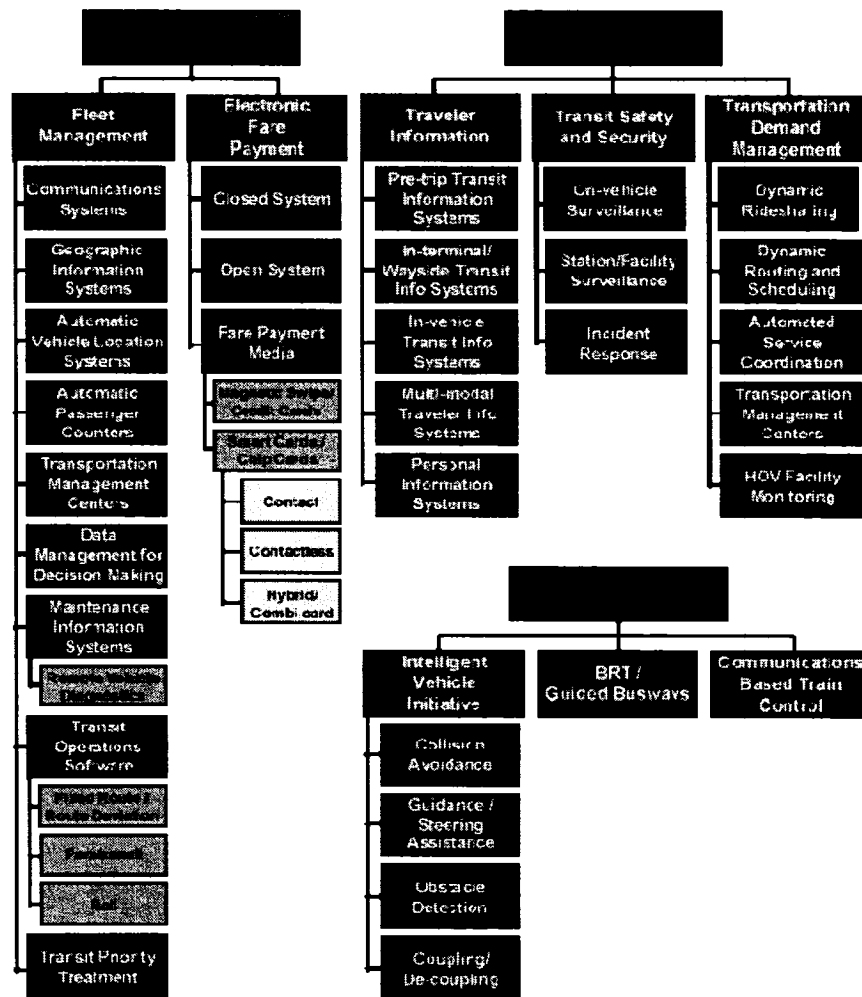


Figure 1.6 – Taxonomie des technologies STI répertoriées par le Advanced Public Transportation Systems (Source : <http://web.mitretrek.org/its/aptsmatrix.nsf/taxonomy>)

L'avènement des technologies présentées ultérieurement, toutes incluses sous la catégorie du déploiement opérationnel (schéma ci-dessus), a souvent été effectué dans une optique de répondre à un besoin informationnel traité en temps réel. Furth (2004) note en effet que le design des systèmes embarqués a été dominé principalement par les systèmes assistés d'information aux usagers en temps réel, se limitant à annoncer le prochain arrêt ou à en afficher l'heure d'arrivée prévue. À cet effet, Brydia et al (1998), se questionnant sur l'optimisation des données issues des STI, déplorent que les données issues de

ces systèmes aient souvent été supprimées, après avoir fait l'objet de traitement en temps réel.

Bien que cette situation s'améliore continuellement, les auteurs identifient cinq problématiques associées à l'utilisation des systèmes STI, soit :

1. La conservation des données
2. Le montage des bases de données associées
3. L'accès aux données
4. La différence entre données et information
5. La problématique reliée à la vie privée

Le quatrième point est particulièrement pertinent en ce qu'il met en garde contre une utilisation trop directe des données brutes. Trépanier et al (2005) parlent à cet égard de « traitement aveugle » des données. En effet, pour des données de type opérationnel (particulièrement issues de traces GPS ou de cartes à puce), un certain nombre de validations s'avèrent requis afin « d'assurer la validité et l'utilisabilité de l'information ».

Cependant, cette information peut grandement gagner en pertinence si elle passe à travers certains filtres représentant des contraintes physiques afin de débusquer certaines situations qui n'ont pas de sens dans le monde réel. Par exemple, une vitesse véhiculaire (dérivée) étant anormalement élevée entre deux positions dans l'espace-temps.

De surcroît, à l'intérieur d'un réseau de transport collectif, une série de validations supplémentaires s'ajoutent comme règles inhérentes à ce type de réseaux. Le chapitre suivant s'attarde à présenter ces classes de règles plus en détail et à présenter diverses anomalies rencontrées suite à leur application.

Pour l'instant, la prochaine section s'attarde de manière plus générale à la validation de l'information issue de systèmes STI.

1.2 Validité de l'information

Avant de procéder à l'analyse et aux traitements de données opérationnelles, cette section s'interroge sur les processus d'acquisition de l'information ainsi que la notion de validité de données.

1.2.1 Création d'une donnée

À la manière des récepteurs GPS évoqués plus tôt, les systèmes STI peuvent se diviser en deux catégories, passif ou actif. Les systèmes passifs engendrent de l'information indépendamment des événements et ne font qu'aligner des séries de données, souvent par intervalle temporel régulier. Bullock, Jiang et Stopher (2005) précisent que les fichiers de sorties résultants de systèmes d'acquisition de données passifs, procurent un flot continu de données spatiales et temporelles (i.e. coordonnées géographiques, heure et date), mais ne procurent aucune référence significative au système de transport (ex : la ligne en opération, le temps de départ, d'arrivée, changement de chauffeur).

En revanche, les systèmes actifs réagissent à des événements précis, telle l'arrivée à proximité d'une balise, l'ouverture/fermeture de portes, le démarrage/arrêt d'un moteur, etc. Furth et al (2004) identifient à cet égard une hiérarchie des niveaux de résolution lors d'une acquisition automatique de données. (Figure 1.7)

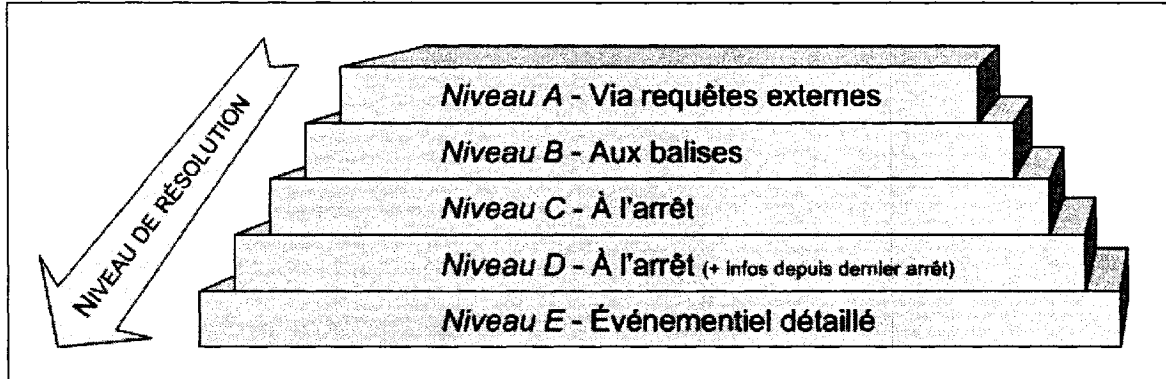


Figure 1.7 – Hiérarchie des niveaux de résolution spatiale des applications visant l'acquisition automatique de données (Furth et al, 2004)

Cette échelle commence au niveau A où les véhicules n'enregistrent (et/ou communiquent) l'information que lorsque réclamé par l'élément central. À l'opposé, au niveau E, on obtient un enregistrement d'une manière multiévénementielle où plusieurs déclencheurs sont combinés (souvent en comportant un élément fréquentiel) pour produire un maximum d'enregistrements. On pourrait qualifier ce dernier niveau de système hybride actif/passif car combinant à la fois les caractéristiques des deux profils.

Un exemple de ce type de système a été étudié dans le cadre d'un autre projet de recherche (évoqué dans Bergeron et Gagné, 2005) où les traces GPS d'une ligne de train de la région de Montréal étaient analysées. Le système embarqué émettait une position lorsqu'un de ces déclencheurs survenait :

- Le cap du 3 mi/h (4,8 km/h) est franchi (accélération ou décélération),
- 800 pi (244 mètres) ont été parcourus depuis la dernière position ou
- 25 secondes se sont écoulées depuis la dernière position.

Dans ce dernier cas, le système devenait passif dans la mesure où il n'y avait pas eu d'événement depuis les dernières 25 secondes.

La commodité d'une information liée aux éléments de transport pourrait rendre tentant l'enrichissement des données par des manœuvres manuelles (i.e. par le

chauffeur). Bullock et al (2005) mettent toutefois en garde contre ce genre de procédure car considérée non fiable. Dans leur environnement relativement stressé, les opérateurs risqueraient d'oublier le début/fin de leurs parcours ou l'arrivée/départ du garage. Nous verrons au chapitre suivant, que c'est exactement cela qui cause le plus grand nombre de transactions erronées lors de l'examen par validations des données opérationnelles étudiées.

Compression de données

Afin de condenser l'information obtenue de systèmes passifs, un traitement utilisant des techniques de filtrage ou synthèse peut être utilisé afin de compresser l'information en format événementiel. Cette technique suppose un système informationnel basé autour d'emplacements significatifs (i.e. arrêts, balises) où l'information des positions précédentes est résumée (i.e. vitesse moyenne, temps écoulé depuis arrêt précédent, etc.).

Un exemple de compression de données est donné par Bergeron et Gagné (2005) où la table des positions GPS d'un train est réduite en une table des arrivées en gare. Ainsi, pour un départ sur une ligne comportant entre 10 et 19 arrêts, un nombre équivalent d'enregistrements synthétise l'information originale, qui pouvait renfermer entre 300 et 1000 enregistrements.

1.2.2 Notion de validation

« [...] plusieurs efforts sont requis pour construire un ensemble de données structurées à partir d'un état informationnel chaotique [...] » (Chapleau et al, 2000). Dans le cas de données brutes issues de systèmes STI, même si elles témoignent de situations réelles, rien ne garantit qu'elles soient cohérentes avec la réalité. Des techniques de validation peuvent venir augmenter la qualité de ces données.

Elles renvoient ainsi à des processus stricts visant à identifier des données erronées, suspectes, voire manquantes. Les objets de transport étant soumis à

des limitations d'ordre physique et opérationnel, les données devront ainsi exister en respect de ces lois physiques et des règles opérationnelles en vigueur à l'intérieur d'un système de transport collectif.

L'application de ces règles permet de débusquer plusieurs données invraisemblables, qu'il est possible de catégoriser selon Burrough (cité par Miller et Shaw, 2001) concernant les données issues de systèmes STI, soit les erreurs :

- évidentes (positions avec coordonnées nulles, heure nulle, etc.),
- de mesure (parcours TC inexistant, vitesse dérivée trop élevée, etc.) et
- de traitement (durant l'application de méthodes)

La validation de données opérationnelles peut aussi passer par une conformité aux logiques d'acquisition de ces données, c'est-à-dire qu'une compréhension des processus de création de l'information peut mener à des validations complémentaires. Ces logiques d'acquisition seront développées au chapitre suivant.

Données manquantes

L'application de techniques de validation peut arriver à exposer certains trous dans la couverture informationnelle. Que ce soit dans le domaine de traces GPS ou de données de carte à puce, ces données manquantes sont habituellement causées par des défaillances matérielles, logicielles ou communicationnelles (Bagchi, 2003).

On a notamment mentionné les problèmes de positionnement des récepteurs GPS dans les environnements urbains denses ou souterrains. Sur le plan des données de carte à puces, Bagchi (2003) relate plusieurs causes de transactions manquantes :

- un bris matériel (carte ou borne);
- une suite de transactions qui n'est pas acheminée au serveur central;

- un oubli de l'usager et/ou négligence du chauffeur;
- une transaction effectuée trop vite (carte retirée avant la fin de la transaction)

Du point de vue de l'analyse de la demande, on doit aussi prendre en compte que l'échantillon n'est pas complet étant donné que certains usagers paient comptant et n'utilisent pas de carte à puce. De plus, lorsque seul le numéro de série des cartes est utilisé afin de suivre les déplacements des usagers, la perte/remplacement d'une carte entraîne une rupture du rythme de déplacement de l'usager concerné, qui sera perçu comme deux utilisateurs distincts.

Manutention des erreurs

L'examen des données à travers des filtres de validation donne aussi lieu à la découverte de données erronées ou suspectes. Étant donné que ces enregistrements peuvent ne comporter qu'un seul attribut douteux, ils ne sont pas nécessairement supprimés mais préférablement identifiés et possiblement corrigés. À cet effet, plusieurs techniques d'imputation et de correction des données issues de systèmes STI existent. Des exemples relatifs à des positionnements GPS et des transactions carte à puce sont exposés aux sections 1.3 et 1.5.

1.2.3 Assistance par ordinateur

Les enquêtes transport utilisent déjà une série de normes strictes visant à valider l'information recueillie afin de garantir une certaine qualité des données. Bien que ce processus puisse se faire via le partage de l'espace cognitif par le répondant et l'analyste (concordance du vocabulaire utilisé), il peut s'avérer subjectif et relativement lent. C'est pourquoi on a développé de nombreuses méthodes d'assistance informatique afin de seconder et valider la saisie d'information (un éventail est présenté dans Greaves, 2004).

Dans le cas de la grande région de Montréal, on utilise un type d'enquête à base d'entrevues téléphoniques assistées par ordinateur (Computer Assisted Telephone Interview - CATI). L'intervieweur utilise un questionnaire intelligent qui, parmi d'autres fonctionnalités, géocode en temps réel tous les emplacements (origines, destinations, jonctions) et valide interactivement les déplacements en terme d'accès, de distances parcourues, de temps de parcours, de l'heure de départ et des correspondances effectuées (Chapleau, 2003). Ce système de validation des réponses inclut la validation des déplacements sur le réseau de transport en commun.

Dans cette optique, la validation des données issues de systèmes STI assure une cohérence à une représentation virtuelle de situations réelles; ces données devant respecter certaines contraintes imposées par les lois de la physique et par les objets du réseau de transport.

1.3 Traitement et imputation

Par un traitement se basant ou non sur l'application de logiques spatio-temporelles, de l'information nouvelle (imputation) peut souvent être dérivée à partir de données brutes. Malgré que toute imputation implique nécessairement une probabilité et demeure ainsi contestable, cette information déduite peut agir comme supplément ou remplacer certains éléments erronés, suspects ou carrément manquants. À cet égard, une revue des méthodes d'imputation (simples et multiples) est d'ailleurs exposée par Ni et al (2005).

Au-delà des traitements en temps réels où les objectifs concernent majoritairement la localisation de véhicules ou l'information à la clientèle, les prochains paragraphes exposent plusieurs articles rencontrés où une utilisation a posteriori de données opérationnelles est utilisée, souvent dans une optique d'aide à la planification.

1.3.1 Segmentation de traces GPS

Plusieurs études portent sur la décomposition des positions provenant de récepteurs GPS passifs. Entre autres, Stopher et al (2003) utilisent une technique qui fragmente ces longues séries de positions en déplacements distincts en utilisant des règles identifiant les terminaisons des déplacements. Dans leur cas, l'hypothèse retenue est qu'un déplacement se termine après deux minutes d'immobilisation (vitesse nulle, distance parcourue nulle) du véhicule, équipé de ce type de récepteur. Pour leur part, Quiroga et Bullock (1999) réduisent plutôt l'information des positions GPS en segments caractérisés, où chaque segment fait partie d'un réseau de vecteurs bidirectionnels (délimités par des discontinuités physiques) couvrant le réseau routier.

1.3.2 Caractérisation des déplacements

De nombreuses études visent à utiliser des récepteurs GPS afin de confirmer ou bonifier certains résultats issus d'enquêtes transport. Murakami et Wagner (1999) furent parmi les premiers à s'intéresser à la comparaison de données issues d'un véhicule personnel équipé d'un récepteur GPS avec une enquête de type auto-rapportées, en comparant les résultats (nombre de déplacements, heures, durée, distance parcourue).

Dans la même veine, Wolf (2004) présente une série de statistiques à propos des ratios des déplacements sous-rapportés en comparant avec des données GPS. Yalamanchili, Pendyala et al (1999) s'intéressent quant à eux aux déplacements à motif autre que le travail (souvent non recensés) qui semblent constituer une part de plus en plus importante des déplacements quotidiens. Étant donné que les sondés peuvent facilement oublier certains de ces petits trajets, les auteurs ont utilisé des récepteurs GPS afin de comparer les réponses fournies aux données. Ainsi, celles-ci ont fourni plus du double de chaînes à multiples déplacements que celles rapportées par les sondés, dans le contexte d'une période de temps d'une journée.

1.3.3 Taille de l'échantillon

La littérature comporte plusieurs études utilisant des données acquises automatiquement mais elles se limitent souvent à des petits ensembles. On constate de plus en plus de recherches qui attaquent de gros ensembles de données. Wolf et al (2004) ont entrepris d'étudier les données GPS sur une longue période en utilisant 186 voitures personnelles équipées de récepteurs passifs, enregistrant lorsque le moteur tourne. L'échantillon résultant comporte près de 50 000 véhicules-jour fournissant plus de 240 000 déplacements (à l'intérieur de la zone d'étude).

Les données de positionnement GPS de trains de l'étude déjà évoquée (Bergeron et Gagné, 2005) consistent aussi en un grand bassin informationnel. En effet, pour la ligne de 64 km, un jour de semaine génère environ 10 000 positions quotidiennes (220 000 par mois). Dans le cas des données de carte à puce étudiées dans cette recherche, nous verrons qu'un mois de données contient entre 400 000 et 800 000 transactions (déplacements non-liés).

1.3.4 Imputation du mode

Une méthode de reconstruction des déplacements est documentée par Chung et Shalaby (2005), qui en plus de dériver (par map-matching) les liens empruntés par les individus portant un récepteur GPS personnel, proposent une dérivation du mode de transport utilisé, uniquement à l'aide des traces GPS. Les modes possibles étant la marche, le vélo, l'automobile et le bus.

Tous les déplacements sont classés selon l'une des combinaisons de mode suivantes :

- Marche seulement
- Marche ► Vélo ► Marche
- Marche ► Automobile ► Marche

- Marche ► Bus ► Marche

Après avoir segmenté les déplacements en différents modes par les biais d'intervalles de vitesses, on utilise des règles de validation spatio-temporelles afin de départager les modes utilisés.

Parmi ces règles, on prend en considération le type de lien emprunté (i.e. un vélo ou un piéton n'emprunte pas d'autoroute mais peut emprunter un sens-unique à contresens), les comportements d'attente (autour des arrêts de bus, géo-référencés), et des comportements routiers (i.e. les bus ont des arrêts intermittents et des trajectoires précises).

En comparant avec les comportements réels et documentés, 78,5 % des liens furent correctement segmentés. Les liens mal segmentés le sont principalement à cause d'une trop petite distance parcourue sur ce lien ou une piètre qualité des positions GPS (canyons urbains, temps d'initialisation des récepteurs entre 15 sec et 5 min). Une grande proportion des liens mal isolés appartient à des bus, où le toit du véhicule peut interférer avec les signaux GPS lorsque l'utilisateur se tient loin des fenêtres du véhicule.

Concernant l'identification du mode utilisé, un taux de succès de 91,7 % est rapporté. Les erreurs sont principalement causées par une insuffisance concernant les règles de validation ainsi qu'à des erreurs relatives à la carte GIS de référence. Face à ces erreurs, les auteurs proposent entre autres de considérer les accélérations entre les positions GPS ainsi que de relier les chaînes de position brisées par des estimations basées sur des algorithmes de plus court chemin.

1.3.5 Imputation du motif

Alors que la plupart des études impliquant des données GPS les utilisent comme bonification à d'autres sources d'information, celle de Wolf et al (2001) décrit

l'imputation des motifs de déplacement uniquement à l'aide d'une analyse a posteriori de ce type de données.

Cette étude implique ainsi des voitures particulières équipées de récepteurs GPS passifs reliés à des mini-ordinateurs de poche, emmagasinant les coordonnées lorsque la voiture est en marche. Parallèlement, les répondants devaient noter les motifs de leurs déplacements pour une période de trois jours.

Après une première étape où l'on isole chaque déplacement effectué, on emploie un inventaire géo-référencé d'utilisation du sol pour en dériver le motif. Cet inventaire comporte plusieurs attributs tels le type de lot, ses dimensions physiques, son zonage, etc. Pour chaque type d'utilisation du sol est attribué une vocation principale et dans certains cas, une vocation secondaire et tertiaire. Le Tableau 1.1 illustre cette classification.

**Tableau 1.1 – Motifs de déplacement possibles selon utilisation du sol (Wolf et al, 2001).
Le code « 99 » équivaut à une utilisation mixte, ambiguë ou inconnue.**

Land Use Description	Purpose 1	Purpose 2	Purpose 3
Airport	Drop Off / Pick Up	Change Mode	Go to Work
Bank	Personal Business		
College or University	Go to Work	Go to School	Drop Off / Pick Up
Commercial	99		
Convenience Store / Gas	Conv Store / Gas		
High School	Drop Off / Pick Up	Go to School	
Improved Gov't Exempt	99		
Residence, Multifamily	Return Home	Social / Recreation	Drop Off / Pick Up
Office Building	Personal Business	Go to Work	
Shopping Mall (Deck)	Shop		
Post Office	Personal Business		
Religious, Church	Social / Recreation		
Residence, One Family	Return Home	Social / Recreation	Drop Off / Pick Up
Restaurant	Eat		
Restaurant - Fast Food	Eat		
Retail, Multiple Occupancy	99		
Retail, Single Occupancy	Shop		
Shopping Center	99		
Strip Mall	99		
Supermarket	Shop		
Unknown - on Road	99		
Vacant Exempt Land	99		
Vacant Land / Railroad	99		
Vacant Lot, Commercial	99		
Warehouse	99		

Les coordonnées de la destination de chaque déplacement sont attribuées à une région de l'inventaire. Ainsi, un type d'activité est dérivé pour chaque déplacement à l'aide de logiques temporelles et du déplacement précédent.

Après épuration des déplacements problématiques et des participants ayant connu d'importantes difficultés techniques, les motifs des 151 déplacements restants furent comparés aux renseignements fournis par les répondants. On note que 7 % des motifs furent incorrectement dérivés. Tous ces cas sont le résultat d'erreurs d'assignation territoriale (erreur de position GPS, information sur utilisation du sol manquante ou incomplète, etc.). Notons que plusieurs erreurs de positionnements auraient été évitées si l'étude avait été effectuée après l'élimination du « brouillage sélectif » du système GPS.

Selon les auteurs, l'étude comporte plusieurs désavantages (les déplacements sont limités à un véhicule et non à un individu, l'information sur les passagers est inexistante) et des améliorations potentielles (ajout de pré-validations des données, meilleur algorithme de détection des déplacements, positionnement GPS non brouillé). Précisons qu'une part importante des processus évoqués tient de manœuvres manuelles.

Une étude subséquente (Wolf et al, 2004) tente d'automatiser au maximum ce type de procédé avec un ensemble de données nettement supérieur (50 000 véhicules-jour, 240 000 déplacements). Afin d'identifier les destinations, on utilise une technique d'agrégation de points (clustering) sur les dernières positions avant l'arrêt du moteur, puis on associe le centre d'agrégat à un point d'intérêt issu d'une carte d'utilisation du sol, d'une manière similaire à l'étude de 2001. Ici, on assigne des probabilités aux motifs selon la proximité au point de destination (max 300 m).

Finalement, connaissant diverses caractéristiques des utilisateurs des véhicules équipés de GPS, les données d'une enquête transport sont utilisées afin de perfectionner l'association des motifs selon la pondération de plusieurs

paramètres (genre, statut, possession automobile, type de jour, heure et durée du déplacement, etc.)

Les auteurs concluent à des résultats cohérents relativement à l'analyse des comportements de mobilité, malgré un potentiel de renforcement des données d'utilisation du sol ainsi que l'absence de détection de nombreux petits déplacements où le moteur n'a pas été arrêté.

1.3.6 Qualité du service rendu

Dans un article récent, Bullock et al (2005) utilisent les données provenant de récepteurs GPS passifs ainsi que des heures d'arrivée planifiées aux points de contrôle le long des parcours, afin de vérifier l'adhérence des véhicules à l'horaire. Une segmentation des traces GPS est faite selon le type de trajectoire effectuée (parcours d'une ligne, sortie d'un garage, retour à un garage, garage à garage, indéfinie). Lorsqu'un parcours de ligne est identifié, les temps de parcours du véhicule entre les points de contrôle sont alors comparés avec les temps de parcours planifiés.

Une étude similaire de Hammerle et al (2005) utilise des véhicules équipés d'un système de positionnement actif sur le parcours d'une ligne de 13,7 km (8,5 mi) avec points de contrôle. L'analyse porte sur les positions enregistrées lors des événements suivants : passage à un arrêt (avec immobilisation ou non) et passage à un point de contrôle. Des statistiques sur l'adhérence à l'horaire et la régularité des créneaux sont présentées.

1.3.7 Comportement de la clientèle

L'élément le plus singulier de données provenant de systèmes de perception électronique est que l'on puisse suivre les déplacements d'un individu, à travers l'usage qu'il fait de sa carte. À cet égard, Hofmann et O'Mahony (2005a) effectuent une analyse des correspondances réalisées où l'arrêt de descente du premier déplacement est déduit. Les auteurs étudient plusieurs aspects dont

l'intervalle temporel entre les deux transactions des correspondances, le temps d'attente ainsi qu'une analyse spatiale des transferts, agrégée par zone.

Dans le même ordre d'idées, Okamura et al (2004) analysent les données provenant d'un système où des cartes magnétiques (de type porte-monnaie) sont utilisées par la clientèle afin de payer plus rapidement leur passage sur le réseau de la ville d'Hiroshima (multi-opérateurs). Malgré le fait que les cartes ne soient ni personnalisées ou rechargeables, on peut néanmoins arriver à suivre les usagers sur de courtes périodes et mesurer certains indicateurs telles les habitudes comportementales relatives à l'utilisation de leurs cartes, procéder à des analyses du temps d'attente ainsi que des comparaisons d'indicateurs en un nœud du réseau. On traite aussi d'une classification des types de client selon les rythmes d'activité. Morency et al (2006) exploitent d'ailleurs ce sujet avec les mêmes données que celles étudiées dans cette recherche.

1.3.8 Développement d'une matrice OD

À l'aide de données de tourniquets équipés de lecteurs magnétiques, Barry et al (2002) décrivent une méthodologie afin de dériver une matrice origine-destination entre les stations du métro de New York. L'information transférée à un serveur central inclut l'heure d'entrée sur le réseau, le numéro du tourniquet utilisé ainsi que le numéro de série de la carte de l'utilisateur.

Afin de valider et d'isoler les déplacements des individus, la méthodologie est basée sur deux hypothèses :

- La station de sortie d'un usager sera la même que la station d'entrée de cet usager lors du prochain déplacement.
- La dernière station de sortie (de la journée) d'un usager sera la même que la première station d'entrée (de la journée) de cet usager.

À la manière des données de carte à puce, l'information concernant la sortie du réseau est inconnue. En revanche, la plupart des tourniquets de sortie étant

munis de compteurs, les quantités d'entrants/sortants peuvent être balancées pour chaque station. Des matrices origine-destination ont pu être édifiées sur la base de ces résultats.

Utilisant une méthodologie similaire, Tranchant (2005) utilise des données carte à puce afin de déterminer la matrice origine-destination d'une ligne de bus pour un jour donné. En effet, en isolant les transactions des usagers concernés, il détermine les arrêts de destination de chaque montée à l'aide de l'arrêt d'embarquement de la transaction suivante de l'usager ainsi que de la régularité de ses transactions. S'en suit le montage des profils de charge et des matrices OD associées. D'autres travaux concernant des données carte à puce seront exposés en fin de chapitre.

1.4 Cadre et contraintes spatio-temporelles

D'une manière générale, la mobilité des individus est limitée par des contraintes de temps et d'espace physique. Cependant, dans un espace informationnel, il est possible que les données, qui représentent pourtant des situations réelles, ne respectent pas ces limites. Il importe donc de les valider par les logiques spatio-temporelles qui s'appliquent aux objets de transport afin de confirmer (ou d'infirmer) leur cohésion.

Cette section présente brièvement les concepts de la géographie temporelle et la théorie des activités qui utilisent certaines contraintes spatio-temporelles évoquées.

1.4.1 Géographie temporelle

Un individu doit moyenner du temps en échange d'une certaine distance via le mouvement (ou la communication) afin de participer à ses activités (Pred, cité par Miller, 2004). Cette affirmation représente bien l'interrelation entre le temps et l'espace, omniprésente dans tous nos déplacements, et qui fournit le concept

de base de la géographie temporelle. Développée dans les années 50 et 60 par Hägerstrand, elle met l'accent sur les contraintes spatio-temporelles des déplacements, c'est-à-dire comprendre ce qu'une personne ne peut pas faire au lieu d'essayer de comprendre ce qu'elle va faire (Miller, 2004).

Comme préambule aux prochains paragraphes, la Figure 1.8 expose un aquarium spatio-temporel où les déplacements d'un ménage sur 24 heures sont présentés. Ce type de représentation, avec le temps sur l'axe vertical, met en évidence les cheminements dans l'espace-temps des membres d'un ménage (ici trois) à l'intérieur d'une certaine zone géographique.

En faisant intervenir des contraintes, on peut s'imaginer un individu devant quitter un emplacement A au temps i avec une vitesse constante V , on peut représenter l'éventail des cheminements spatio-temporels possibles (son espace accessible) par un cône débutant à cette origine et s'ouvrant à mesure que le temps progresse, cet espace étant contraint par sa vitesse. Si cet individu doit se rendre au point B avant le temps j , c'est le même type de cône (mais inversé) qui représentera son espace accessible afin d'arriver à l'heure. Une intersection de ces deux espaces contraints résultera en un espace fermé représentant l'éventail des cheminements spatio-temporels possibles selon les contraintes de temps, d'espace et de vitesse.

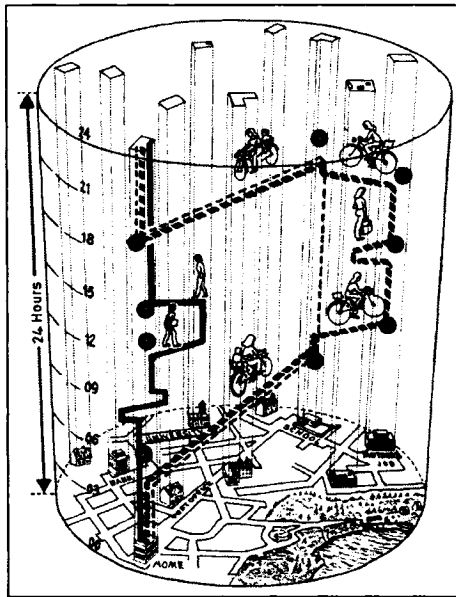


Figure 1.8 – Cheminements spatio-temporels sur 24 heures des éléments d'un ménage (tiré de présentation de Kwan, 2002)

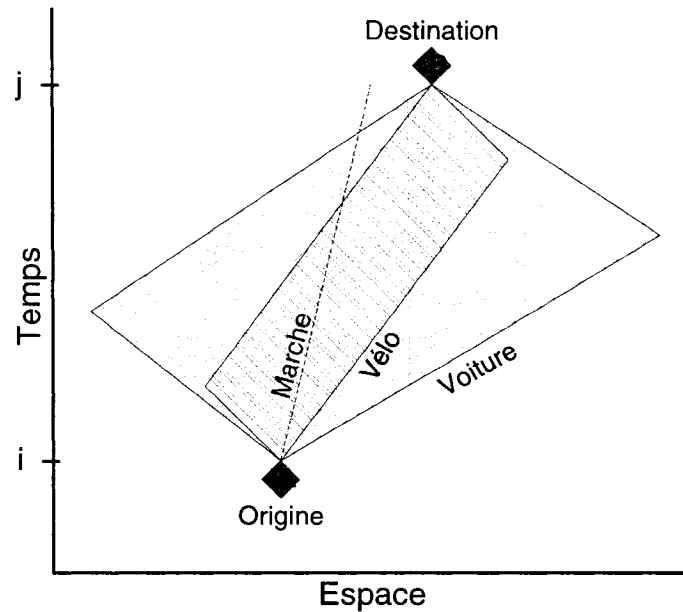


Figure 1.9 – Prismes spatio-temporels selon mode, entre une origine et une destination

Si on linéarise la dimension spatiale, on retrouve un diagramme bidimensionnel (Figure 1.9) mettant de l'avant ce type de contraintes physiques, représentées par des prismes spatio-temporels, qui encadrent les limites spatiales atteignables par un déplacement entre deux points en un temps donné, ici selon différents modes (différentes vitesses). Ainsi, dans la figure, un individu devant se rendre d'un point A (origine) à partir du temps i à un point B (destination) avant le temps j , aura plus de portée s'il prend sa voiture que s'il prend son vélo (vitesse moins élevée). Ainsi, le prisme spatio-temporel du vélo sera plus limité, du fait qu'il dispose de moins de temps et d'espace s'il veut arriver à l'heure. Quant à l'option où l'individu marcherait, il ne serait pas arrivé au temps j , sa vitesse V_{marche} étant inférieure à la vitesse minimale requise, $\frac{B-A}{j-i}$.

1.4.2 Théorie des activités

Contrairement à l'analyse classique de distribution des déplacements qui prend comme hypothèse qu'un déplacement est une entité indépendante (Kitamura, 2004), la théorie des activités met plutôt de l'avant une approche de planification basée sur les activités des individus où, notamment, l'origine spatio-temporelle d'une activité est dépendante de la destination de l'activité précédente.

Cette théorie voit les patrons d'activité comme tendances comportementales de la population à l'intérieur de sous-systèmes tels le ménage, l'institution ou l'entreprise. Au lieu de percevoir les séquences comportementales des déplacements uniquement comme un phénomène de demande, la consommation d'une activité est aussi dépendante d'une considération de l'offre disponible (Chaplin, cité par Harvey, 2004).

La théorie des activités se trouve en lien avec la géographie temporelle en termes des contraintes imposées par les activités des individus dans l'espace-temps. Harvey (2004) rappelle la catégorisation des contraintes envisagée par Hägerstrand, soit

- les contraintes de capacité (biologiques ou matérielles d'un individu),
- les contraintes de jonction (définissant le lieu, l'instant et la durée d'une activité) et
- les contraintes de tutelle (tierce personne ou groupe, i.e. heures d'ouverture, congestion, etc.)

Certaines enquêtes de mobilité s'appuient sur des modèles d'activité basés sur ces types de contraintes. Ainsi, au lieu d'interroger les individus sur leurs déplacements, les enquêtes d'activité se préoccupent plutôt des activités qui génèrent ces déplacements.

1.4.3 Enrichissement par données STI

Avec une acquisition automatique de données géolocalisées décrivant des trajectoires spatio-temporelles, les itinéraires de déplacements gagnent considérablement en terme de quantité (nombre de données) et de qualité (précision). Ces percées technologiques permettent d'améliorer l'acquisition de données des systèmes d'activité (Miller, 2005).

"It is clear that the development of full-fledged activity-based models [...] calls for the collection, synthesis, organization, manipulation, integration, visualization, analysis, and modeling of detailed data about people's activity-travel patterns by time of day, mode and route choices, activity-sequencing and trip chaining, spatio-temporal constraints and opportunities, and interactions in time and space." (Pendyala, 2005)

L'auteur identifie de nombreux collègues qui s'appliquent à l'exploration des patrons d'activité via le forage de données (data mining) ainsi que par diverses techniques de visualisation. Parmi ces auteurs, Kwan précise qu'avec le grand nombre d'attributs disponibles afin de caractériser les patrons d'activité et étant donné la capacité de la visualisation scientifique à rendre compte de ces attributs, la visualisation incarne une direction prometteuse pour l'exploration et l'analyse de larges et complexes ensembles de données d'activité-déplacement (Kwan, 2000). De plus, cette visualisation peut être interactive sur le plan de l'angle de vision, des attributs présentés, etc.

Enfin, ces techniques de visualisation permettent de conserver la complexité des données originales tout en les rendant accessibles à l'œil humain. Comme exemple de ce type de visualisation, la Figure 1.10 tente de montrer que les espaces d'activité des femmes afro-américaines de l'échantillon sont plus restreints que ceux des individus des autres groupes ethniques ou de genre (non illustrés).

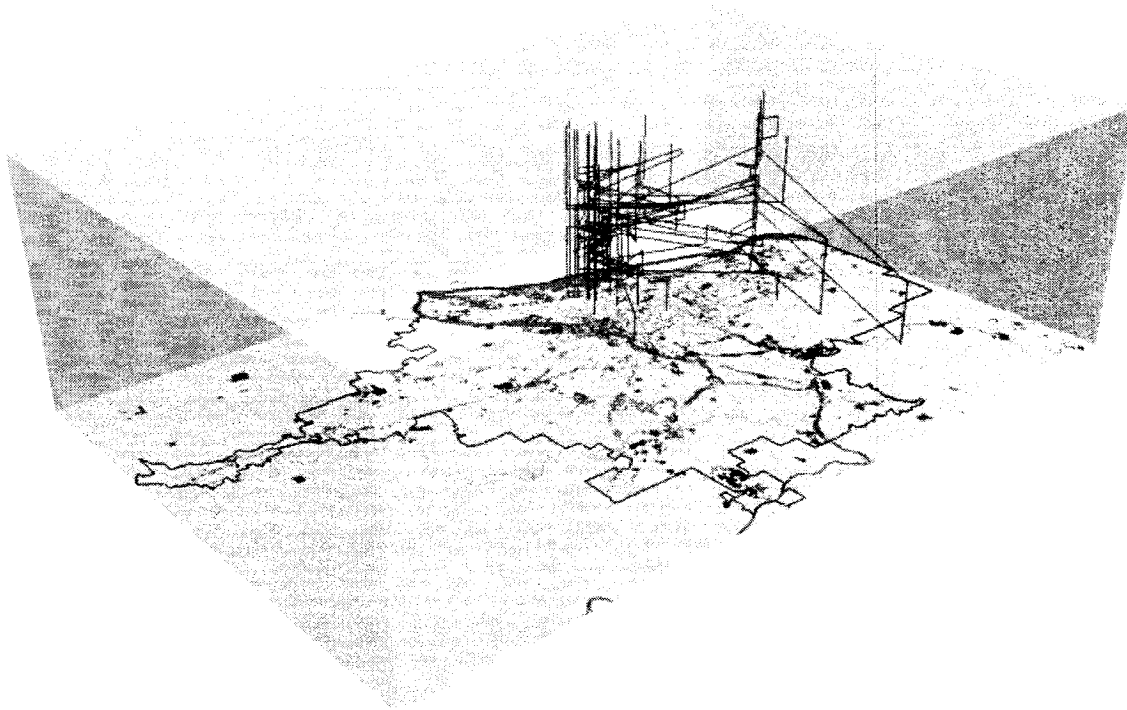


Figure 1.10 – Cheminements dans l'espace-temps des femmes afro-américaines, provenant des données de l'enquête « Activity-Travel » de Portland de 1994-95 (Source : Kwan, <http://geog-www.sbs.ohio-state.edu/faculty/mkwan/Gallery/STPaths.htm>)

1.4.4 Application au transport collectif

En raison de la structure de son service, le transport collectif n'offre pas la même indépendance de déplacement que la voiture personnelle, le vélo ou la marche. En effet, pour un individu devant se rendre d'un point A au point B entre les temps i et j en transport collectif, ses déplacements sont limités aux départs offerts (Figure 1.11).

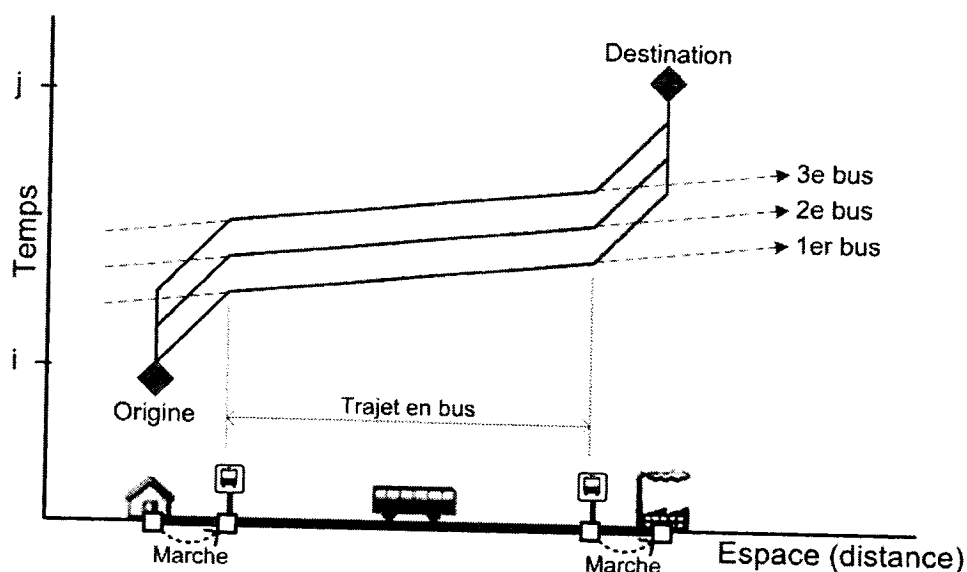


Figure 1.11 – Cheminements espace-temps possibles afin de se rendre d'un point A au point B entre les temps i et j, en utilisant un service de transport collectif

Lorsqu'on applique ce type de contrainte aux prismes spatio-temporels, ceux-ci se retrouvent discrétisés étant donné le nombre fini d'éléments (véhicules-tracé-sens-départ) accessibles par l'utilisateur. Ainsi, en considérant plus d'un parcours, on peut concevoir une représentation des prismes où les cadres sont associées à ces contraintes. Dans le schéma de la Figure 1.12, plusieurs possibilités de déplacements s'offrent à l'utilisateur afin de se rendre d'un point A au point B entre les temps i et j.

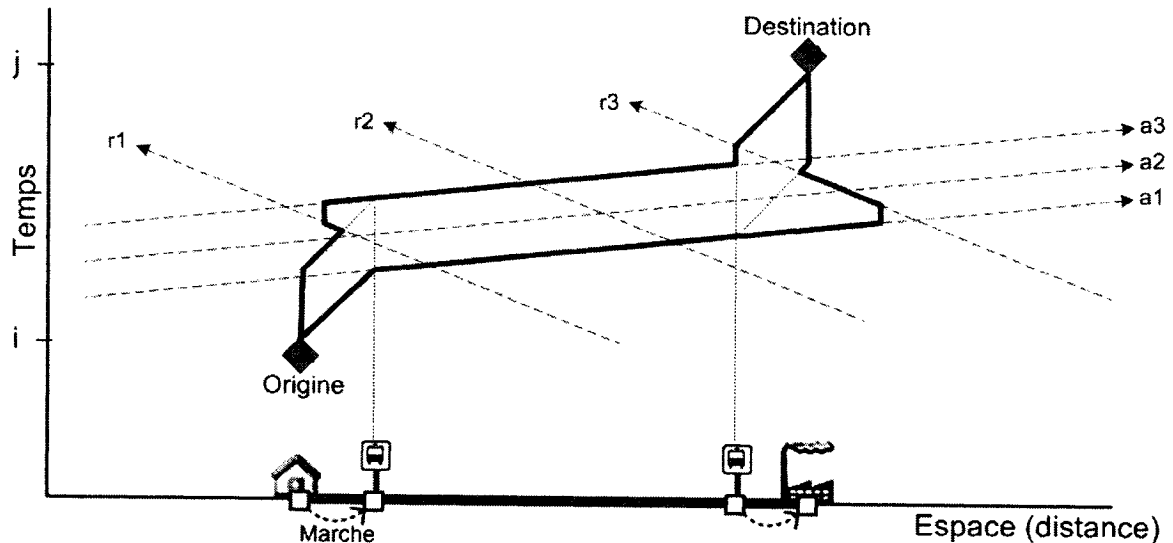


Figure 1.12 – Prisme spatio-temporel délimitant l'arbre de potentialités discret associé à un réseau de transport collectif

Ici, l'arbre limité des possibilités se situe à l'intérieur des deux trajectoires extrêmes, soit :

- Marche à l'arrêt ► Bus a1 ► Transfer ► Bus r3 ► Marche à la dest.
- Marche à l'arrêt ► Bus r1 ► Transfer ► Bus a3 ► Marche à la dest.

Notons que cette représentation s'avère considérablement simplifiée par la linéarisation de la dimension spatiale, la représentation de seulement deux parcours et la non-représentation de l'ensemble des arrêts de transfert. D'autres contraintes pourraient être considérées telles la non-uniformité des vitesses de parcours, l'attente due aux feux de circulation, la congestion, les perturbations de service, etc.

1.4.5 Cartes d'accessibilité

On peut représenter de manière plus intuitive les contraintes liées au transport collectif à travers des cartes géographiques. Miller (2004) illustre ces contraintes par des arbres de cheminements potentiels depuis une origine fixe. Ils représentent les nœuds accessibles, par réseau TC et par marche, à l'intérieur

d'un certain laps de temps. Une approche par lignes isochrones peut aussi être utilisée (O'Sullivan et al, 2000) où, toujours depuis une origine fixe, on représente les temps d'accès sur le réseau par des lignes de contour. Chaque ligne expose des points de destination à égale distance temporelle.

À la manière de cônes spatio-temporels, on peut projeter l'intersection de cônes isochrones sur un espace géographique (plan X-Y, Figure 1.13), afin d'identifier la zone d'accessibilité lors d'un déplacement entre deux points.

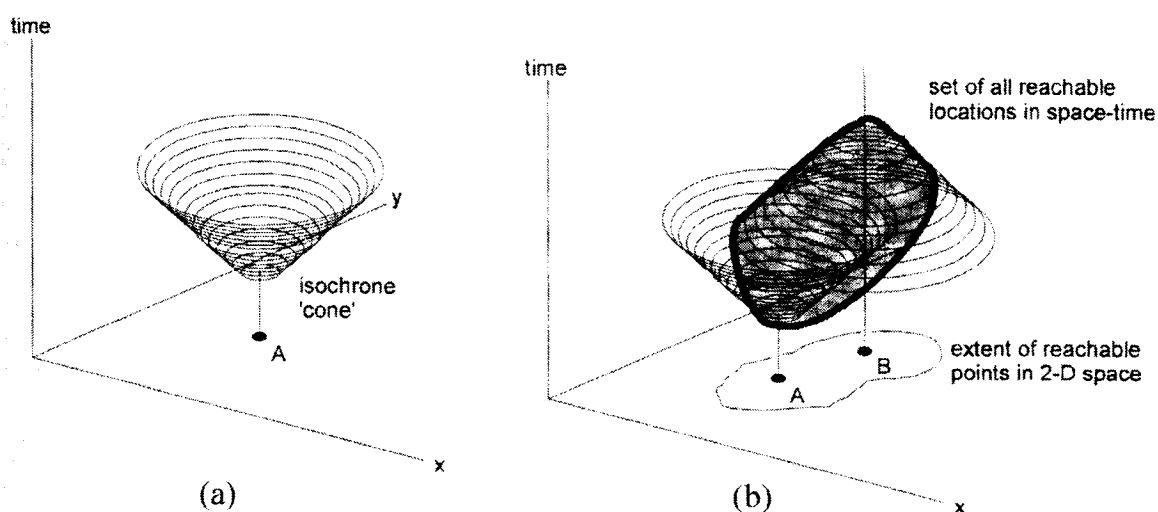


Figure 1.13 – Prismes spatio-temporels tridimensionnels (Source: O'Sullivan, 2000)

En projetant les intersections entre les différentes lignes de contour, on conçoit des cartes d'accessibilité où les contraintes peuvent être définies selon les fréquences de passage (O'Sullivan et al, 2000) ou par l'horaire planifié (Church et al, 2005). Les figures suivantes illustrent quelques applications utilisant les contraintes de temps d'accès issues de réseaux de transport collectif.

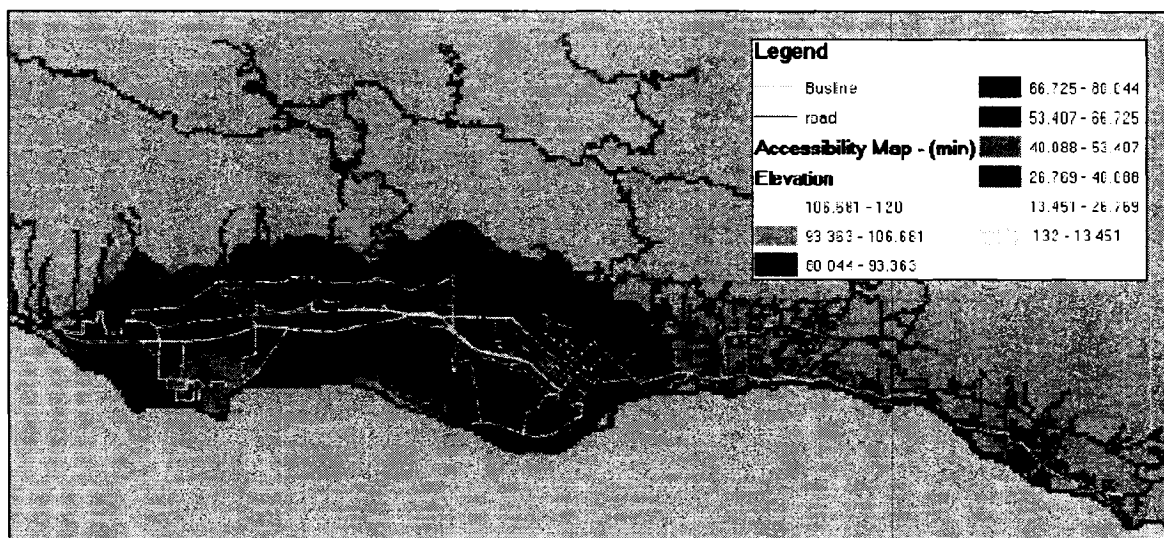


Figure 1.14 – Carte d'accessibilité TC du campus de l'UCSB (à gauche) et du centre-ville (au centre) afin d'arriver à 8h00 (Source : Church et al, 2005)

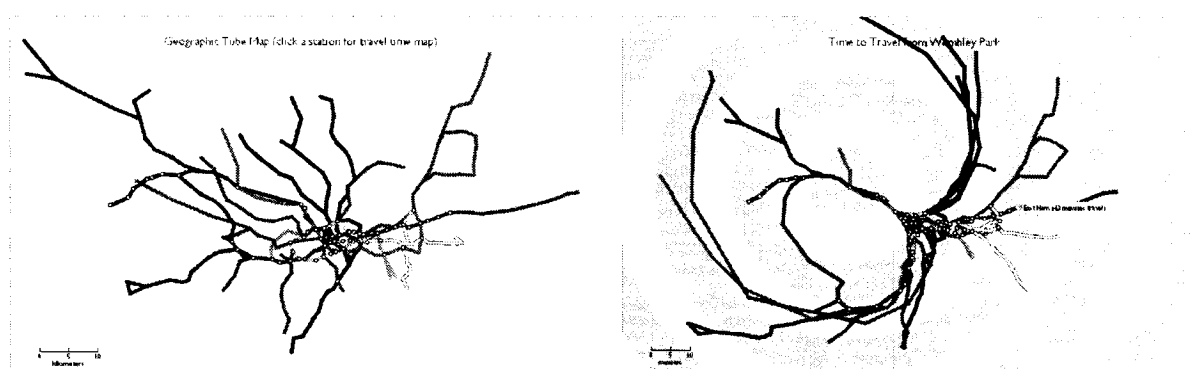


Figure 1.15 – Représentation géographique (à gauche) du réseau de métro de Londres, puis réorganisée selon les temps de parcours depuis la station Wembley Park (à droite) (Source interactive : http://www.tom-carden.co.uk/p5/tube_map_travel_times/applet)

1.5 Régularité et données carte à puce

Afin de réduire la dimensionnalité de l'information liée à la régularité des chaînes de déplacement et de la représenter sous une forme plus accessible, Kitamura (2004) évoque plusieurs recherches procédant à des regroupements de caractéristiques de mobilité (dispersion des destinations, la fréquence de

déplacement, etc.) ainsi que certains indices tels des mesures de centralité (Horton et Hultquist), mesure de circuiterie (Kansky), etc. Cependant, d'autres auteurs (Hanson et Huff, Pas) sont cités à l'effet que la plupart de ces travaux se basaient sur des données d'enquêtes où un seul jour de déplacement était considéré afin de dériver ces mesures.

Avec les données automatisées issues de systèmes de perception automatique, notamment ceux de type carte à puce, le suivi peut désormais se faire sur de longues périodes. Au-delà des considérations reliées à la fraude et à la sécurité des transactions, un potentiel énorme est relié à ce type de données puisqu'il permet d'effectuer un suivi sur plusieurs objets de transport, notamment les transactions associées à la carte d'un usager afin d'en étudier sa mobilité.

De plus, lorsque couplés à un système AVL, les systèmes de type carte à puce sont enrichis d'une information à dimension spatiale, multipliant les potentialités d'analyse. Les prochains paragraphes retracent l'état de la recherche associée à l'étude de données opérationnelles carte à puce.

1.5.1 État des recherches

« [...], l'approche totalement désagrégée s'est peu penchée sur les données à caractère régulier et cyclique car les analyses touchent habituellement le concept de "jour moyen de semaine" obtenu des enquêtes origine-destination. » (Trépanier et al, 2005)

Domaine en expansion, la littérature sur l'apport des données carte à puce à des fins opérationnelles est limitée à quelques auteurs. Notons l'apport particulier de Paula Bagchi (Bagchi, 2003, Bagchi et White, 2004, 2005) qui en a fait l'objet de sa thèse de doctorat et qui poursuit ses recherches pour une firme privée.

Bien qu'abordant des problématiques plus générales quant à l'usage de la carte à puce dans les sociétés de transport collectif, la thèse de Bagchi (2003) procède aux analyses quantitatives suivantes :

- calcul des taux de déplacement,

- calcul des taux de déplacements liés (correspondances),
- détermination de profils de jours avec déplacement et
- segmentation de la clientèle

Plusieurs travaux (Tranchant, 2005, Trépanier et al, 2005, Morency et al, 2006) provenant de l'École Polytechnique de Montréal portent sur les données étudiées dans cette recherche. À l'aide de techniques de validation et de la régularité des habitudes de la clientèle, Tranchant applique certains traitements et imputations aux données carte à puce, notamment :

- une reconstitution de données (voyages planifiés, séquences de transaction),
- la détermination des destinations (correction des transactions, notamment des arrêts, et dérivation des points de descente),
- une caractérisation in extenso des départs/tracés (profils de charge, diagrammes espace-temps, surface d'achalandage, matrice OD)

Pour leur part, Morency et al (2006) proposent une étude des rythmes d'activité à la lumière de ces mêmes données.

CHAPITRE 2. ÉVALUATION DES DONNÉES, VALIDATION ET LOGIQUES

Un système de transport en commun renferme certaines règles et régularités faisant en sorte qu'il est organisé et structuré. Par contre, à l'intérieur de bases de données sensées représenter un tel système, rien ne garantit que ces règles soient respectées. Une analyse de ce type d'ensembles informationnels est présentée à l'aide de certains filtres et validations afin d'authentifier les données en tant que comportements réels, car mus par des objets physiques réels.

Avant de procéder à l'élaboration de ces validations et de leur application sur des données carte à puce, ce chapitre présente d'abord brièvement la Société de Transport de l'Outaouais, d'où proviennent les données étudiées, ainsi que son contexte géographique, tarifaire, etc. Le processus d'acquisition des données via les montées à bord est expliqué ainsi que la structure du système informationnel. S'en suit les méthodes de validation des données, classées par objet de transport (usagers, véhicules, chauffeurs). Diverses problématiques et données suspectes sont ainsi dégagées.

2.1 Société de Transport de l'Outaouais

Les données étudiées dans cette recherche sont gracieusement fournies par la Société de Transport de l'Outaouais (STO), pionnière québécoise dans l'implantation et l'utilisation d'un système carte à puce dans un contexte de transport collectif.

2.1.1 Présentation du réseau

Le réseau de la STO dessert la ville de Gatineau (comprenant les secteurs de Hull, Gatineau, Aylmer, Masson-Angers et Buckingham) mais est surtout orienté vers le centre-ville d'Ottawa, situé en Ontario, de l'autre côté de la rivière

Outaouais. Ottawa, capitale canadienne et pôle économique principal de la région, possède son propre réseau de transport public (OC Transpo).

Tableau 2.1 – Statistiques du réseau de la STO, juillet 2006 (Source : www.sto.ca)

Véhicules :	249 autobus
Nombre d'abribus :	268 abribus
Nombre de panneaux d'arrêts d'autobus :	1 722 panneaux d'arrêts
Passages à bord des autobus annuellement :	16 millions
Kilomètres parcourus chaque année :	Plus de 9 millions
Nombre de lignes d'autobus :	55 lignes (comprenant : 44 lignes régulières, 9 lignes express et 2 lignes interzones)
Cartes à puce en circulation :	Plus de 31 000 cartes à puce
Population desservie :	Plus de 240 000 résidents
Superficie du territoire desservi :	589 km ²



Figure 2.1 – Réseau de la STO superposé aux secteurs de la ville de Gatineau (Source : www.ville.gatineau.qc.ca et www.sto.ca)

La totalité des véhicules de la STO sont dotés du système carte à puce et la majorité (80 %) des usagers valident leurs passages avec un titre de type carte à puce, qui est physiquement composée :

- d'une puce agissant comme carte-mémoire ;

- d'une antenne permettant d'interagir à proximité (10 cm) de la borne de validation d'un bus;
- d'une surface visible comprenant photo et information sur l'utilisateur (nom et numéro d'identification de l'utilisateur).

Ainsi, dans le cas de la STO, chaque carte est personnalisée de façon à ce qu'il ne soit pas possible de transférer sa carte à un autre usager. Nous prendrons donc comme hypothèse que chaque carte est associée à un seul usager et qu'ainsi, on puisse parler indifféremment d'une carte ou d'un usager lorsqu'on suivra les déplacements effectués à travers les numéros de série de ces cartes. Notons que les usagers de la STO n'ont pas à valider leur carte à puce lors de la descente des véhicules.

2.1.2 Modes de paiement et échantillon

La STO comporte plusieurs modes de paiement. Bien que l'abonnement mensuel soit majoritairement utilisé sur le réseau (68,9 % des passages) (Société de Transport de l'Outaouais, 2004), la carte à puce supporte aussi d'autres produits tel l'abonnement annuel, saisonnier, estival, etc. (10,2 % des passages). Les autres modes de paiement sont les billets (16,1 %) vendus par carnets ainsi que l'argent comptant (4,8 %). Par conséquent, nous avons un peu plus de 79 % de tous les passages qui s'effectuent en utilisant la carte à puce. Ce sont ces passages, exprimés via les enregistrements d'une base de données, qui constituent l'échantillon étudié dans cette recherche. Il est pertinent de souligner que toute la clientèle munie d'une carte à puce est ainsi pourvue au minimum d'un abonnement d'un mois.

2.1.3 Abonnements et tarification

La STO offre une tarification selon la catégorie d'utilisateur. Afin de bénéficier du tarif réduit, une carte à puce est généralement requise et fait office de pièce d'identité. Ainsi, à l'exception des étudiants de 6 à 12 ans, un usager désirant

payer son passage à tarif réduit par argent comptant ou billet est requis de présenter une carte à puce personnalisée.

Tableau 2.2 – Tarification par catégorie d'usager (Source : Société de Transport de l'Outaouais, 2006)

Catégorie	Âge	Tarif	Preuve requise pour tarif réduit	Renouvellement de la carte
Enfants	0-5 ans	Gratuit	N/A	N/A
Étudiants	6-11 ans	Réduit étudiant	Non	N/A
	12-20 ans	Réduit étudiant	Oui (annuellement)	Oui (annuellement)
	21 ans et plus	Réduit étudiant	Oui (chaque semestre)	Non
Adultes	-	Régulier	N/A	Non
Aînés	65 ans et plus	Réduit aîné	Oui (initialement)	Non

De plus, la STO offre plusieurs types de parcours (régulier, express et interzone). Un usager peut ainsi disposer d'un abonnement pour chacune de ces catégories. Cependant, afin de monter à bord d'une ligne à tarification supérieure, les usagers (dont ceux possédant une carte à puce) doivent payer un supplément en argent comptant.

2.2 Structure des données

Chaque montée à bord est traduite par un enregistrement, éventuellement transféré à une base de données centrale, qui fait partie intégrante d'un système informationnel. Cette section expose les mécanismes de création de l'information à partir de la montée d'un usager, puis décrit les données engendrées.

2.2.1 Processus de validation d'un embarquement

C'est au moment où l'usager monte à bord d'un bus et approche sa carte de la borne de validation que l'ordinateur du véhicule conçoit la transaction. L'usager approche ainsi sa carte à moins de 10 cm du lecteur à radiofréquence qui valide l'embarquement. On peut considérer qu'il se produit alors une collision virtuelle entre l'usager et le véhicule assigné dans la mesure où ces deux entités de transport se rencontrent dans l'espace-temps.

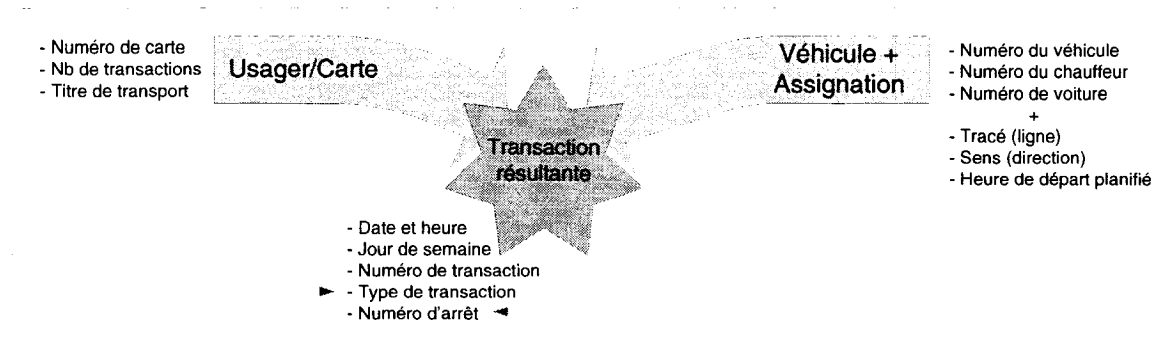


Figure 2.2 – Caractérisation des objets engendrant une transaction ainsi que la transaction engendrée, qui complète l'information sur un embarquement

Ainsi, pour chaque embarquement, une transaction est générée avec une caractérisation décrivant les deux objets impliqués, ainsi que la conjecture spatio-temporelle du moment (Figure 2.2). Tandis que l'objet « usager » est caractérisé par une information sur la carte, le terme « véhicule-assigné » est employé pour marquer la démarcation entre le véhicule physique et l'assignation qui s'y rapporte. En effet, le véhicule se caractérise par des éléments se rapportant aux ressources physiques (bus, chauffeur) tandis que l'assignation traite d'information relative au service que pourvoit ce véhicule (ligne, direction et heure de départ planifiée).

L'espace de transaction créé renferme l'information sur la temporalité de l'embarquement ainsi que deux attributs dérivés : le type de transaction (validation simple ou correspondance) ainsi que le numéro de panneau d'arrêt. Alors que le type de transaction est essentiellement déterminé par l'heure de la dernière transaction figurant sur la carte à puce de l'usager, nous verrons que le numéro d'arrêt est évalué par l'ordinateur de bord selon l'assignation du véhicule.

2.2.2 Processus d'assignation du véhicule

Au garage, lorsqu'un chauffeur prend possession d'un véhicule, il compose son numéro de chauffeur afin de s'identifier au système. Lorsqu'il quitte le garage avec son véhicule, la liste ordonnée des départs associés à ce chauffeur est

transférée à l'ordinateur de bord du véhicule. Ainsi, à la fin de chaque départ, le chauffeur doit manifester l'entrée en vigueur du départ suivant dans la liste par l'appui d'un bouton sur sa console. Lorsqu'un usager monte à bord d'un véhicule, son embarquement est ainsi marqué de l'assignation en vigueur à cet instant dans l'ordinateur de bord.

2.2.3 Dérivation de l'arrêt d'embarquement

Chaque véhicule de la STO est équipé d'un module de localisation GPS ainsi que de la séquence et des coordonnées des arrêts à être desservis selon le départ assigné. Au moment d'un embarquement, le numéro d'arrêt dérivé sera celui qui, dans la séquence des arrêts du départ en cours, est le plus rapproché de la position GPS à cet instant. Notons que sur un même départ, l'arrêt déduit ne peut pas précéder l'arrêt de l'embarquement précédent. Nous verrons que la détermination du numéro d'arrêt comporte une marge d'erreur significative, introduite par une assignation erronée.

2.2.4 Création et organisation des données

Le processus de création d'une donnée est initié lorsqu'un bus quitte le garage et qu'il accomplit une suite de départs où il collecte un certain nombre de transactions représentant des usagers qui montent à bord (Figure 2.3). Lorsque ce bus revient au garage faire le plein d'essence, une borne de téléchargement permet de décharger ses transactions sous forme d'enregistrements au serveur central qui attribue un numéro d'enregistrement unique à chaque transaction. Ainsi, un bus produira une séquence d'enregistrements successifs contenant les embarquements qu'il aura recueillis depuis son dernier déchargement.

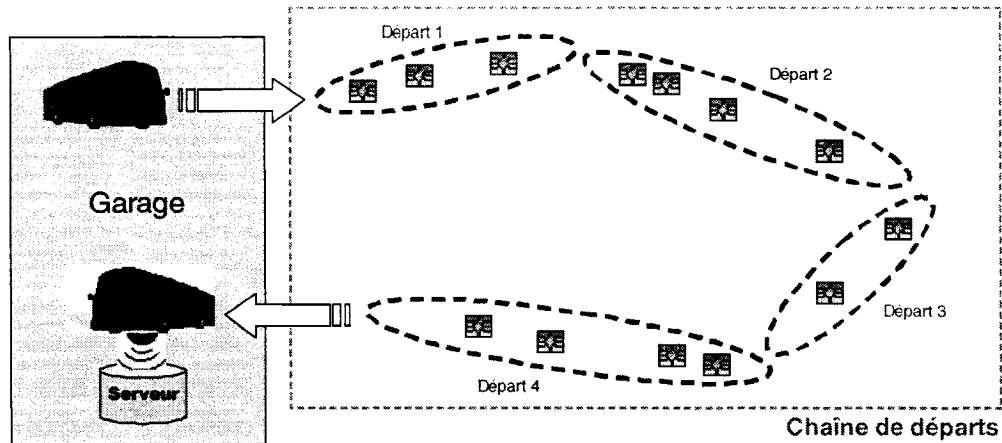


Figure 2.3 – Processus d’acquisition et de sauvegarde des transactions pour une chaîne de départs

À mesure que les véhicules viennent décharger leurs transactions, un nouveau lot de transactions à numéros successifs s’ajoute à la base de données. Nous définirons ces lots de transactions par « récoltes » qui, bien que fréquemment équivalentes aux tournées (entre une sortie et un retour au garage), peuvent s’échelonner sur plusieurs jours. Les transactions, telles qu’exprimées à travers la base de données, sont ainsi segmentées par bus, par jour puis par départ effectué. Une emphase est mise sur cette structure puisque l’application de logiques de cohérence de l’information en tiendra compte.

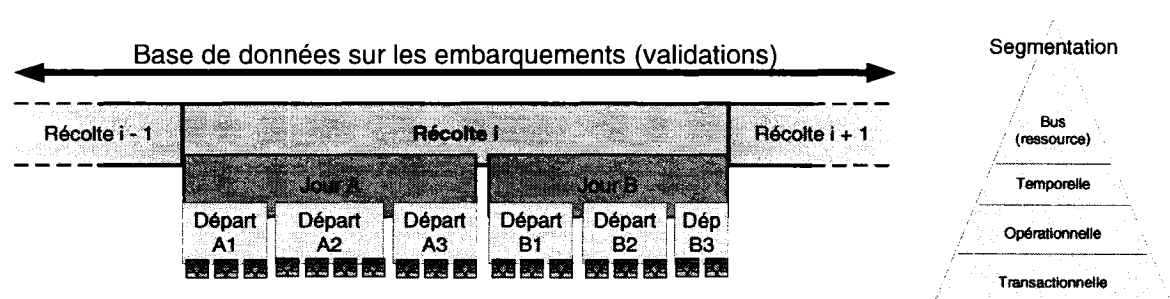


Figure 2.4 – Décomposition des séquences de transactions en récoltes, jours et départs

Plus de détails sur le système informatisé de validation des titres (SIVT) ainsi que la structure globale du système de paiement carte à puce utilisé à la STO sont décrits dans Tranchant (2005) et Trépanier et al (2005).

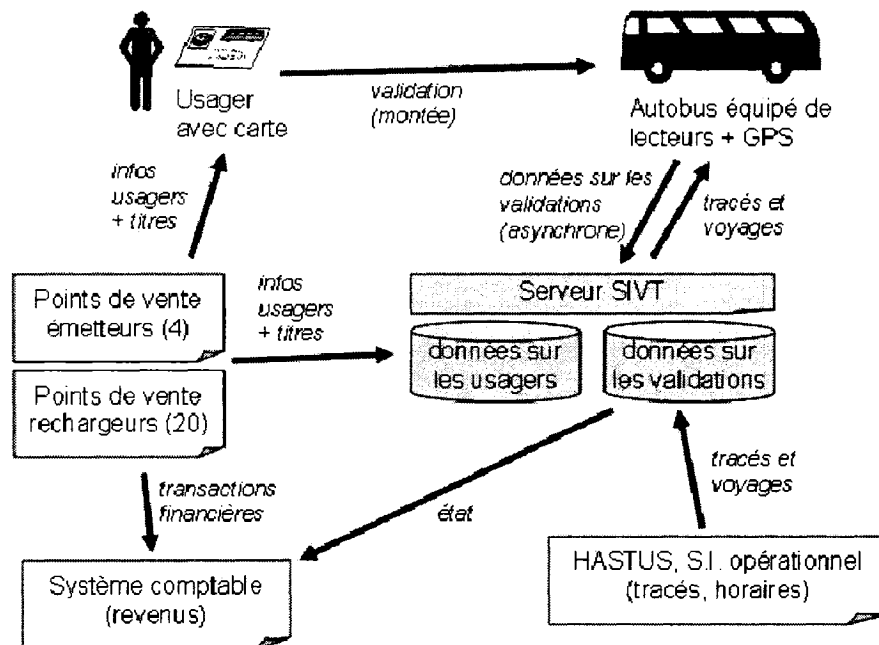


Figure 2.5 – Diagramme fonctionnel du paiement par carte à puce à la STO (Source : Trépanier et al, 2005)

2.2.5 Données disponibles

Les données étudiées dans cette recherche concernent les transactions de montées à bord de véhicule effectuées durant les mois de juillet 2003 (387 588 transactions), octobre 2003 (771 239 transactions) ainsi que février 2005 (763 570 transactions).

2.2.6 Description des champs

Afin de mieux comprendre les méthodes analytiques appliquées plus loin, cette section présente les attributs des enregistrements de la base de données. D'abord, le Tableau 2.3 présente un cliché d'enregistrement avec les descriptions sommaires des champs caractérisant chaque transaction. Les prochains paragraphes décrivent ces champs, groupés par objet.

Tableau 2.3 – Description des champs rattachés à chaque embarquement

Champ	Exemple	Mesure	Description	Objet
NumDéplacement	22811142	Num. ordinal	Un numéro unique d'enregistrement (clé primaire)	Transaction
DateDéplacement	2/1/2005	Date	La date de la transaction	
NumJour	2	Num. nominal	Le jour de la semaine de la transaction (1 à 7, plus 9 pour férié)	
HeureDéplacement	0726	Texte	L'heure de la transaction	
TypeTransaction	1	Num. nominal	Le type de transaction (validation, correspondance ou refus)	
NUM_MESS_REFUS	25	Num. nominal	Message sur transaction (oct 2003 et février 2005)	
NumArrêt	3712	Num. nominal	Le numéro du panneau d'arrêt de la transaction	Usager
NumCarte	3001088459	Num. nominal	Le numéro de la carte d'usager utilisée	
NbrTransactions	124	Num. ordinal	Le numéro de la transaction de la carte utilisée	
NumTitre	1	Num. nominal	Le titre de transport en vigueur sur la carte utilisée	
NumBus	9123	Num. nominal	Le numéro du véhicule employé	Véhicule
NumChauffeur	3359	Num. nominal	Le numéro de chauffeur en poste	
NumVoiture	187	Num. nominal	Le numéro de voiture (code opérationnel)	
NumTrace	82	Num. nominal	Le numéro de tracé assigné	Assignment
SensLigne	0	Num. nominal	Le sens du tracé assigné	
HeureLigne	0720	Texte	L'heure de départ planifiée de ce tracé-sens	

Attributs de l'usager

Trois champs concernent cet objet. Le numéro de carte est en fait son numéro de série. Notons qu'un usager qui doit physiquement remplacer sa carte se verra attribuer un nouveau numéro de carte.

Le numéro de transaction est lié à une carte et se veut un compteur incrémental augmentant d'une unité à chaque transaction. Il est cyclique sur 256 unités. Ainsi, une transaction #0 devrait suivre la transaction #255. Nous verrons que des transactions autres que des embarquements à bord de véhicules agissent aussi sur ce compteur.

Le numéro de titre équivaut à l'abonnement en vigueur sur la carte à puce. Malgré une cinquantaine de titres disponibles, seulement une vingtaine sont utilisés durant le mois de février 2005. De ce nombre, quatre titres se répartissent sur plus des trois quarts de transactions (Tableau 2.4).

Tableau 2.4 – Titres utilisés au mois de février 2005, triés selon nombre de transactions

Numéro	Titre	% transactions	Numéro	Titre	% transactions
1	ADULTE REGULIER	41.69%	39	ETUDIANT EXPRESS 2004	0.62%
38	ETUDIANT REGULIER 2004	17.15%	17	CF INTERZONE	0.59%
2	ADULTE EXPRESS	10.67%	35	CAMPUS EXPRESS	0.21%
41	INTEGRE SCOLAIRE 2004 (5)	6.79%	18	CF AINE	0.16%
15	CF REGULIER	5.23%	10	EMPLOYE STO	0.12%
37	ETUDIANT INTEGRE 2004 (7)	4.78%	40	ETUDIANT INTERZONE 2004	0.10%
34	CAMPUS REGULIER	3.43%	36	CAMPUS INTERZONE	0.02%
16	CF EXPRESS	3.24%	42	ARCHIVE	0.02%
8	AINE	2.85%	31	ETUDIANT REGULIER 2003	0.00%
3	ADULTE INTERZONE	2.31%	30	ETUDIANT INTEGRE 2003	0.00%

Attributs du véhicule

Les attributs concernant le numéro de véhicule et le numéro de chauffeur sont affiliés aux ressources internes de la STO. Ils correspondent au véhicule physique et au chauffeur présent lorsque l'embarquement est effectué.

Le numéro de voiture identifie la tournée (chaîne de départs) pour un véhicule-chauffeur. Après certains constats d'incohérence, où le numéro de voiture ne décrivait pas systématiquement les mêmes tournées, il a été déterminé que cet attribut ne serait pas utilisé dans cette recherche.

Attributs de l'assignation

Le tracé correspond à la ligne de bus sur laquelle se produit l'embarquement. Par contre, on retrouve des tracés qui ne font pas partie du service. Le plus fréquent (900) équivaut à un parcours « haut le pied ». On retrouve aussi d'autres tracés (>300) qui correspondent à des lignes spéciales et pour lesquelles un numéro d'arrêt incongru sera associé par l'ordinateur de bord. La proportion de transactions renfermant ces faux tracés est d'environ 4 % pour les trois mois de données.

Le sens est un attribut binaire (0 ou 1) qui correspond à la direction du tracé.

L'heure du départ planifiée correspond à l'heure de départ du véhicule au terminus pour ce tracé-sens. Pour environ 2 % des transactions des trois mois de données, on retrouve des heures de départ planifiées erronées affichant « 0000 ».

Attributs de la transaction

Tel qu'évoqué plus tôt, le numéro d'enregistrement est un identificateur unique de chaque transaction (clé primaire). Il est attribué séquentiellement par le serveur de données lors du téléchargement des transactions effectuées par un bus depuis son dernier déchargement.

Chaque transaction est caractérisée temporellement par les champs date, heure d'embarquement et jour de semaine. Ce dernier champ est décomposé selon les jours de semaine (1 à 5), jours de fin de semaine (6 et 7) et jours fériés (9).

Le type de transaction possède trois catégories : validation (1), correspondance (3) et refus (9). Une transaction sera de type correspondance lorsqu'une validation simple a été effectuée préalablement à l'intérieur d'un certain laps de temps. À partir de l'ensemble de données de février 2005, on ne retrouve plus le type « refus » étant donné qu'un message de transaction décrit les embarquements problématiques (présent depuis octobre 2003). Le Tableau 2.5 fait état de la fréquence des messages d'erreurs relatifs aux deux ensembles de données concernés.

Tableau 2.5 – Messages d'erreur relatifs aux transactions refusées (octobre 2003 et février 2005)

Octobre 2003			Février 2005		
Message de refus	%	% relatif	Message de refus	%	% relatif
Aucun message	98.0%		Aucun message	98.2%	
Supplément exigé	2.0%	85.6%	Supplément exigé	1.8%	95.4%
Titre invalide		9.1%	Carte expirée		2.2%
Carte expirée		4.3%	Jour invalide		1.3%
Carte invalide		0.6%	Carte invalide		0.6%
Carte refusée LN		0.3%	Titre invalide		0.3%
Carte refusée LN		0.2%	Carte refusée LN		0.1%
			Carte refusée LN		0.1%
			Ligne invalide		0.0%

Le message « supplément exigé » concerne un usager possédant un abonnement qui ne lui permet pas d'emprunter une ligne de type express ou interzone. Dans ce cas, la borne de validation homologue sa carte à puce mais exige un supplément en argent comptant.

Tel que précisé à la section 2.2.3, le numéro de l'arrêt de l'embarquement est dérivé à partir de l'assignation (connue) en vigueur au moment où l'utilisateur monte à bord, ainsi que de la position GPS (inconnue) du véhicule au moment de l'embarquement. 1406 arrêts valides sont utilisés au cours du mois de juillet 2003.

2.2.7 Autres tables de données

D'autres tables de références sont aussi utilisées dans cette recherche (exemples à l'annexe A). Elles incluent une table contenant les descriptions nominales des titres de transport disponibles, une table renfermant les 1856² arrêts valides (géocodés) du réseau ainsi qu'une table des 7255³ arrêts-ligne valides définissant le service. La distinction entre ces deux concepts suit.

2.2.8 Confidentialité des données

Les données provenant des ensembles étudiés de la STO ne comportent aucune information nominative ou personnelle. Seul un numéro de carte (numéro de série) anonyme définit un usager et ne peut pas être mis en relation avec un particulier.

2.3 Concepts du transport collectif

Plusieurs notions propres au transport en commun sont utilisées dans ces pages. Cette section en définit quelques-unes.

2.3.1 Arrêts et arrêts-ligne

On peut visualiser les arrêts comme des nœuds (unidimensionnels) et les arrêts-lignes comme des nœuds et des liens (bidimensionnels) entre ces nœuds, qui comportent une notion de séquence où l'ordre est important.

² Table des arrêts de la base de données de juillet 2003. Respectivement 1853 et 1869 pour octobre 2003 et février 2005.

³ Idem. Respectivement 7195 et 7333 pour octobre 2003 et février 2005.

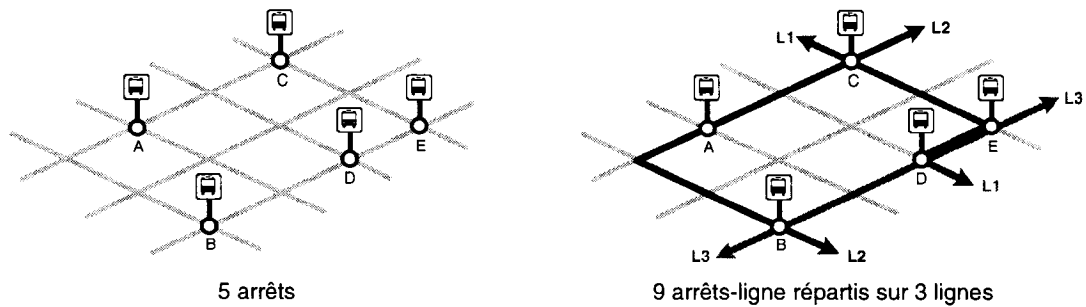


Figure 2.6 – Distinction entre arrêts et arrêts-ligne

Un tracé-sens, tel que défini dans la table des arrêts-ligne, comporte ainsi une liste ordonnée des numéros des arrêts desservis. Un arrêt peut donc être commun à plusieurs lignes. Étant donné que la table des arrêts de la STO comporte les coordonnées XY (projection UTM) des arrêts, il est possible de les repérer géographiquement. La Figure 2.7 utilise ainsi les données des deux tables afin de donner un aperçu des lignes desservant les arrêts du secteur d'Aylmer.



Figure 2.7 – Parcours et arrêts associés couvrant une partie du secteur d’Aylmer

2.3.2 Déplacements et chaînes de déplacement

Une transaction est catégorisée comme correspondance si elle se fait à l’intérieur d’un certain intervalle de temps après une première transaction (validation simple). Nous définirons ainsi un déplacement comme une suite de transactions successives effectuées à l’intérieur de cet intervalle temporel. En conséquence, chaque validation simple constitue en fait le premier (et possiblement unique) maillon d’un déplacement qui peut inclure une ou plusieurs transactions de correspondance. De ce fait, l’identification des déplacements associés à une carte passera par la localisation de ses validations simples.

Parallèlement, un déplacement comportant plusieurs transactions fait appel à la notion de déplacement lié (linked trip) qui est défini par Bagchi (2003) comme deux ou plusieurs embarquements sur des véhicules différents par un individu lors d’un même jour à l’intérieur d’un certain laps de temps.

Enfin, une chaîne de déplacement est normalement définie par l'ensemble des déplacements réalisés sans revenir au domicile. Ainsi, afin de reconnaître et départager les chaînes de déplacement d'un usager, nous établirons l'arrêt-domicile affilié à chaque carte (section 4.1.5) et admettrons l'hypothèse selon laquelle un usager revient à son domicile à la fin de chaque journée.

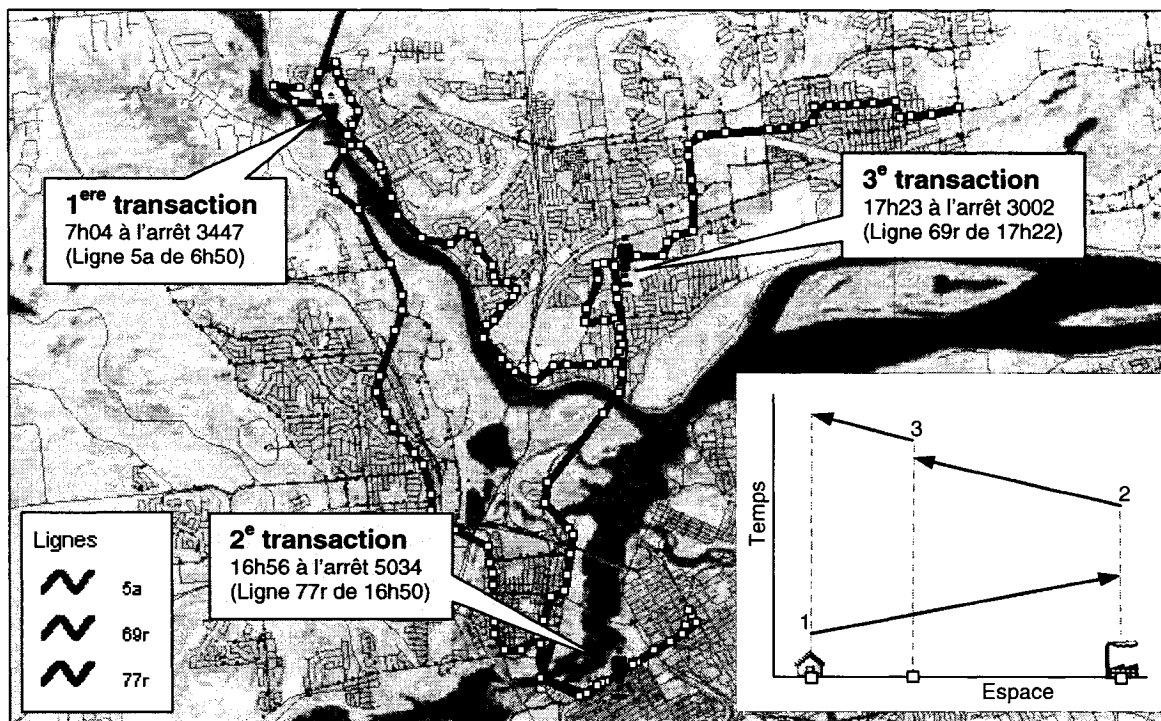


Figure 2.8 – Représentation d'une chaîne de déplacement composée de deux déplacements, où trois transactions ont été observées pour un usager

Pour le mois de février 2005, un jour moyen de semaine compte en moyenne 36 000 transactions, dont environ 15 % de correspondance. Ce taux est identique pour les deux autres mois étudiés, bien que juillet 2003 comporte nettement moins de transactions (Figure 2.9). Les concepts de déplacement, déplacement lié et chaîne de déplacement seront exploités plus en détail dans le chapitre portant sur la caractérisation de la demande.

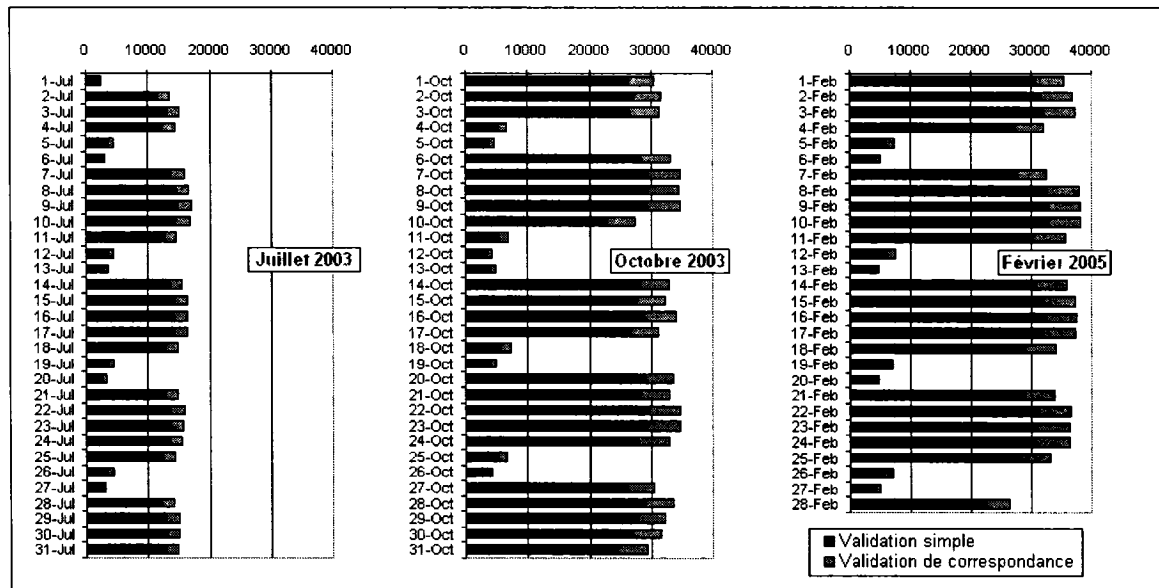


Figure 2.9 – Nombre de transactions de validation et de correspondance par jour (juillet 2003, octobre 2003 et février 2005)

2.3.3 Véhicules, tournées et départs

On a vu qu'une récolte consiste en une série d'enregistrements correspondant à l'ensemble des transactions emmagasinées par un véhicule, depuis le dernier déchargement au garage. Bien que les enregistrements d'une récolte soient parfois étalés sur plusieurs jours, il est fréquent de retrouver plus d'une récolte par jour pour un véhicule. Conséquemment, ces situations traduisent que le véhicule est revenu au garage au cours de cette journée.

Cela rejoint le concept de tournée, définie comme le moment entre la sortie d'un véhicule jusqu'à son retour au garage. Ainsi, pour un véhicule-jour, le fait de retrouver des enregistrements qui proviennent de différentes récoltes, exprime que plus d'une tournée a été effectuée. Les transactions d'un véhicule-jour peuvent ainsi être segmentées en tournées, puis en départs. Un départ correspond à un trajet sur un parcours orienté (tracé-sens) débutant à une certaine heure planifiée.

On peut quantifier ces concepts en examinant les données à l'aide des attributs temporels et l'assignation des transactions ainsi qu'un attribut ajouté identifiant les récoltes (décrit à la section 2.4). Ainsi, à la STO, un jour moyen de semaine de février 2005 utilisera 201 véhicules, répartis sur 290 tournées et effectuant 632 départs (Figure 2.10). Ces résultats sont très stables, à l'image que l'on se fait du service. En effet, pour ces trois valeurs, le ratio de l'écart-type à la moyenne ne varie que de 1,2 % à 1,6 %.

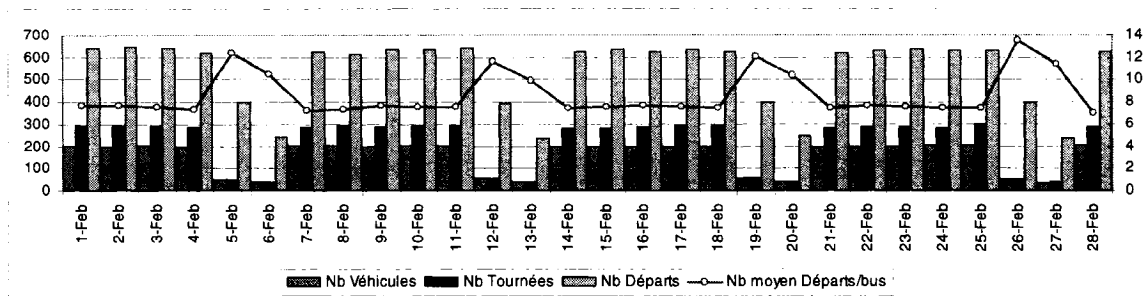


Figure 2.10 – Nombre de bus, tournées, départs et ratio moyens de départs par bus pour les jours de février 2005

Afin de visualiser l'étalement temporel de ces tournées à l'échelle d'une journée, les graphiques de la Figure 2.11 utilisent l'heure de la première et la dernière transaction de chaque tournée afin d'estimer le début et la fin de chaque tournée.

Dans le graphique du haut, l'ordonnancement des numéros de tournée se fait selon l'arrivée au garage. On peut y distinguer les tournées n'ayant desservi que la pointe du matin (vers le bas), les tournées s'étalant sur une partie de la journée (milieu et haut) et celles n'ayant desservi que la pointe du soir (haut).

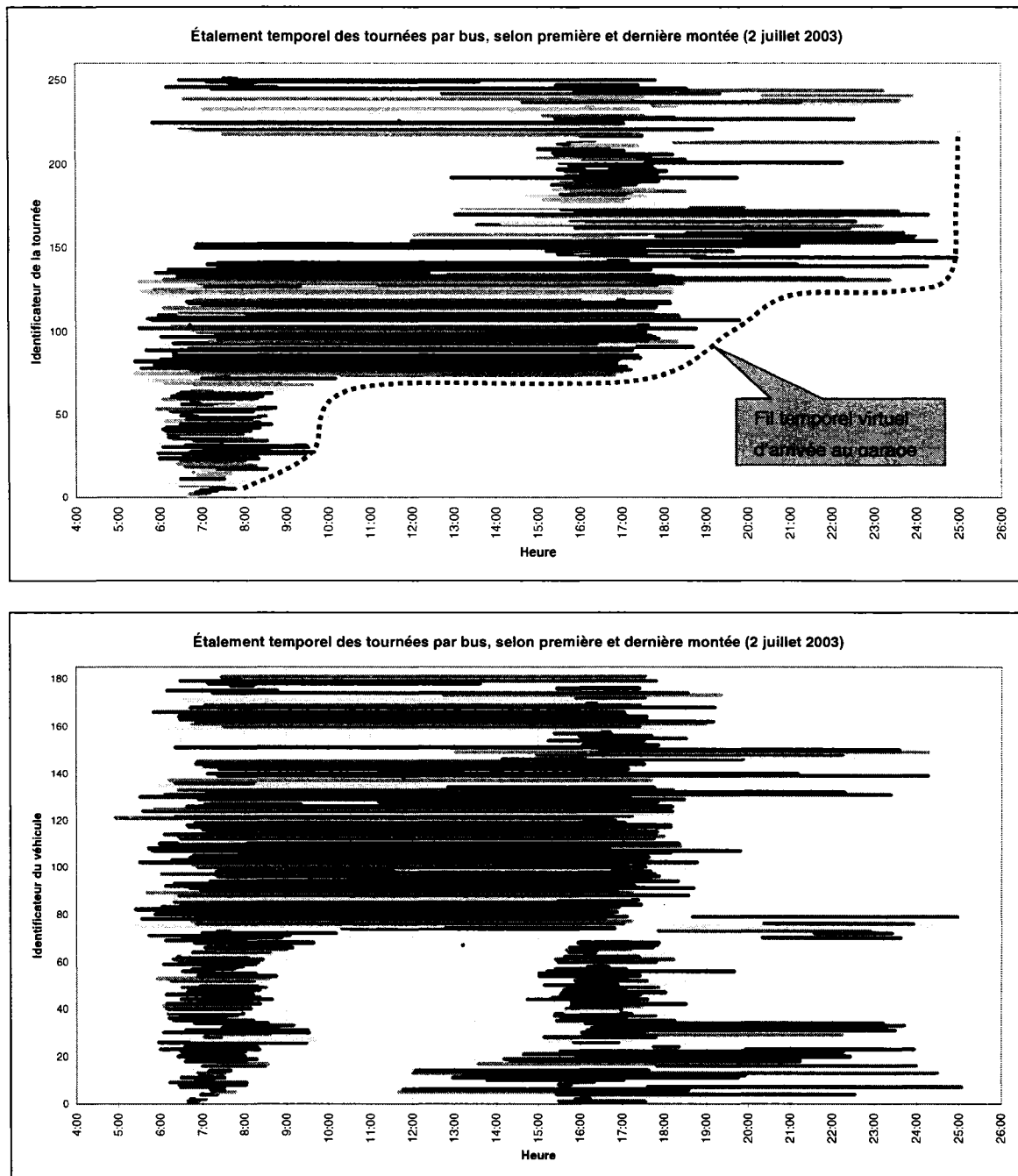


Figure 2.11 – Intervalles temporels des tournées pour un jour moyen de semaine (haut : une ligne par tournée, bas : une ligne par bus)

Puisqu'il est fréquent qu'un même bus desserve la pointe du matin, rentre au garage puis reparte effectuer la pointe du soir, il peut être aussi intéressant de

regrouper les tournées selon le véhicule utilisé. Le graphique du bas illustre donc les plages occupées par chaque bus au cours de la journée. Pour cette journée, la répartition des 181 bus ayant effectué une ou deux tournées est respectivement de 61 % et 39 %.

2.3.4 Occupation du réseau et pointe

Avec des données de carte à puce où seule l'information sur les embarquements est connue, il est difficile d'évaluer précisément l'occupation du réseau. La Figure 2.12 estime cette occupation en agrégeant le nombre de transactions par tranche de quinze minutes.

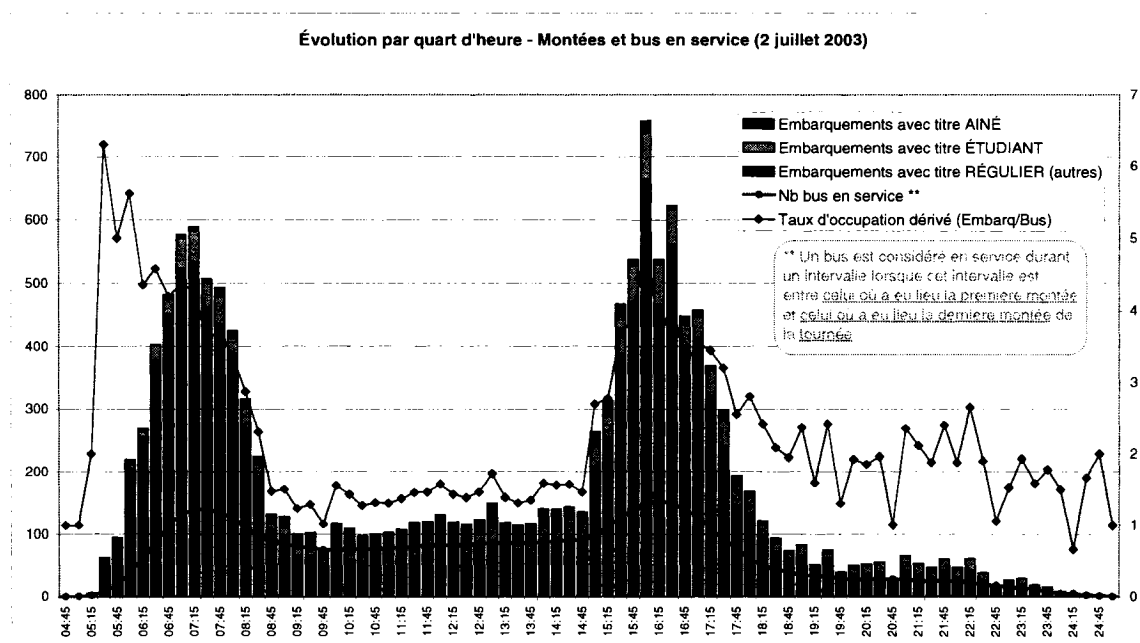


Figure 2.12 – Estimation de l'occupation du réseau (ventilé par catégorie de titre), des bus en service et du taux d'occupation dérivé par quart d'heure (2 juillet 2003)

On distingue l'intensité du nombre de déplacements débutant durant les pointes AM et PM. Lorsqu'on considère tout un mois de données (février 2005), on confirme que plus des trois quarts des déplacements en semaine débutent entre 6h et 9h (39 %) ou entre 15h et 18h (37 %).

2.4 Préparation des données

Deux éléments de prétraitement des données sont brièvement présentés car utilisés lors de manipulations subséquentes.

2.4.1 Identificateur de récoltes

Afin de distinguer chacune des récoltes effectuées par les véhicules, un algorithme traverse séquentiellement les numéros d'enregistrement (triés). Un numéro d'identification de récolte est attribué pour chaque séquence de transactions consécutives avec le même numéro de véhicule. Cet identificateur sera principalement utilisé par l'algorithme de détermination du service réalisé (section 3.1).

2.4.2 Séquence des transactions d'un usager-jour

Nous verrons que l'heure de transaction n'est pas complètement fiable à cause de problèmes de synchronisation des horloges des systèmes embarqués. On ne peut donc pas s'y rapporter afin d'établir la séquence réelle des transactions d'un usager-jour. D'autre part, le numéro de transaction est fiable mais potentiellement désordonné numériquement. En effet, il arrive fréquemment qu'à l'intérieur d'une même journée, cette valeur cyclique (0 à 255) qui doit normalement s'incrémenter lors d'une transaction, passe de la valeur 255 à 0, se répète ou « recule » subitement à la valeur 2, probablement dû à une réinitialisation de la carte dans un point de service.

Il est question ici de créer un nouveau champ qui évalue les séquences réelles des transactions d'un usager-jour. Ainsi, un script parcourt les enregistrements (triés par usager-jour, numéro de transaction puis heure d'embarquement) et attribue un numéro de séquence selon le numéro de transaction rencontré. Pour la première transaction d'un usager-jour, l'enregistrement est toujours marqué du numéro de séquence « 1 ». Pour les transactions subséquentes, le numéro de

séquence est habituellement augmenté dans la même proportion que la variation du numéro de transaction. Ainsi, il est accru

- de la valeur de l'incréméntation $\Rightarrow$ si le numéro de transaction s'incréménte ou
- de 1 $\Rightarrow$ si le numéro de transaction ne change pas ou recule à 2.

Notons que dans le premier cas, la valeur de l'incréméntation est conservée afin de pouvoir revenir sur le phénomène de données manquantes. Les détails du script décrit se retrouvent à l'annexe C.

Afin d'illustrer la différence entre le numéro de transaction et la séquence de transaction, la Figure 2.13 expose un exemple des transactions d'un usager qui part de son domicile (A), se rend à son lieu de travail (B) puis revient chez lui en passant recharger sa carte dans un point de vente (C). On pourrait observer des numéros de transaction associés 34, 34, 35, 2 et 3 alors qu'une séquence de transaction qui progresse de 1 à 5 est plus facile à utiliser.

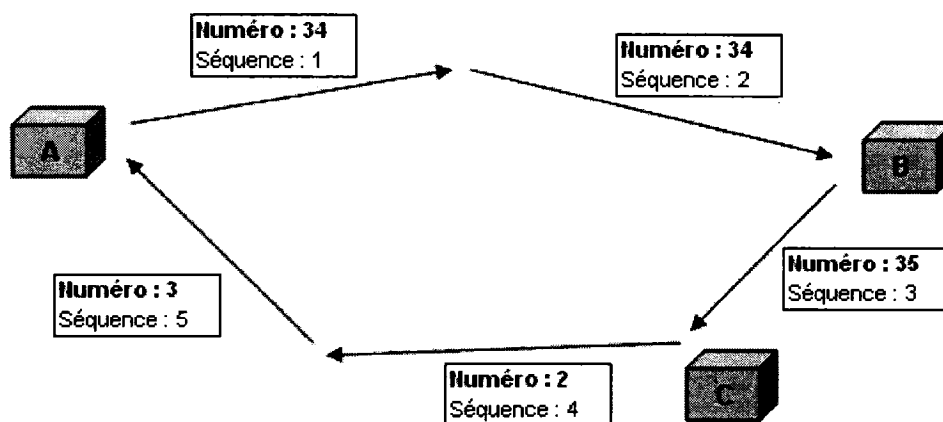


Figure 2.13 – Distinction entre numéros de transaction originaux et numéros de séquence d'un usager-jour

Ce numéro de séquence sera employé à plusieurs égards, dont celui d'évaluer directement si une transaction est la première de la journée.

2.5 Application de logiques de validation

À l'aide de logiques spatio-temporelles, les données opérationnelles du mois de juillet 2003 ont été testées afin de dégager certaines anomalies. Les logiques présentées sont classées selon l'usager, le véhicule, le chauffeur ou l'assignation opérationnelle.

2.5.1 Usager

L'usager du transport collectif, à travers l'usage de sa carte à puce, obéit à plusieurs contraintes physiques et opérationnelles. En effet, une carte :

- ne devrait pas être validée deux fois à l'intérieur d'un très court intervalle;
- ne devrait pas être validée au même moment dans deux bus distincts;
- ne devrait pas être validée au même moment à deux arrêts distincts;
- devrait voir ses correspondances être accomplies à l'intérieur d'un certain laps de temps, après une transaction préalable;
- devrait avoir ses numéros de transaction séquentiels et continus;
- devrait avoir ses numéros de transaction quotidiens chronologiques;

Lorsque ces validations opérationnelles sont appliquées systématiquement aux données, plusieurs constats ressortent :

Enregistrements manquants

On arrive à cette conclusion suite au fait qu'on retrouve de nombreux petits sauts⁴ dans l'ordre séquentiel des numéros de transaction par carte, notamment lorsqu'une validation simple préliminaire à une correspondance est introuvable. Dans le cas général, les transactions manquantes pourraient n'être que des

⁴ 4212 occurrences pour juillet 2003, dont 97 % où le saut est d'une unité.

transactions administratives ne dénotant pas un embarquement (i.e. recharge de carte). Pour les cas particuliers où la transaction suivante à l'enregistrement manquant est une correspondance (correspondance orpheline), nous verrons à la section 3.1 qu'elles ne représentent que 122 (0,2 %) des 59674 correspondances du mois de juillet 2003. On remarque aussi des sauts à grande échelle dans les numéros d'enregistrement, comme si des récoltes entières étaient absentes. Ces « trous » dans les données ne représentent pas nécessairement des enregistrements manquants, étant donné leur processus d'extraction (pour fins de recherche) depuis le serveur de la STO. En effet, malgré que la date de déchargement des transactions d'un véhicule au serveur central soit inconnue (on peut toutefois la déduire avec les déchargements des véhicules précédents et suivants), les numéros d'enregistrement sont attribués selon cette date de déchargement et non selon la date où les transactions ont été effectuées. On peut constater ce fait à la Figure 2.14 où on constate que plusieurs récoltes accumulent des transactions pour plusieurs jours avant de se faire décharger sur le serveur. On voit cependant que ce phénomène n'est pas si fréquent ; 96 % des récoltes ne comportent qu'une ou deux journées.

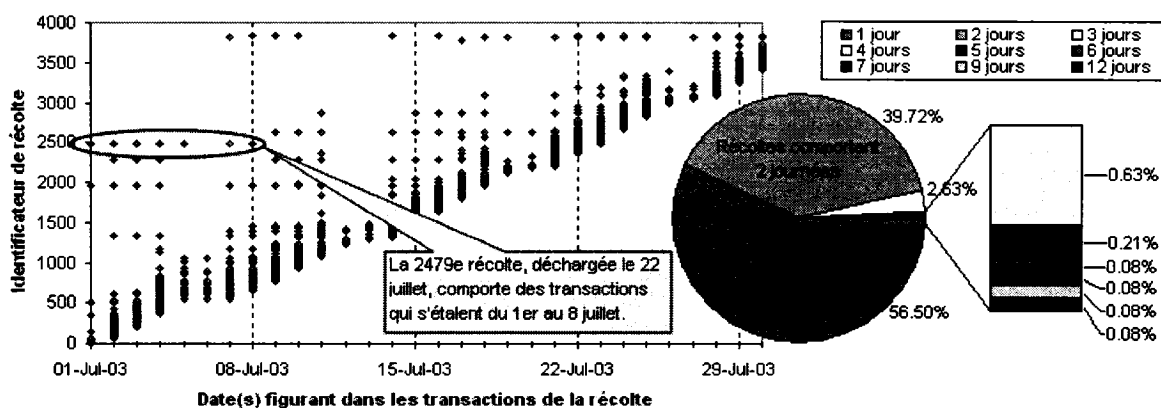


Figure 2.14 – Dates d'embarquement présentes dans les enregistrements des 3837 récoltes et répartition du nombre de jours par récolte (juillet 2003)

Ainsi, les trous dans les numéros d'enregistrement s'expliquent par le fait que lorsqu'un certain mois de données est extrait du serveur, l'extraction ne

considère que les transactions dont la date d'embarquement figure dans ce mois. Ces trous représentent ainsi des embarquements s'étant effectués dans d'autres mois. Il est même possible que si l'extraction est faite trop tôt, des embarquements du mois concerné n'aient pas encore été déchargés sur le serveur.

D'ailleurs, le plus grand numéro d'enregistrement appartenant à juillet 2003 est supérieur (11901542) au plus petit numéro d'octobre (11828666). C'est-à-dire que la transaction du mois de juillet fut déchargée après celle du mois d'octobre.

Mauvais arrêt assigné

En vérifiant qu'un usager ne devrait pas se trouver à deux arrêts simultanément, on retrouve des anomalies (sur le plan de la distance) qui laissent présager qu'un des arrêts n'est pas correctement déduit. Comme nous le verrons dans la section sur les logiques de l'objet véhicule, cette situation survient lorsqu'une assignation erronée contamine la dérivation de l'arrêt d'embarquement.

Erreurs dans l'horloge interne du système embarqué

En faisant appel au champ dérivé indiquant les séquences de transactions pour un usager-jour (décrit à la section 2.4.2), une vérification de la chronologie des numéros de transaction (pour un usager-jour) détecte certaines anomalies en ce qui a trait à l'heure d'embarquement. On les remarque dans les situations où deux transactions sont effectuées dans deux bus différents à quelques minutes d'intervalle et où la 2^e transaction apparaît temporellement avant la 1^{re}.

Pour le mois de juillet 2003, on répertorie 22 occurrences où la chronologie est déficiente entre deux transactions consécutives. Parmi ces cas, on en retrouve 14 où la deuxième transaction a lieu une minute avant la transaction précédente. Nous sommes ainsi probablement en présence d'une légère désynchronisation de l'horloge interne de l'un des deux véhicules. Une quantification exacte de ce

phénomène est difficile, étant donné les circonstances exceptionnelles où un usager monte dans deux véhicules différents durant un très court intervalle.

Dans les 8 cas restants, on est en présence de différences temporelles importantes allant de 28 à 152 minutes. Après observation de la régularité des déplacements des usagers (exemple à la Figure 2.15) touchés par ces transactions problématiques, on confirme que la désynchronisation affecte aussi les heures des autres transactions de ces véhicules.

	NumDéplacement	NumTransactions	NumCarte	NumTire	NumJour	DateDéplacement	HeureDéplacement	TypeTransaction	NumTrace	SensLigne	HeureLigne
Lu	10561493	210	1923469362	1	1	14-Jul-03	0748	1	45	0	0725
	10568488	211	1923469362	1	1	14-Jul-03	0811	3	37	1	0756
	10570394	212	1923469362	1	1	14-Jul-03	1537	1	33	0	1510
	10581546	213	1923469362	1	1	14-Jul-03	1601	3	45	1	1534
Ma	10583673	214	1923469362	1	2	15-Jul-03	0736	1	45	0	0712
	10581182	215	1923469362	1	2	15-Jul-03	0759	3	37	1	0740
	10583761	216	1923469362	1	2	15-Jul-03	1530	1	900	0	1249
	10596126	217	1923469362	1	2	15-Jul-03	1540	3	45	1	1520
Me	10615841	218	1923469362	1	3	16-Jul-03	0750	1	45	0	0725
	11891208	219	1923469362	1	3	16-Jul-03	0815	3	37	1	0000
	10600575	220	1923469362	1	3	16-Jul-03	1531	1	37	0	1514
	11289078	221	1923469362	1	3	16-Jul-03	1259	3	45	1	0000
Je	10613859	222	1923469362	1	4	17-Jul-03	0748	1	45	0	0725
	10611975	223	1923469362	1	4	17-Jul-03	0810	3	37	1	0756
	10620781	224	1923469362	1	4	17-Jul-03	1532	1	33	0	1510
	10630168	225	1923469362	1	4	17-Jul-03	1539	3	900	0	1508
Ve	10639708	226	1923469362	1	5	18-Jul-03	0756	1	45	0	0733
	10640963	227	1923469362	1	5	18-Jul-03	0811	3	37	1	0756
	10637794	228	1923469362	1	5	18-Jul-03	1535	1	49	1	1407

L'heure de transaction semble erronée... lorsqu'on regarde les déplacements similaires le reste de la semaine

Figure 2.15 – Transactions de la carte 1923469362, du 14 au 18 juillet 2003

Néanmoins, les ensembles d'enregistrements ayant des heures erronées ne se limitent que sur trois récoltes. C'est-à-dire que l'heure d'embarquement n'est erronée que pendant les départs de ces récoltes, et pas nécessairement sur toute la journée. Pour le mois de juillet 2003, cette problématique importante affecte 4 départs et 86 transactions (Tableau 2.6).

Tableau 2.6 – Récoltes où l'on constate un décalage important de l'horloge interne

Date	Id	Charnier	Transactions affectées	Départ affecté	Décalage constaté
mardi 8 juillet	9113	a	22	1	43 minutes
jeudi 24 juillet	4	b	29	2	116 minutes
mercredi 16 juillet	9002	c	35	2	n/a

Le décalage de l'horloge du bus concerné est estimé en utilisant la différence entre l'heure d'embarquement de la 1^{re} transaction d'un départ avec l'heure planifiée de ce départ au terminus. Le décalage de l'horloge est ainsi supérieur au maximum de ces différences pour chaque départ affecté. Dans le cas de la troisième récolte, on ne peut estimer le décalage car les deux départs concernés sont marqués de l'heure de départ planifiée « 0000 ».

Délai de correspondance

Avec la logique concernant le délai de correspondance, on déduit expérimentalement que la STO définit une transaction comme étant une correspondance lorsque cette transaction est effectuée moins de 120 minutes après une première validation simple.

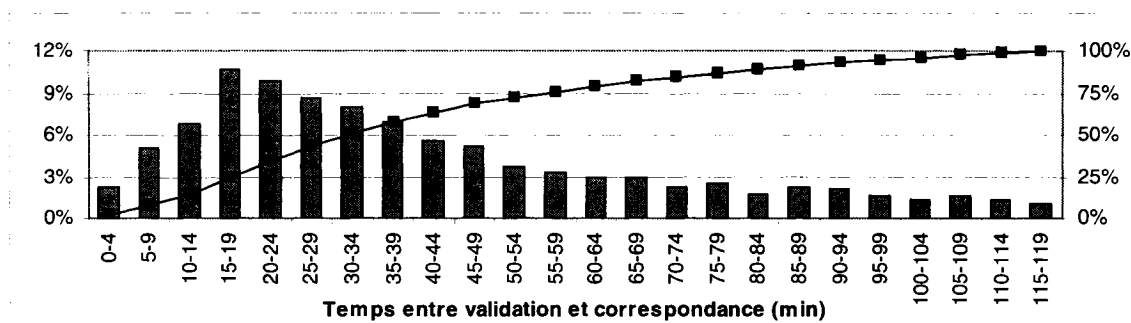


Figure 2.16 – Répartition standard et cumulée des intervalles de temps entre une correspondance et sa validation simple associée (juillet 2003)

2.5.2 Véhicule

Le véhicule de transport en commun est soumis à plusieurs règles opérationnelles. En effet, un véhicule :

- ne devrait pas rester trop longtemps au même arrêt;
- ne devrait pas se déplacer au-delà d'une certaine vitesse;
- ne devrait pas être à deux arrêts différents en même temps;
- devrait voir ses transactions quotidiennes survenir chronologiquement;

Ici aussi, en confrontant ces logiques aux données, on y déduit certaines anomalies :

Erreurs dans l'assignation du véhicule

En repérant des transactions où pour une longue période, les transactions d'un véhicule s'effectuent toujours au même arrêt, on constate des erreurs d'assignation qui sont généralement imputables aux chauffeurs. Afin d'appuyer cette affirmation, rappelons d'abord le mécanisme selon lequel l'arrêt est déterminé.

L'ordinateur de bord du véhicule connaît la liste des départs assignés au chauffeur ainsi que la séquence des arrêts faisant partie de ces départs. Ainsi, chaque fois qu'un usager monte à bord et qu'une transaction est générée, l'ordinateur y associe l'arrêt du départ en vigueur le plus rapproché de sa position GPS à cet instant. Donc, lorsqu'un chauffeur ne signale pas à l'ordinateur de bord qu'il effectue un départ différent, l'arrêt assigné aux prochains embarquements sera un arrêt appartenant à la séquence du départ obsolète. De plus, chaque départ étant constitué d'une liste ordonnée d'arrêts à desservir et que l'ordinateur de bord ne peut pas assigner un arrêt précédent (dans la liste) à l'arrêt de la dernière transaction, les arrêts erronés se situent donc généralement vers la fin de parcours du tracé-sens-départ précédent. On se retrouvera ainsi avec des séquences d'arrêts identiques erronés.

Afin d'illustrer cette conjoncture et son implication, la Figure 2.17 distingue les départs réalisés du point de vue opérationnel avec ceux perçus du point de vue informationnel. Ainsi, un véhicule effectuant deux départs (respectivement cinq et sept arrêts) verra trois usagers monter à bord du 1^{er} départ (aux arrêts 1A, 1B et 1D) et cinq autres lors du 2^e départ (aux arrêts 2B, 2C, 2D, 2E et 2G).

Normalement, en débutant son 2^e départ, le chauffeur aurait dû le signaler à l'ordinateur de bord. Ayant omis de le faire, l'ordinateur de bord assigne aux transactions subséquentes l'assignation du départ précédent. De plus, l'arrêt 1E

étant l'arrêt (faisant partie de la séquence du 1^{er} départ) le plus près des positions GPS de ces embarquements subséquents, l'ordinateur de bord alloue ainsi cet arrêt à ces transactions.

Suivant l'embarquement à l'arrêt 2D, le chauffeur constate son erreur et signale à l'ordinateur qu'il effectue le 2^e départ. Par conséquent, les embarquements effectués aux arrêts subséquents (2E et 2G) possèdent l'assignation et l'arrêt correct. C'est ainsi que dans la base de données, le 1^{er} départ comportera six embarquements (dont trois à l'arrêt 1E) et le 2^e départ en comportera deux, ce qui ne correspond pas à la réalité.

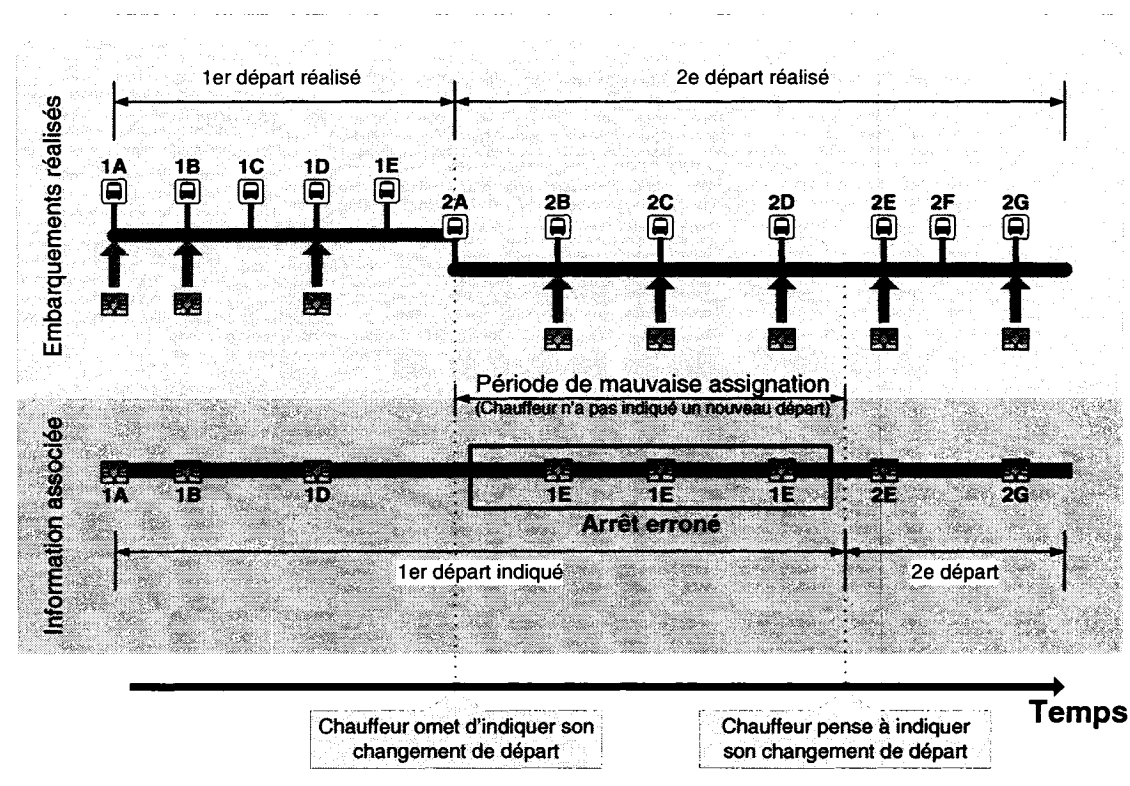


Figure 2.17 – Répercussion de l'omission d'un changement de départ sur l'information des transactions

Ainsi, lorsqu'un chauffeur oublie pour une certaine période de signaler ses changements de départ, une conséquence observable est que les transactions réalisées durant cette période peuvent renfermer une série d'arrêts identiques.

Une vérification systématique de ces séquences d'arrêts identiques montre aussi qu'il existe quelques situations où un chauffeur appuie sur son bouton de changement de départ plus d'une fois. Par le même raisonnement, cette situation amène également des séquences d'arrêts identiques erronés, mais faisant parti du départ planifié suivant. Bien que les séquences d'arrêts erronés provenant d'un oubli de signallement produisent des arrêts erronés se situant généralement vers la fin de la séquence d'arrêts du parcours obsolète et que celles d'un signallement redoublé produisent des arrêts erronés se situant plutôt vers le début du parcours suivant, la géométrie des tracés demeure un facteur déterminant de la déduction de l'arrêt erroné.

Afin de quantifier les transactions affectées par ces mauvaises assignations, on recherche des séquences d'enregistrements renfermant le même arrêt durant un intervalle temporel minimal (ici quinze minutes). Il est difficile de s'assurer d'un bon seuil car un intervalle temporel minimal trop bas risque de prendre en compte les situations où un usager monte à bord avant l'heure de départ (le chauffeur ayant laissé les portes de son véhicule ouvertes au terminus) tandis qu'un intervalle temporel minimal trop élevé risque de ne pas tenir compte d'un bon nombre de courtes périodes d'oubli de changement de départ.

Pour chacun des mois étudiés, le nombre d'occurrences où l'on retrouve un véhicule immobilisé plus de quinze minutes est ainsi approximativement évalué (Tableau 2.7). Cependant, Il est difficile de mesurer la variation relative de ces occurrences étant donné la dissemblance du niveau d'activité des mois comparés. Par contre, sur le plan des taux de transactions affectées par ces mauvaises assignations, il passe de 15 % à 8 % en dix-neuf mois, traduisant vraisemblablement une amélioration des comportements des chauffeurs.

Tableau 2.7 – Indicateurs relatifs à la qualité des données étudiées : juillet 2003, octobre 2003 et février 2005

	Juillet 2003	Octobre 2003	Février 2005
Nombre de transactions totales	387588	771239	763570
Proportion sur intervalle des nos de transaction	26.07%	19.72%	36.74%
Proportion de transactions refusées	2.41%	0.001%	0%
Proportion de transactions avec message d'erreur	0%	1.963%	1.83%
Proportion de transactions refusées ou avec message d'err	2.41%	1.96%	1.83%
Proportion de transactions avec arrêt invalide	4.19%	4.53%	3.47%
Proportion de transactions avec tracé invalide	4.22%	4.52%	3.47%
Proportion de transactions avec heure de départ invalide	2.30%	2.48%	1.98%
Occurences où un bus est resté à un arrêt < 15min	3562	4094	2486
Proportion de transactions affectées (mauvaise assignation)	15.14%	13.65%	8.35%

Vitesse anormalement élevée d'un véhicule

On désire valider la vitesse entre les transactions consécutives d'un même véhicule. Ainsi, pour chaque transaction, on calculera la vitesse avec la transaction précédente seulement si :

- le véhicule est le même;
- l'écart temporel n'est pas nul;
- l'écart temporel est inférieur à un certain seuil;
- les deux arrêts sont valides.

On utilise un écart temporel minimal dans le but d'éliminer les vitesses entre deux transactions où il y aurait eu un changement de départ, de chauffeur ou une perturbation de service. Afin de mesurer la sensibilité à l'écart temporel maximal, des seuils de 5, 10, 15 et 20 minutes ont été utilisés. La Figure 2.18 illustre la répartition des vitesses dérivées des 14283 transactions du 2 juillet 2003 en fonction de ces seuils.

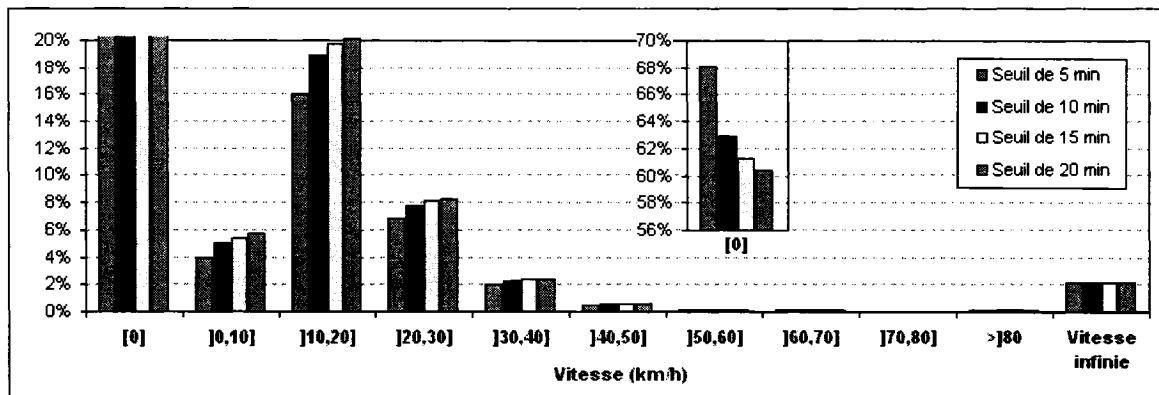


Figure 2.18 – Distribution des vitesses dérivées entre deux transactions consécutives selon seuil temporel maximal entre les deux transactions (2 juillet 2003)

On s'aperçoit qu'une grande majorité des vitesses sont exactement nulles car elles représentent une situation où les deux transactions ont été effectuées au même arrêt ($\Delta\text{distance}=0$).

À l'autre limite, une « vitesse infinie » indique une distance parcourue à l'intérieur de l'unité de temps minimale (1 min). Les transactions successives ont eu lieu à différents arrêts mais durant la même unité temporelle ($\Delta\text{temps}=0$). On doit donc vérifier que la distance parcourue par les véhicules durant cette minute est plausible. À l'exception d'une distance de 5145 m, attribuable à une erreur d'assignation (corrigée par le chauffeur, entre deux embarquements la même minute), la plus grande distance parcourue durant une minute est 705 m, ce qui équivaut à une vitesse, vraisemblable, de 42 km/h.

Entre les deux extrêmes ($v=0$ et $v=\infty$), seulement quelques vitesses dérivées s'avèrent suspectes (supérieures à 80 km/h). Pourtant, l'examen manuel de ces (15) occurrences ne révèle pas de problématique apparente. À cet effet, le Tableau 2.8 présente les enregistrements des deux transactions successives, ayant la plus grande vitesse dérivée (358 km/h).

Tableau 2.8 – Éléments informationnels de deux transactions successives où une vitesse de 358 km/h est dérivée

ID Trans	Date	HrDép	NumTrace	SensLigne	HrLigne	NumBus	Chauf	NumArrêt	COORD X	COORD Y	Dist(m)	ΔT(min)	Vit (km/h)
10441489	02-Jul-03	0644	5	0	0620	9109	3146	3445	438917	5046653	5962	1	358
10441490	02-Jul-03	0645	5	0	0620	9109	3146	3451	439996	5040789			
Success.	Date et heure	Assignment identique						Arrêts et coordonnées UTM			Champs calculés		

Dans ce cas, tout semble cohérent. En effet, après vérification à la table des arrêts-ligne, les deux arrêts sont successifs et la distance dérivée (euclidienne) de 5962 m est confirmée par une différence de chaînage de 6,5 km. Même cohérence du point de vue de la régularité des heures d'embarquement et des arrêts des deux usagers concernés. En fait, une analyse systématique de tous les intervalles de temps entre les embarquements d'un véhicule-jour-départ entre ces deux arrêts du tracé-sens donne des temps qui varient entre 1 et 9 minutes, avec une moyenne de 4:07 minutes pour juillet 2003 (33 obs.) et de 2:35 minutes pour février 2005 (38 obs.). Afin de mettre en perspective, on pourrait déterminer qu'un décalage minimal de 4 minutes serait plausible (dans la mesure où l'on a potentiellement 4:59 minutes de disponible) avec une vitesse (instantanée) de 72 km/h.

Cette problématique reste donc entière. Malgré une cohérence de plusieurs points de vue dans les occurrences où la vitesse dérivée est anormalement élevée, une absence de mise à jour ou un mauvais géopositionnement de l'un des arrêts concernés est suspecté.

2.5.3 Chauffeur

Le chauffeur à l'intérieur d'un véhicule est aussi soumis à certaines règles opérationnelles dont qu'un chauffeur :

- ne devrait pas être dans deux bus en même temps (ou à l'intérieur d'un certain intervalle);
- ne devrait pas avoir un quart de travail dérivé par tournée (amplitude) ou par jour (pièce de travail) dépassant un certain laps de temps;

L'application de ces logiques aux données conduit à certains constats.

Chauffeur simultanément dans deux véhicules

On retrouve des séquences problématiques de transactions simultanées, dans deux bus différents, qui sont associées au même chauffeur.

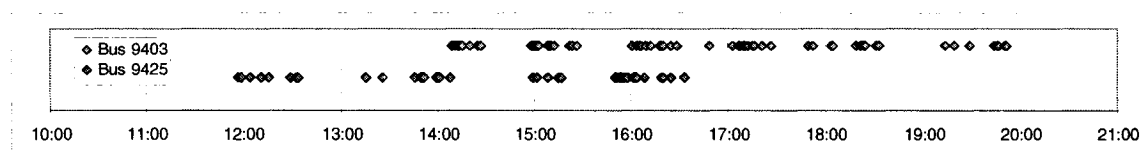


Figure 2.19 – Étalement temporel des transactions des bus 9403 et 9425, le 31 juillet 2003 (même chauffeur)

La cause exacte de ce type d'erreur est inconnue mais elle n'est survenue qu'une seule fois dans le mois étudié. Il pourrait s'agir de l'erreur d'un des deux chauffeurs qui, au moment de la prise du véhicule, aurait indiqué le numéro d'identification d'un autre chauffeur (en service à ce moment).

Durée des pièces de travail et amplitudes

On estime le temps de travail d'un chauffeur à l'aide de l'heure d'embarquement des premières et dernières transactions d'une tournée ou d'une journée. La Figure 2.20 illustre la juxtaposition des durées des pièces de travail par chauffeur-jour-bus avec les durées des amplitudes par chauffeur-jour. La grande majorité (95 %) des pièces de travail sont d'une durée de moins de 7h30.

Il est intéressant de constater le chevauchement des durées autour de 7h00, qui signifie que les durées de cet ordre s'effectuent avec un seul véhicule. Cinq pièces de travail vont au-delà de 14h, soit une de 17h30 et quatre autour de 15h20. Ces dernières sont quasi identiques, étant caractérisées le même jour de semaine (dimanche), le même chauffeur et presque la même durée. Ils ont ainsi une régularité qui laisse présager une situation ordinaire.

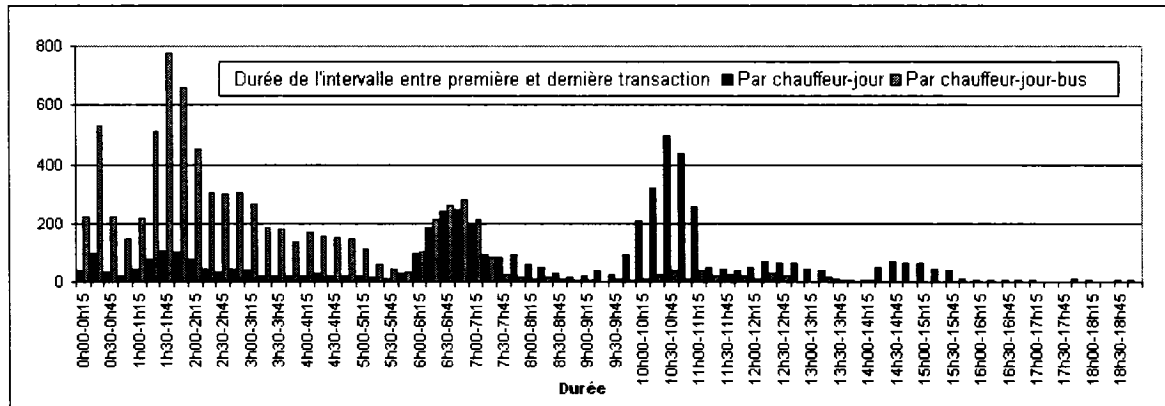


Figure 2.20 – Distribution (aux quinze minutes) des intervalles entre premier et dernier embarquement, par chauffeur-jour-bus (pièces de travail) et chauffeur-jour (amplitudes) (juillet 2003)

Cette régularité est illustrée à la Figure 2.21, qui présente l'étalement temporel des transactions de ces cinq plus grandes pièces de travail rencontrées.

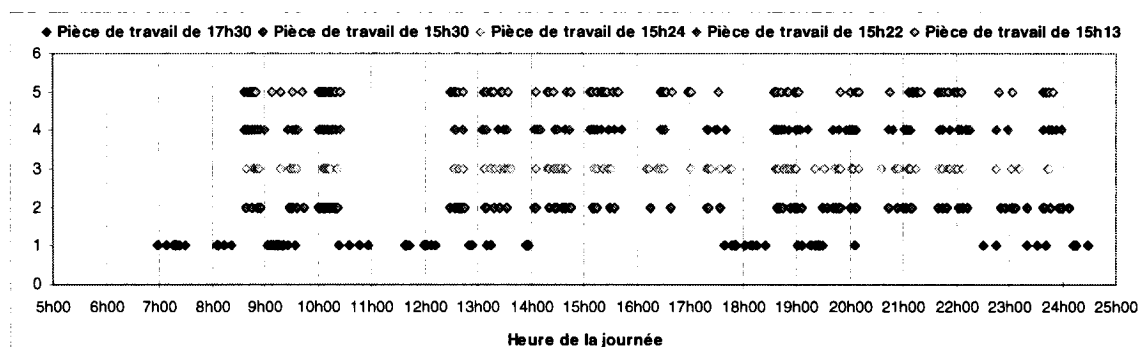


Figure 2.21 – Étalement temporel des transactions pour les cinq pièces de travail dérivées les plus importantes (juillet 2003)

Dans le cas où l'amplitude est de 17h30, les regroupements de transactions pourraient laisser présager qu'il s'agit d'une erreur d'identification où un deuxième chauffeur aurait omis de s'identifier lors du changement de chauffeur et que cette pièce de travail constitue en fait l'assemblage de deux pièces de travail concernant deux chauffeurs.

2.5.4 Assignment

En pouvant isoler et suivre certains ensembles d'enregistrements par leur assignation (tracé-sens-départ), des logiques peuvent s'appliquer :

- un départ ne devrait être assigné qu'à un seul véhicule;
- une transaction ne devrait pas se produire trop longtemps avant (ou après) l'heure de départ planifiée du départ assigné;
- l'arrêt dérivé d'une transaction devrait faire partie de la séquence d'arrêts du tracé-sens assigné;
- après avoir déduit les séquences des arrêts des tracé-sens, les arrêts d'embarquement sur un tracé-sens devrait respecter l'ordre séquentiel des arrêts de ce tracé-sens.

Étant donné le nombre important d'erreurs d'assignation, explicitées précédemment, ces logiques n'ont pas été extensivement testées. À cet effet, une première étape vers la correction de ces assignations est entreprise au chapitre suivant.

Deux véhicules sont associés au même départ

On a pourtant observé plusieurs (69) départs étant présents plus d'une fois à l'intérieur d'une même journée. La majorité (41) de ces occurrences concerne deux tracés-sens-départ (39-0-0655 et 67-1-1520) qu'on retrouve sur presque tous les jours de semaine avec deux véhicules et deux chauffeurs, ce qui laisse supposer que le service est planifié de cette façon.

Parmi les autres (28) occurrences examinées, toutes renferment des transactions qui, même si elles se retrouvent sur deux véhicules et/ou deux chauffeurs, sont chronologiquement distinctes (non entrelacées); donc il semble s'agir de situations où l'on a procédé à des remplacements ponctuels de véhicule et/ou de chauffeur.

CHAPITRE 3. CARACTÉRISATION DE L'OFFRE DE SERVICE

Le propos de ce chapitre est de montrer qu'à partir des déplacements de la clientèle, on peut dériver de l'information détaillée sur l'offre de service. À cette fin, il serait possible de caractériser presque complètement une ligne de bus en n'utilisant que les données d'embarquement. Toutefois, ce chapitre se limitera à présenter deux méthodologies susceptibles de bonifier notre connaissance de l'offre.

La première s'attardera sur les correspondances effectuées par la clientèle. Il est ici question de caractériser le chaînon constituant le transfert entre une première validation et sa correspondance. La validité de ces chaînons est aussi mise à l'épreuve par des logiques spatio-temporelles. L'objectif à long terme de cette caractérisation est d'aboutir à une modélisation réseau. La deuxième méthode présentée utilisera les mêmes données d'embarquement afin de déterminer l'horaire complet pour un jour de semaine, à travers la régularité des chaînes de départs (tournées) rencontrées. Avec les fruits de cette méthode, il peut être possible de revenir sur ces données d'embarquement et de corriger les assignations jugées erronées au chapitre précédent.

3.1 Évaluation des correspondances réalisées

C'est dans une perspective de modélisation réseau que l'on désire approfondir notre connaissance des correspondances effectuées par la clientèle. Un réseau simplifié et codifié peut permettre de calibrer un modèle d'affectation et d'aborder une tentative d'optimisation de ce réseau. Les données opérationnelles de type carte à puce, par l'entremise de l'information sur les déplacements effectués, nous fournissent l'opportunité d'effectuer cette codification.

Ainsi, le fait qu'un enregistrement fasse état de l'embarquement par un usager à un certain arrêt confirme l'existence de cet arrêt en tant que point d'accès au réseau. De la même manière, une correspondance réalisée par un usager entre les arrêts de deux lignes exprime qu'un transfert est possible (et dorénavant permis) entre ces deux lignes, via ces arrêts. On peut ainsi concevoir un métaréseau regroupant les nœuds des arrêts où se produisent des correspondances. Ce nouveau réseau simplifié conserve les éléments d'information du réseau original via une table de concordance entre les nœuds et les arrêts couverts.

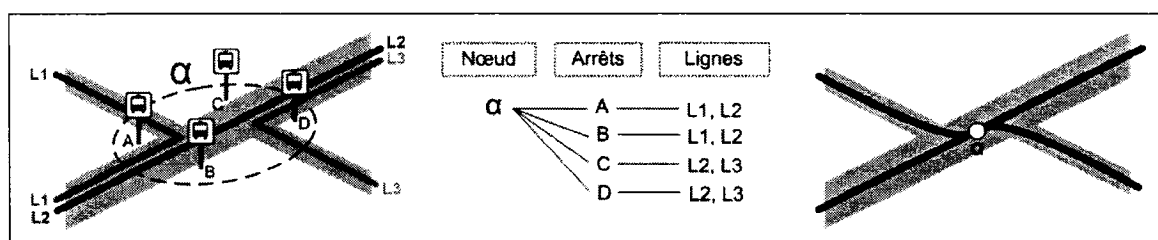


Figure 3.1 – Simplification des arrêts du réseau

Cette section entreprend donc d'aborder les premières étapes de la modélisation réseau en confrontant les correspondances réalisées avec des logiques de validations dans le but de les homologuer ou de faire ressortir les ambiguïtés qu'elles pourraient comporter.

3.1.1 Méthodologie

Plusieurs étapes sont nécessaires afin de caractériser systématiquement chacune des correspondances. On présente ici la méthodologie, exécutée via un script VBA (présenté à l'annexe D).

Extraction de l'échantillon

On a vu au chapitre précédent que les données opérationnelles de la STO comportent un champ indiquant le type de transaction (validation simple, correspondance ou refus). En filtrant les enregistrements du mois de juillet 2003,

on en extrait 59 674 correspondances (15 % des 387 588 transactions), qui composeront notre échantillon pour le reste de cette section.

Identification de l'embarquement précédent

En utilisant le champ ajouté décrivant la séquence des transactions par usager-jour (décrit à la section 2.4.2), le script procède facilement à l'identification de la validation simple précédant chaque correspondance. L'information retenue concerne la temporalité de l'embarquement, son assignation (tracé-sens-départ) et son numéro d'arrêt.

Identification de l'arrêt de débarquement

Maintenant que l'on connaît le tracé-sens des deux embarquements consécutifs, on utilisera une fonction suivant la ligne de fuite et qui estime l'arrêt de descente. Cette fonction prend comme entrée le tracé-sens et l'arrêt du 1^{er} embarquement (Ligne A) ainsi que l'arrêt du 2^e embarquement (Ligne B). À l'aide de la table de référence des arrêts-ligne, elle détermine l'arrêt subséquent (à l'arrêt d'embarquement) de la ligne A le plus rapproché de l'arrêt de la ligne B où a eu lieu la correspondance. La fonction (aussi présentée à l'annexe D) retourne donc l'arrêt de descente le plus probable ainsi que la distance (euclidienne, donc minimum) que l'utilisateur aurait théoriquement marché pour correspondre (Figure 3.2).

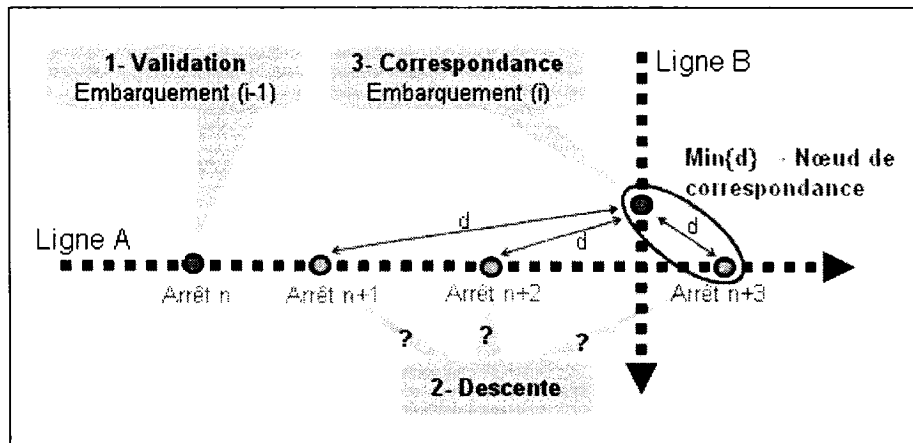


Figure 3.2 – Détermination de l’arrêt de descente le plus probable lors d’une correspondance

Tests de validation

À mesure que le script s’exécute, les correspondances subissent plusieurs tests de validation. Ces tests, indépendants et non exclusifs, s’articulent autour des principes suivants :

- La transaction de correspondance doit être précédée d’une transaction de validation
- Les arrêts de validation et de correspondance doivent faire partie de leurs tracés-sens respectifs
- Les tracés-sens doivent être valides
- L’arrêt de validation ne doit pas être le dernier arrêt du tracé-sens
- Les arrêts ou tracés doivent être valides
- La distance de transfert dérivée ne devrait pas être supérieure à 1km
- Le laps de temps entre l’embarquement de validation et de correspondance devrait être inférieur à 2h

Chaque test manqué verra la correspondance se faire ajouter le code attribué à ce test.

3.1.2 Résultats

Selon les tests échoués, le traitement affecte ainsi une ou plusieurs lettres aux correspondances. Près de 80 % des correspondances sont ainsi déclarées « valides », n'ayant échoué aucun test. La Figure 3.3 illustre la répartition des tests (anomalies) détectés sur les correspondances ainsi que la distribution des distances entre les paires d'arrêts (descente/correspondance) issue de correspondances valides. Le Tableau 3.1 précise la légende utilisée pour chaque anomalie.

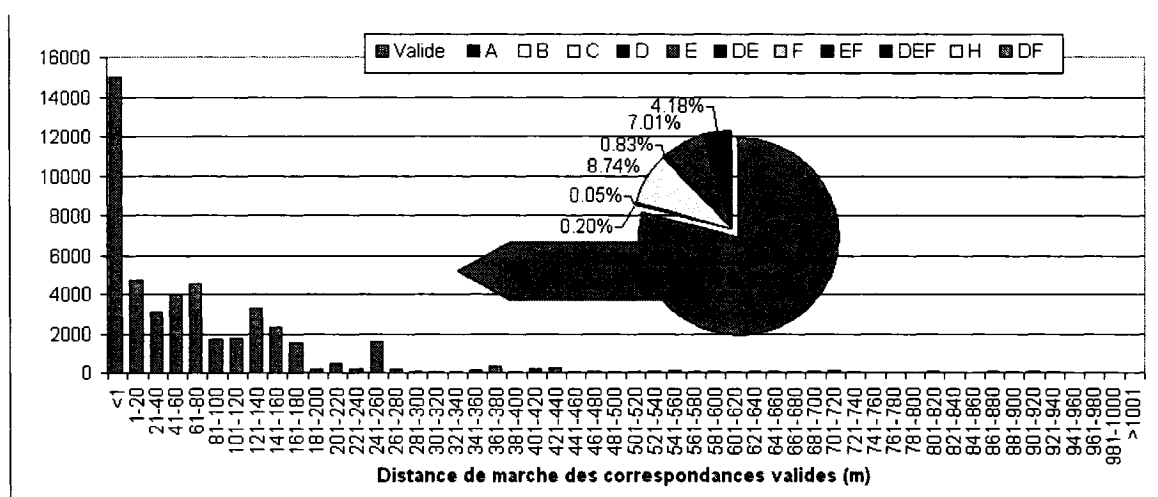


Figure 3.3 – Répartition des correspondances par code d'anomalie, ainsi que distribution de la distance entre paires d'arrêts issues de correspondances valides

Tableau 3.1 – Légende des codes d'anomalie et de leur répartition

Code	Type d'anomalie	%	Code	Type d'anomalie	%
Valide	CORRESPONDANCE VALIDE	78.81%	F	Délai avant correspondance non-inférieur à 2h	0.11%
C	Invalidité d'un arrêt ou tracé-sens	8.74%	A	Arrêt de validation inconnu au tracé-sens	0.05%
E	Distance dérivée supérieure à 1km	7.01%	EF	Anomalies E et F	0.03%
DE	Anomalies D et E	4.18%	H	Le tracé-sens est inconnu à la table des arrêts-ligne	0.02%
D	Arrêt de validation est le dernier arrêt du tracé-sens	0.83%	DEF	Anomalies D, E et F	0.01%
B	Pas de transaction précédant cette correspondance	0.20%	DF	Anomalies D et F	0.00%

Élaborons sur les trois tests qui causent les quatre principales catégories d'anomalies formant 98% des correspondances non valides. Ces tests sont tous reliés à la même cause.

- Invalidité d'un arrêt ou tracé-sens. Tel qu'explicité au chapitre précédent, on retrouve des tracés-sens ou arrêts invalides lorsque le chauffeur oublie d'indiquer le début d'un départ;
- Distance dérivée supérieure à 1km. Une distance de marche dérivée trop élevée est habituellement reliée aux situations où le chauffeur oublie d'indiquer le début d'un départ. Puisque l'assignation erronée contamine l'imputation des arrêts, on arrive à des distances dérivées incongrues entre les arrêts de correspondance;
- L'arrêt de validation est le dernier arrêt du tracé-sens. On a vu que l'arrêt de validation est le dernier arrêt du tracé-sens lorsque le chauffeur oublie d'indiquer le début d'un départ. L'ordinateur de bord s' imagine alors toujours sur le départ précédent et a de bonnes chances d'imputer le dernier arrêt de ce départ, alors que l'embarquement s'effectue sur le nouveau départ.

D'autre part, pour les correspondances valides, on retrouve que 32 % (25 % du total) ont leur distance de transfert égale à zéro signifiant que la correspondance se fait au même arrêt. En termes de nœuds de correspondance, c'est 15 % des paires d'arrêts qui sont en fait constituées d'un arrêt unique.

On remarque aussi que 70 % des correspondances valides ont un segment de transfert (distance euclidienne entre les paires d'arrêts) sous 100 m, 90 % sous 200 m et 95 % sous 300 m. La Figure 3.4 représente géographiquement les 2262 segments de transfert ainsi que leur intensité, représentée par le nombre de correspondances effectuées dans le mois sur le segment. Le plus grand vecteur de correspondance se situe entre deux arrêts des Promenades de l'Outaouais, où convergent plusieurs tracés, avec une intensité de 3269 correspondances dans le mois (7 % des correspondances valides).

On a vu qu'afin d'augmenter significativement le ratio de correspondances valides, il serait souhaitable de corriger au préalable les assignations erronées des transactions. Cette correction est amorcée à la section suivante.

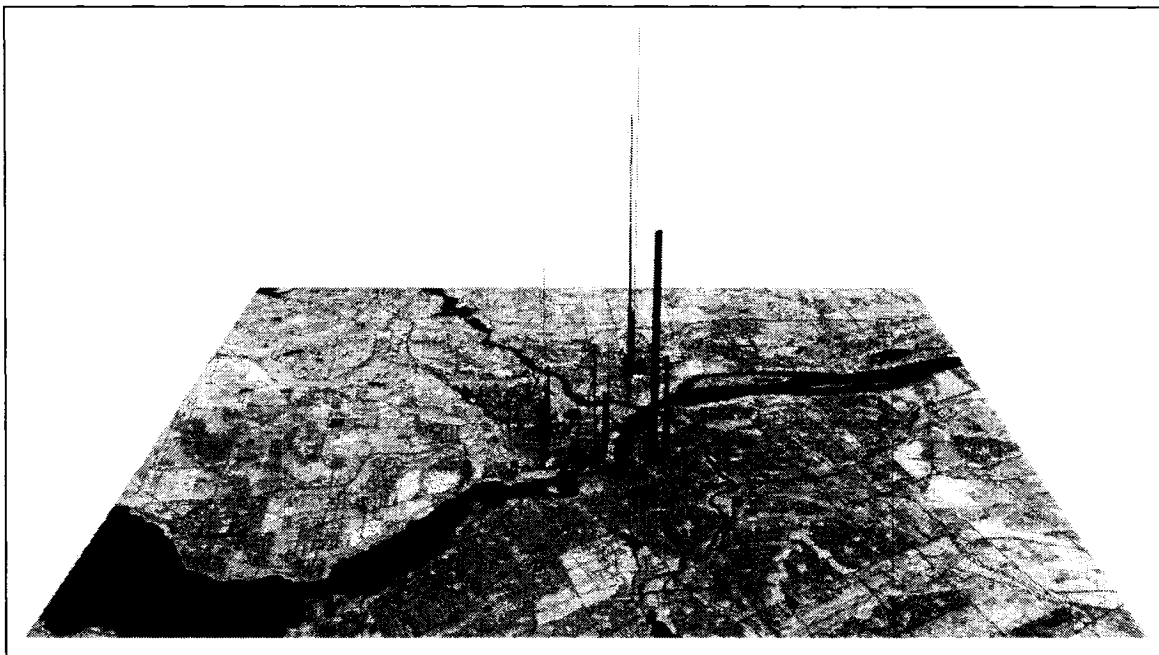


Figure 3.4 – Segments de transfert entre paires d'arrêts avec leurs intensités respectives (juillet2003)

3.1.3 Simplification du réseau

Revenant à la modélisation réseau évoquée en début de section, il serait ainsi possible de regrouper les d'arrêts des segments de transfert, en nœuds de correspondance, permettant de simplifier le réseau de nœuds dans le but, ultime, d'effectuer une modélisation réseau.

Cependant, dans le but d'avoir un réseau plus efficace, une étape supplémentaire consiste à agréger les arrêts « rapprochés », qui pourraient constituer un nœud. Afin d'avoir une vue d'ensemble sur la distribution de ces arrêts de proximité, la Figure 3.5 illustre le résultat du calcul des 13 309 distances inférieures à 500 m entre les arrêts.

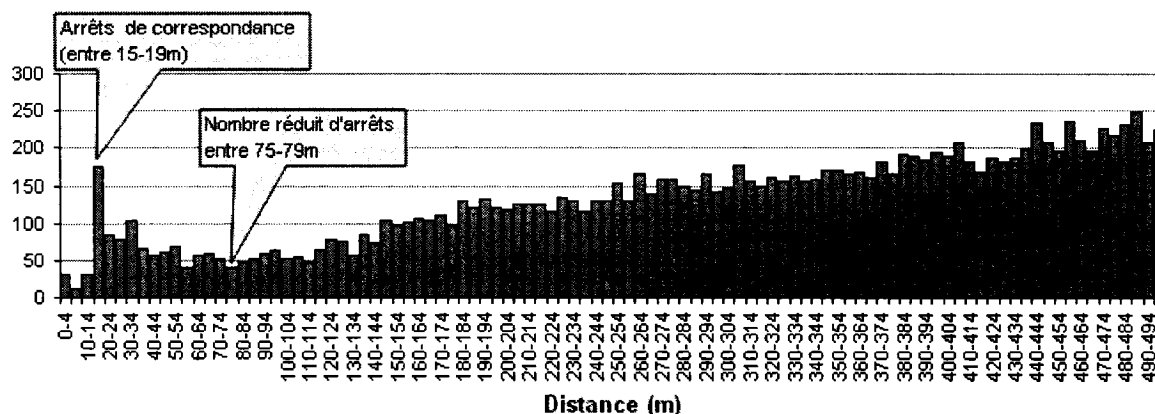


Figure 3.5 – Distribution des distances (< 500 m) entre tous les arrêts

La démarcation des arrêts distants de 15-19 m représente vraisemblablement une géométrie où les deux arrêts sont du même carrefour. D'autre part, on note un creux pour les arrêts distants de 75-79 m. Cette distance semble marquer une charnière entre les arrêts permettant facilement une correspondance et les autres arrêts. Un seuil de 75 m est ainsi utilisé comme seuil d'agrégation afin de regrouper les arrêts de proximité en nœuds (Figure 3.6).

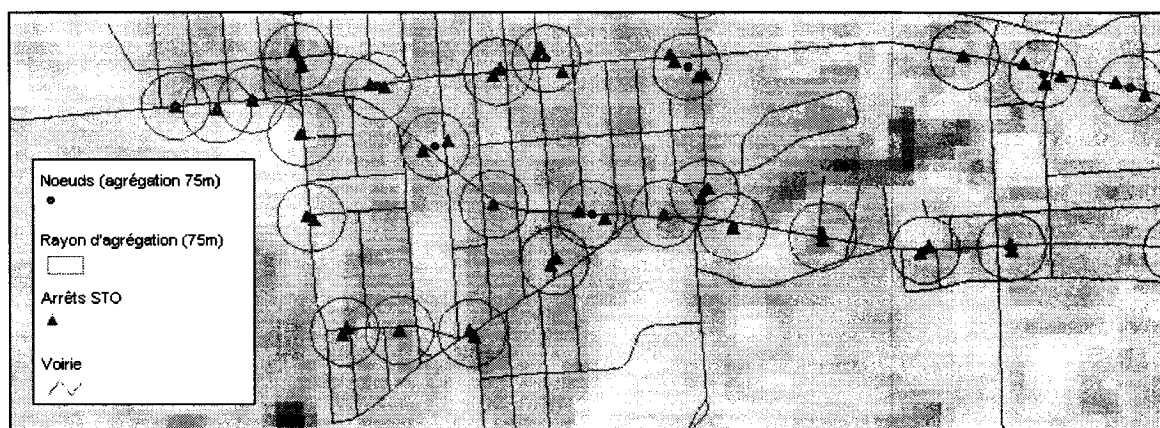


Figure 3.6 – Détail de l'agrégation à 75 m des arrêts

Cette agrégation spatiale, couplée avec l'agrégation de correspondance décrite plus tôt, permet une simplification de la codification du réseau. Cependant, afin

de compléter cette simplification, la déduction des arrêts de descente s'avérerait nécessaire.

3.2 Déduction du service réalisé

On a vu que l'assignation erronée des enregistrements est la plus grande source d'anomalies dans les données étudiées. L'objectif à long terme de la démarche présentée dans cette section est ainsi de corriger l'assignation erronée en déduisant le service planifié via la régularité des tournées. Bien qu'une méthode de correction de l'assignation via la régularité des déplacements des usagers ait été réalisée par Tranchant (2005), la qualité de cette méthode varie selon la constance de l'usager et de son nombre de déplacement. La présente section utilise plutôt l'hypothèse selon laquelle un jour de semaine est présumé comporter un horaire fixe afin de retrouver la régularité des chaînes de départs à travers l'information des embarquements réalisés, puis de reconstruire l'horaire des tournées.

3.2.1 Méthodologie

Nous prenons ainsi comme hypothèse que le service est le même pour tous les jours de semaine⁵. On aspire à ce qu'une régularité soit observée sur le plan des séquences de départs (tournées). Autrement dit, à chaque jour, un certain bus sera assigné à une séquence de départs (tracés-sens-départ ou TSD) prévue au service planifié.

La méthodologie présentée part de la perspective du départ afin de reconstituer le service. La première étape consiste à établir une table de tous les départs actifs assignés aux enregistrements du mois (juillet 2003). Un script VBA

⁵ Par mesure d'allègement, le reste de cette section utilisera dorénavant « jour » pour « jour de semaine »

(présenté à l'annexe E) parcourt cette table et caractérise chaque départ, à l'aide des enregistrements d'embarquement, selon plusieurs attributs décrits plus loin. La méthode consiste essentiellement, pour chacun de ces départs, à déterminer quels sont les départs précédant et suivant les plus habituels.

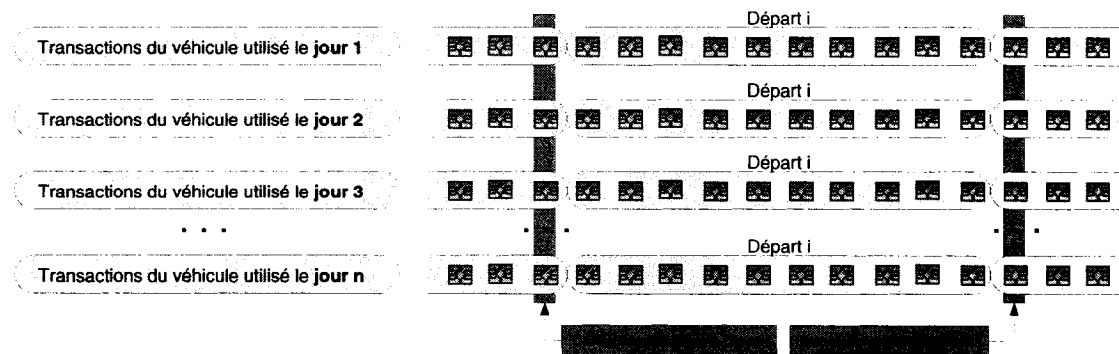


Figure 3.7 – Identification du départ précédent et suivant le plus fréquent au départ i, sur les jours où il apparaît dans les transactions

Ainsi, pour un départ donné, on recherche donc tous les jours où ce départ apparaît et on observe les départs précédents et subséquents à l'intérieur de la même tournée de ce bus.

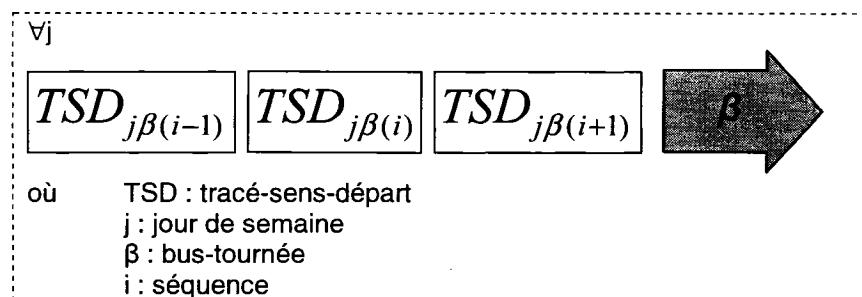


Figure 3.8 – Section d'un bus-tournée type

D'un point de vue procédural, on a :

Pour chaque tracé-sens-départ de semaine;

- Pour chaque jour où ce tracé-sens-départ apparaît;
 1. Détermination du premier [et dernier] embarquement de ce départ;

2. Identification du bus utilisé;
 3. Pour ce bus et cette tournée, détermination de l'embarquement précédent [et subséquent] au tracé-sens-départ concerné;
 4. Identification du tracé-sens-départ de cet embarquement.
- Analyse de la régularité des tracés-sens-départ rencontrés;
 - Calcul d'autres indicateurs tels durée⁶ et régularité du chauffeur.

Théoriquement, le tracé-sens-départ précédent [subséquent] devrait toujours être le même puisque nous émettons l'hypothèse que le service est identique (les séquences de départs sont les mêmes) pour tous les jours. Toutefois, sur le plan des données opérationnelles, le tracé-sens-départ précédent [subséquent] « probable » sera établi en fonction de sa régularité. C'est-à-dire par la fréquence de ce départ au nombre de jours où le départ de référence est apparu. Notons qu'un minimum d'un embarquement par départ est requis afin que ce départ « apparaisse » dans les données.

Table des tracés-sens-départ

Ainsi, pour chaque départ, on établit le tracé-sens-départ précédent [et suivant] le plus probable ainsi que d'autres indicateurs concernant la durée et le chauffeur en poste. Ces indicateurs viennent caractériser une table de données basée sur chaque départ apparaissant dans les données d'embarquement (Tableau 3.2). Chaque enregistrement équivaut ainsi à un TSD présent dans les données d'embarquement qui est identifié par une clé primaire. Le TSD est caractérisé selon plusieurs indicateurs, dont celui du nombre de jours où on le retrouve (**nbSurMois**). Ce nombre, utilisé comme pondérant, sera ainsi égal ou

⁶ La durée est définie ici comme étant l'intervalle de temps entre l'heure du départ planifiée et la dernière montée de ce départ.

inférieur à 22, soit le nombre de jours de semaine du mois de juillet 2003 (le 1^{er} juillet est un jour férié).

Pour chaque jour où apparaît le TSD, on calcule la différence temporelle entre la dernière transaction marquée de ce départ et de l'heure de départ planifiée. Le champ **DureeMoy1** consiste en la moyenne de ces durées. Cependant, étant donné que les erreurs d'assignation vues au chapitre précédant, causée par le non-signallement d'un changement de départ, risquent de gonfler la durée moyenne des départs de manière significative, une deuxième moyenne **DureeMoy2** est calculée en ne tenant pas compte des jours où la durée est supérieure (ou inférieure) à un écart type de DureeMoy1. **nbMoy2** représente ainsi le nombre de jours ayant servi au calcul de DureeMoy2.

Tableau 3.2 – Clichés d'enregistrement et description des champs de la table des tracés-sens-départ (TSD)

Champ	Ex 1	Ex 2	Ex 3	Description	Sujet
TSD_ID	24	996	129	Numéro unique du tracé-sens-départ (le premier)	TSD de référence
Date	82	37	89	Date	
Heure	1	0	1	Heure	
Départ	1700	0000	1637	Nom du départ, possible	
EnSan	1	1	1	Le tracé-sens-départ est-il en Sanjour?	Infos sur TSD de référence
Vehic	1	0	1	Le véhicule est-il un véhicule de transport de personnes?	
Surp	0	1	0	Le tracé-sens-départ est-il sur un pont?	
NbSurMois	22	14	21	Nombre de jours où le TSD apparaît dans le mois	
DureeMoy1	24.41	0	23.43	Durée moyenne des départs (en heures) pour le TSD	
DureeMoy2	20.57	0	21.5	Durée moyenne des départs (en heures) pour le TSD, excluant les jours où la durée est supérieure ou inférieure à un écart type de DureeMoy1	
NbMoy2	21	14	18	Nombre de jours où le TSD apparaît dans le mois, excluant les jours où la durée est supérieure ou inférieure à un écart type de DureeMoy1	
Chaqueur	3398	3166	3188	Chaqueur assigné au départ du TSD (en heures)	
ChSurMois	4	6	11	Chaqueur assigné au départ du TSD (en heures)	Infos sur TSD précédent
PR_Index	21	1849	774	Traceur unique du TSD précédent	
PR_TSD	82_1_1549	81_0_0604	30_0_1600	Traceur unique du TSD précédent	
PR_T_SurNb	20	6	19	Nombre de jours où le TSD précédent apparaît dans le mois	
St_Index	1630	932	0	Traceur unique du TSD suivant	Infos sur TSD suivant
St_TSD	67_1_1807	35_0_0800	DernierTSD	Traceur unique du TSD suivant	
St_T_SurNb	19	4	19	Nombre de jours où le TSD suivant apparaît dans le mois	

Toujours pour chaque jour où apparaît le TSD, on isole le véhicule utilisé cette journée et on note l'assignation de l'embarquement précédant la première transaction du TSD étudié. Pour l'ensemble des jours, l'assignation précédente la plus fréquente est ainsi identifiée par **PR_TSD** (le tracé-sens-départ) et **PR_Index** (l'identificateur unique). D'ailleurs, **PR_T_SurNb** représente le nombre de jours où l'on retrouve cette assignation précédente, et peut se comparer à NbSurMois, le nombre de jours où on retrouve le TSD étudié.

Lorsque l'enregistrement précédant la première transaction du TSD étudié n'existe pas, le TSD est assumé « premier TSD » de la tournée. C'est d'ailleurs l'appellation utilisée pour le champ PR_TSD (PR_Index sera 0) dans le cas où cette situation est la plus fréquente.

De la même manière, on détermine l'assignation suivante la plus fréquente est caractérisée par **SU_TSD**, **SU_Index** et **SU_T_SurNb**. Finalement, le chauffeur le plus fréquent (**Chauffeur**) ainsi que le nombre de fois où il apparaît (**ChSurMois**) est aussi déterminé. À l'aide de ces données, on est donc en mesure de reconstituer les chaînes de départs (tournées) planifiées et d'en identifier certains maillons douteux ou problématiques.

3.2.2 Résultats

Des 1856 tracés-sens-départ de semaine identifiés, on en retrouve 1381 qui s'avèrent valides et qui constituent l'échantillon étudié ici. Une régularité élevée d'apparition est observée. En effet, plus de la moitié des départs sont présents plus de 20 fois sur 22 jours de semaine (régularité de 91 %) et 75 % se retrouvent sur au moins 17 des 22 jours (régularité de 77 %) (Figure 3.9).

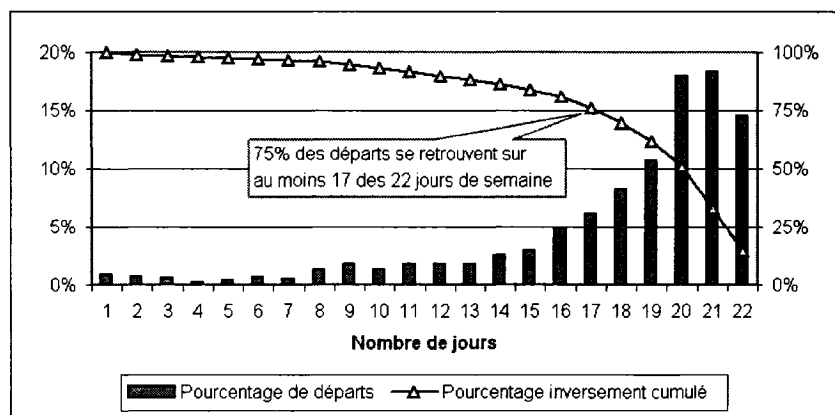


Figure 3.9 – Distribution des départs selon nombre de jours apparus (juillet 2003)

Le fait d'obtenir un faible nombre de jours pour un départ est attribuable aux erreurs d'assignation où, plus souvent qu'autrement, un TSD se fait

systématiquement « oublier » par le(s) chauffeur(s) assigné(s) à sa réalisation. Il est aussi possible que des départs sans embarquement soient courants à certaines heures, possiblement lorsqu'à contresens des flots de déplacement.

Régularité des attributs

Plusieurs mesures de régularité sont mesurées pour chaque TSD. Bien que les plus intéressants demeurent le TSD précédant et le TSD suivant, d'autres mesures sont considérées comme la régularité du chauffeur ainsi que la régularité de la durée du départ. La mesure de la régularité est exprimée par le nombre d'occurrences de l'attribut le plus fréquent du TSD sur le nombre d'occurrences totales de ce TSD. Ces mesures de régularité expriment une quantification quant à la valeur de l'élément dérivé.

Dans le cas de l'indicateur utilisé pour estimer la régularité de la durée dérivée, on utilise le nombre de valeurs retenues lors du calcul de *DureeMoy2*. Rappelons qu'une fois que la durée moyenne de tous les jours concernés est calculée (*DureeMoy1*), l'écart-type est aussi calculé afin d'arriver à une autre moyenne en ne conservant que les durées comprises à l'intérieur de cet écart-type. La régularité de la durée moyenne est donc définie comme le ratio du nombre de valeurs conservées sur le nombre de valeurs (de jours) total. Une régularité de durée parfaite (100 %) représente ainsi qu'aucune des journées où le TSD a été observé ne s'est démarquée quant à la durée calculée (entre l'heure du dernier embarquement sur ce TSD et l'heure de départ planifiée).

La Figure 3.10 illustre la répartition des TSD en regard de la régularité des attributs mentionnés, soit la régularité du départ précédent, du départ suivant, du chauffeur et de la durée. On peut apprécier que 82 % des départs précédents, 79 % des départs suivants et 95 % des durées dérivées, détiennent une régularité supérieure à 75 %. Par contre, la régularité des chauffeurs n'est pas du même niveau, ce qui est conforme à leur type d'horaires (semaine de quatre jours, vacances, etc.).

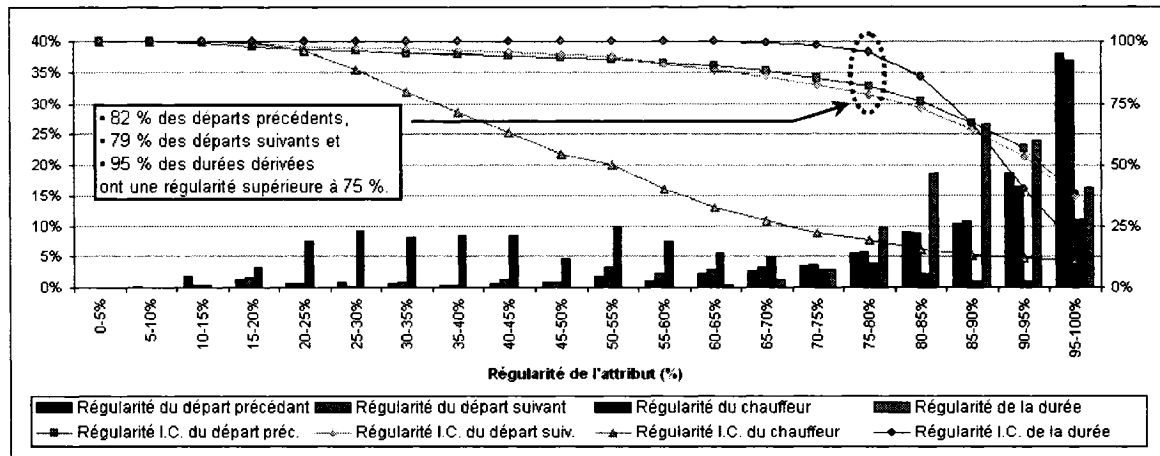


Figure 3.10 – Distribution des départs selon régularité et régularité inversement cumulée (I.C.) du départ précédent, du départ suivant, du chauffeur et de la durée (juillet 2003)

Même si certains TSD observés sur un petit nombre de jours peuvent introduire un certain biais (les attributs d'un TSD observé sur un seul jour auront une régularité de 100 %), nous avons vu à la Figure 3.9 que cette situation n'est pas fréquente.

Parmi les éléments de cette nouvelle base de données du service déduit sur lesquels on pourrait s'interroger, quelques détails sur le niveau de concordance bidirectionnelle des départs précédant et suivant. En effet, lorsqu'on a un TSD_x, précédant d'un TSD_y, est-ce que ce TSD_y sera nécessairement le suivant du TSD_x? Après vérification, on retrouve une concordance de 97 %. Le pourcentage manquant est expliqué par la définition de « plus fréquent » de l'algorithme. En effet, lorsque deux éléments surviennent le même nombre de jours (les deux sont le plus fréquent), l'algorithme n'est pas nécessairement conséquent dans les deux « situations miroir » (détermination du TSD précédant dans un cas et détermination du TSD suivant dans l'autre cas).

La table des TSD est riche en information dérivée. En effet, joignant les logiques spatio-temporelles du chapitre précédant, elle recèle un potentiel intéressant pour débusquer les enregistrements suspects. En effet, elle permet d'estimer la durée maximale d'un départ, au-delà de laquelle un embarquement serait perçu

automatiquement comme possédant une assignation douteuse. De plus, à la lueur du nombre de fois où un TSD apparaît dans le mois, on connaît les départs planifiés manquants.

3.2.3 Vers une correction de l'assignation

En utilisant la table des TSD, il est possible de relier le départ suivant déduit pour chaque TSD. L'ensemble du service d'un jour de semaine peut ainsi être exprimé par 292 chaînes de départs, équivalent à autant de tournées, constituées en moyenne de 4,3 départs se distribuant inégalement au cours de la journée (Figure 3.11).

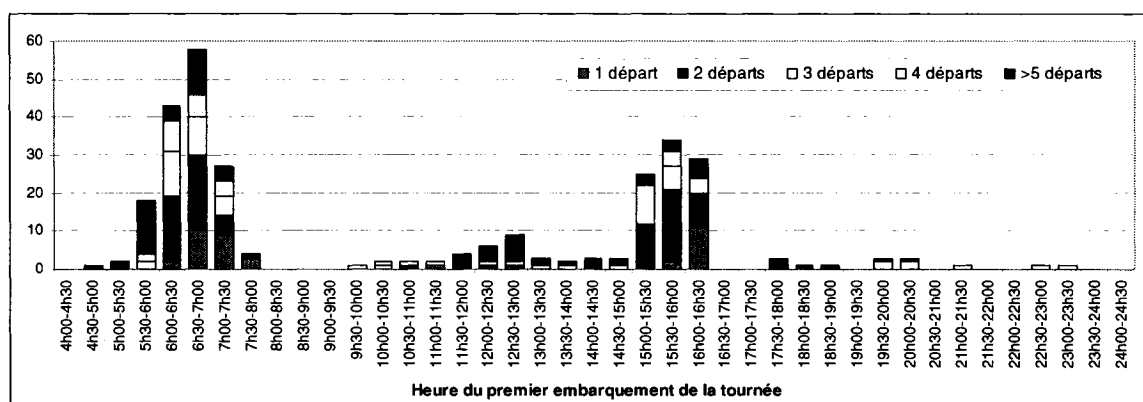


Figure 3.11 – Distribution des débuts des tournées selon l'heure du 1^{er} embarquement, ventilé par le nombre de départs par tournée

Avec le service déduit comme une suite de départs planifiés, une correction des enregistrements comportant une assignation erronée, tel que vu au chapitre précédant, est envisageable. En effet, dès que l'on détecte que l'assignation des enregistrements d'une tournée devient erronée suite à l'oubli d'un chauffeur, cette assignation peut être corrigée pour chaque enregistrement concerné, en se fiant à la suite de départs prévus ainsi qu'à l'heure d'embarquement. On vient donc « calquer » le service planifié sur les enregistrements possédant une assignation erronée. Un élément probabiliste peut être incorporé aux corrections, d'après les pourcentages de régularité calculés plus tôt dans cette section.

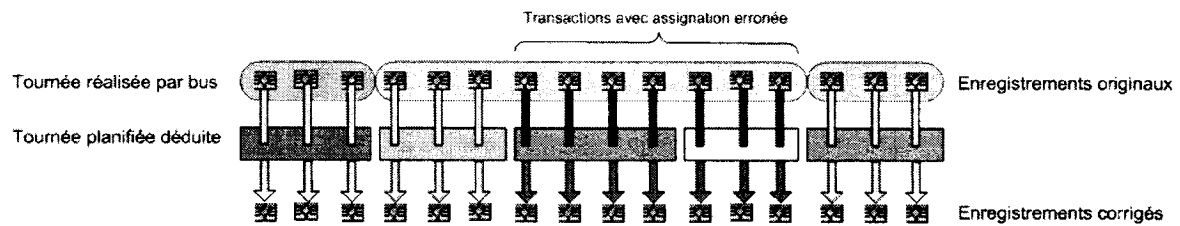


Figure 3.12 – Schéma illustrant la correction des transactions (possédant une assignation erronée) par le service déduit

Ultimement, une combinaison de la méthode de correction par la régularité des tournées avec celle de la régularité de l'utilisateur peut mener à un renforcement de la qualité de la correction lors d'une corroboration. Bien que la régularité de l'utilisateur puisse être moins stable que celle du service planifié, cette méthode comporte l'avantage de pouvoir estimer l'arrêt utilisé par l'utilisateur.

CHAPITRE 4. CARACTÉRISATION DE LA DEMANDE

Ce chapitre analyse diverses facettes de la demande de transport à travers des indicateurs dérivés des comportements réalisés. On présente ainsi la méthode utilisée, suivie de résultats à travers plusieurs indicateurs dérivés. Divers types de variabilité sont ainsi mis en évidence avec l'application de la méthode sur trois ensembles de données, soit les mois de juillet 2003, octobre 2003 et février 2005. Enfin, une analyse de type longitudinale des cartes se retrouvant dans les trois ensembles de données est décrite.

4.1 Détermination des profils d'usager

On souhaite caractériser les profils des usagers à travers les embarquements réalisés. Pour ce faire, une synthèse informationnelle de l'ensemble des transactions de chaque carte active est effectuée et est exprimée à travers un enregistrement qui résume les comportements réalisés de l'individu affilié à la carte.

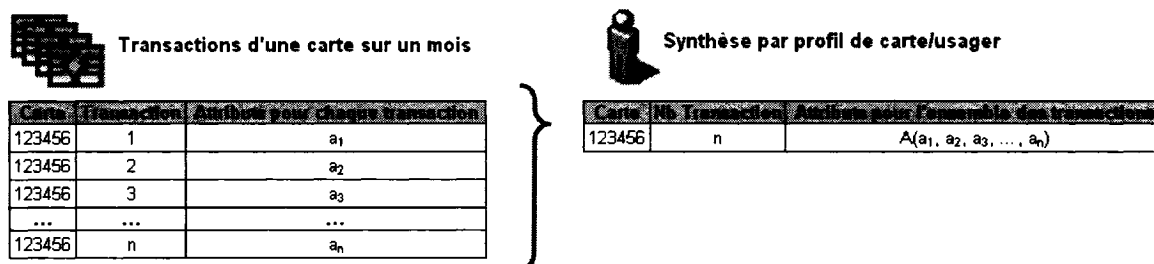


Figure 4.1 – Processus de synthèse de l'information d'une carte/usager

Ces profils sont exprimés à travers des attributs déduits par un algorithme qui prend comme intrant les données d'embarquement des différents ensembles de données. Cette section présente ces nouveaux champs, la méthodologie utilisée pour les composer ainsi que divers résultats en découlant.

4.1.1 Attributs et méthodologie

Afin de représenter les profils des usagers, une nouvelle table est utilisée avec divers champs synthétisant les transactions effectuées par une carte. On attribue ainsi à chaque carte présente dans au moins un des ensembles de données étudiés, un arrêt de résidence dérivé, sa distance au centre ainsi que plusieurs autres attributs, présentés globalement au Tableau 4.1.

Tableau 4.1 – Clichés d’enregistrement et description de la table des profils, caractérisant les cartes en service

Champ	Ex1	Ex2	Description de champ	Champs identiques pour les autres mois
NoCarte	1644860741	2196668353	Clé primaire : Numéro de la carte	
Juil03	1	1	La carte a été utilisée dans ce mois	
Oct03	0	0	La carte a été utilisée dans ce mois	
Fev05	1	0	La carte a été utilisée dans ce mois	
Groupe0307	r	e	Catégorie du titre [Régulier, Etudiant, Aîné]	
RatioGroupe0307	100	94	Fiabilité du groupe (%)	
ArretDomicile0307	2322	2188	Arrêt-domicile dérivé	
RatioArret0307	97	25	Fiabilité de l'arrêt-domicile (%)	
Eloignement0307	6467	6995	Distance au centre (m)	
NbTrans0307	71	16	Nombre de transaction dans le mois	
NbValid0307	59	10	Nombre de validations simples dans le mois	
NbCorr0307	12	4	Nombre de correspondances dans le mois	
NbRefus0307	0	2	Nombre de transactions refusées dans le mois	
DeplJour0307	1.90	0.32	Nb de validations simples / Nb de jours dans ce mois	
DeplSem0307	2.05	0.45	Nb de validations simples en semaine / Nb de jours de semaine dans ce mois	
RatioCV0307	0.20	0.40	Nb de correspondances / Nb de validations simples dans ce mois	
RatioSemCV0307	0.18	0.40	Nb de corr. en semaine / Nb de validations simples en semaine dans ce mois	

La majorité des champs de la table sont répétés pour chaque ensemble de données analysé, leur libellé (attribut+année+mois) les diffère selon le mois concerné. Ces champs sont définis dans les paragraphes suivants.

4.1.2 Numéro de carte

Le numéro de carte constitue la clé primaire de la table, un seul enregistrement renferme ainsi toute l'information relative aux transactions effectuées par celle-ci. Dès qu'une carte figure dans un des ensembles de données d'embarquement, un enregistrement s'y référant existe dans la table des profils. Rappelons que les cartes étudiées ici étant non transférables (car identifiées par le nom et une photo de l'utilisateur), la suite du texte de cette section considérera les objets « carte » et « usager » indifféremment.

4.1.3 Apparition mensuelle

Afin de procéder à un examen de la variabilité entre les différents ensembles de données disponibles, un champ binaire (0 ou 1) pour chaque mois est fixé en fonction de l'utilisation (ou non) de la carte au moins une fois au cours de ce mois. Ces indicateurs sont établis afin de pouvoir filtrer les cartes selon le ou les mois étudiés.

4.1.4 Titre de transport

En ne se basant que sur les transactions d'embarquement, les renseignements concernant la socio démographie de la clientèle sont limités. L'information pouvant être tirée de ces transactions peut provenir du libellé du titre de transport utilisé. En l'occurrence, on peut classifier les titres selon le groupe démographique Étudiant, Aîné ou Régulier puis par la zone tarifaire appliquée (selon le tracé) soit la zone Express, Interzone ou Régulière.

Cependant, seule la catégorie du titre sera prise en compte dans les analyses présentées. C'est sa fréquence d'utilisation qui nous permettra de dériver à quel groupe démographique l'utilisateur (rattaché à la carte) appartient. Étant donné la grande variété des titres de transport disponibles, la classification précédemment énoncée (Étudiant, Aîné ou Régulier) est effectuée selon la grille du Tableau 4.2.

Tableau 4.2 – Catégorisation des titres de transport en groupe démographique

Titre	Groupe	Titre	Groupe	Titre	Groupe	Titre	Groupe
1 ADULTE REGULIER	Rég.	4 ET REGULIER 1999	Étud.	23 ETUDIANT ETE REGULIER	Étud.	35 CAMPUS EXPRESS	Étud.
2 ADULTE EXPRESS	Rég.	5 ET EXPRESS 1999	Étud.	24 ETUDIANT ETE EXPRESS	Étud.	36 CAMPUS INTERZONE	Étud.
3 ADULTE INTERZONE	Rég.	6 ET INTERZN. 1999	Étud.	25 ETUDIANT ETE INTERZONE	Étud.	37 ETUDIANT INTEGRE 2004 (7)	Étud.
9 CARTE A VALEUR	Rég.	7 ETUDIANT INTEGRE	Étud.	26 ETUDIANT INTEGRE 2002	Étud.	38 ETUDIANT REGULIER 2004	Étud.
10 EMPLOYE STO	Rég.	11 ETUDIANT INTEGRE	Étud.	27 ETUDIANT REGULIER 2002	Étud.	39 ETUDIANT EXPRESS 2004	Étud.
15 CF REGULIER	Rég.	12 ET REGULIER 2000	Étud.	28 ETUDIANT EXPRESS 2002	Étud.	40 ETUDIANT INTERZONE 2004	Étud.
16 CF EXPRESS	Rég.	13 ET EXPRESS 2000	Étud.	29 ETUDIANT INTERZONE 2002	Étud.	41 INTEGRE SCOLAIRE 2004 (5)	Étud.
17 CF INTERZONE	Rég.	14 ET INTERZONE 2000	Étud.	30 ETUDIANT INTEGRE 2003	Étud.	44 ETUDIANT INTEGRE 2005 (7)	Étud.
42 ARCHIVE	Rég.	19 ETUDIANT INTEGRE 2001	Étud.	31 ETUDIANT REGULIER 2003	Étud.	45 ETUDIANT REGULIER 2005	Étud.
43 CONGRES ACTU	Rég.	20 ETUDIANT REGULIER 2001	Étud.	32 ETUDIANT EXPRESS 2003	Étud.	46 ETUDIANT EXPRESS 2005	Étud.
8 AINE	Aîné	21 ETUDIANT EXPRESS 2001	Étud.	33 ETUDIANT INTERZONE 2003	Étud.	47 ETUDIANT INTERZONE 2005	Étud.
18 CF AINE	Aîné	22 ETUDIANT INTERZONE 2001	Étud.	34 CAMPUS REGULIER	Étud.	48 INTEGRE SCOLAIRE 2005 (5)	Étud.

Le champ « Groupe » de la table correspond ainsi au groupe démographique le plus rencontré dans les transactions d'une carte. À cet égard, le champ « RatioGroupe » équivaut au pourcentage du titre le plus fréquent sur

l'ensemble des transactions. Cette régularité est quasi parfaite ; pour les cartes utilisées au mois de février 2005, 99,8 % des cartes utilisaient toujours le même titre de transport (RatioGroupe de 100 %).

4.1.5 Arrêt-domicile

Dans l'optique où l'on désire observer les comportements de la clientèle sous différentes perspectives, notamment géographique, on désire établir une référence spatiale à chaque carte/usager. Puisque la résidence du détenteur de la carte nous est inconnue mais que nous connaissons les coordonnées géographiques de chaque arrêt, nous approximerons le domicile d'un usager par son arrêt-domicile. Le champ « ArretDomicile » est ainsi défini comme le "premier arrêt de la journée" le plus fréquent. (Figure 4.2)

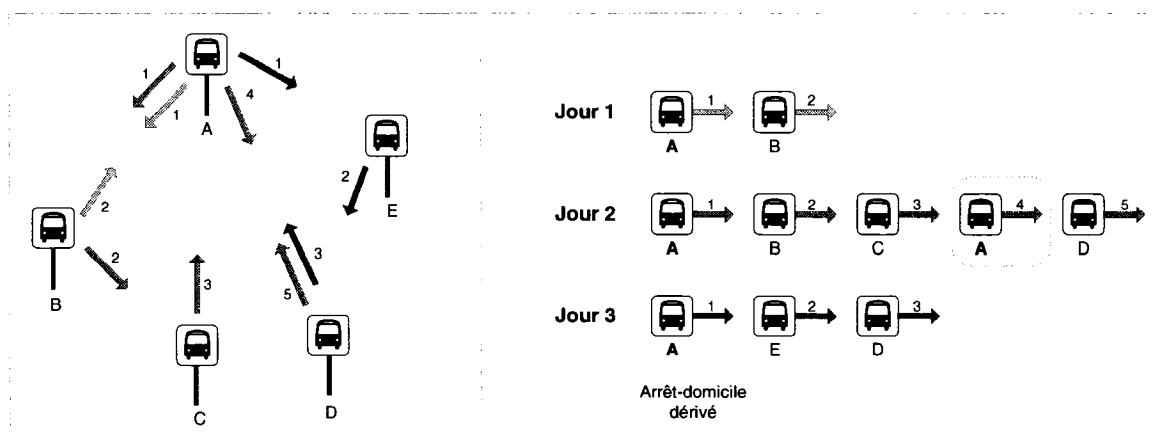


Figure 4.2 – Représentation spatiale et schématique des transactions d'une carte sur trois jours. Pour les trois jours, l'arrêt A est le premier arrêt de la journée

$$AD^c = \text{Mode} \{A_{j1}^c\} \quad \forall j \in \text{Mois}$$

où AD^c = Arrêt-domicile de la carte c

A_{j1} = Arrêt d'embarquement de la première transaction du jour j

De façon similaire à RatioGroupe, le champ « RatioArret » est établi comme étant le pourcentage où l'arrêt-domicile dérivé constitue le premier arrêt de la

journée. Ainsi, pour les profils de février 2005, où on observe la meilleure régularité, 12 % des cartes ont une régularité d'arrêt-domicile parfaite (RatioArret de 100 %), une grande majorité (80 %) des cartes ont une régularité supérieure à 50 %. Cela signifie que pour ces cartes, le premier arrêt de la journée à plus d'une chance sur deux d'avoir été l'arrêt-domicile dérivé.

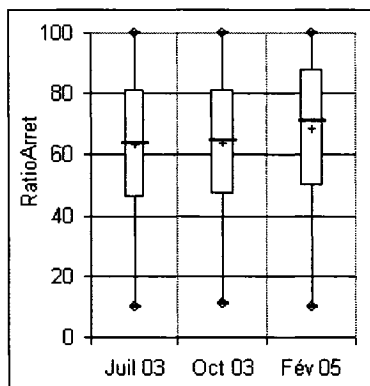


Figure 4.3 – Distribution statistique par boîtes à moustache du champ RatioArret (juillet 2003, octobre 2003 et février 2005)

4.1.6 Facteur d'éloignement

La distance au centre agit souvent comme variable de synthèse de la localisation spatiale des entités (Morency et Chapleau, 2004). Cette variable sera ainsi calculée afin de linéariser la spatialisation des arrêts-domicile et de leurs attributs.

Détermination du centre

Ayant localisé géographiquement les usagers à travers l'usage qu'ils font de leur carte, il reste à identifier un point qui représentera le centre-ville. Étant donné que les lieux de descente des usagers sont inconnus mais que la majorité des déplacements à l'heure de pointe de l'après-midi émanent du côté d'Ottawa, on se fiera à l'intensité de ces embarquements (effectués entre 15h-18h) pour identifier les arrêts recueillant le plus de transactions. Par cette méthode, on valide que 7 des 10 arrêts générant le plus d'embarquements en pointe PM se

retrouvent à Ottawa et que l'arrêt 5042 au coin des rues Wellington et Bank surpasse de loin les autres arrêts (visualisé à la Figure 4.4).

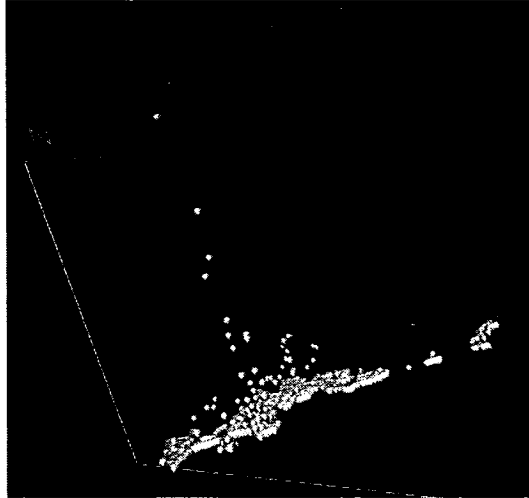


Figure 4.4 – Intensité (axe z) des embarquements aux arrêts en pointe PM (juillet 2003)

Dans ces circonstances, cet arrêt est identifié comme centre et le champ « Eloignement » est défini comme la distance en mètres entre ce centre et l'arrêt-domicile de l'utilisateur.

4.1.7 Indicateurs de mobilité

Les autres champs concernent des indicateurs relatifs à la mobilité des usagers pour chaque mois de données disponible. Pour chaque carte, on détermine essentiellement le nombre de transactions total par mois, le nombre de validations simples ainsi que le nombre de correspondances (Respectivement les champs « NbTrans », « NbValid », « NbCorr »).

On calcule aussi pour chaque carte-mois le nombre de déplacements⁷ par jour, le nombre de déplacements par jour de semaine, le nombre de correspondances par déplacement et le ratio du nombre de correspondances en semaine par le nombre de déplacements en semaine.

Taux de déplacements par jour

Le champ « DeplJour » équivaut au nombre de validations simples divisé par le nombre de jours de ce mois.

Taux de déplacements par jour de semaine

Le champ « DeplJSem » équivaut au nombre de validations simples en semaine divisé par le nombre de jours de semaine de ce mois.

Ratio de correspondances par déplacement

Le champ « RatioCV » équivaut au nombre de transactions de type correspondance divisé par le nombre de validations simples.

Ratio de correspondances en semaine / déplacements en semaine.

Le champ « RatioJSemCV » équivaut au nombre de correspondances en semaine divisé par le nombre de validations simples effectuées en semaine.

4.2 Analyse des résultats

Après la mise en application de la méthode décrite pour chacun des mois étudiés, on dresse le portrait des résultats sous différentes perspectives telles une répartition temporelle des cartes selon les ensembles de données, certains

⁷ Suivant les explications vues au CHAPITRE 1, le terme « déplacement » est équivaut à une « validation simple ».

éléments de démographie ainsi que quelques résultats ayant trait aux attributs de mobilité.

4.2.1 Répartition mensuelle des cartes

Les cartes en service diffèrent selon l'ensemble de données étudié. La Figure 4.5 illustre les appartenances des 38884 cartes distinctes aux ensembles de données selon qu'au moins une transaction y ait été réalisée. On y observe la différence significative entre le nombre de cartes durant le mois d'été (juillet 2003, 11573 cartes) comparé aux mois d'automne et d'hiver (octobre 2003 et février 2005, respectivement 20141 et 21813 cartes).

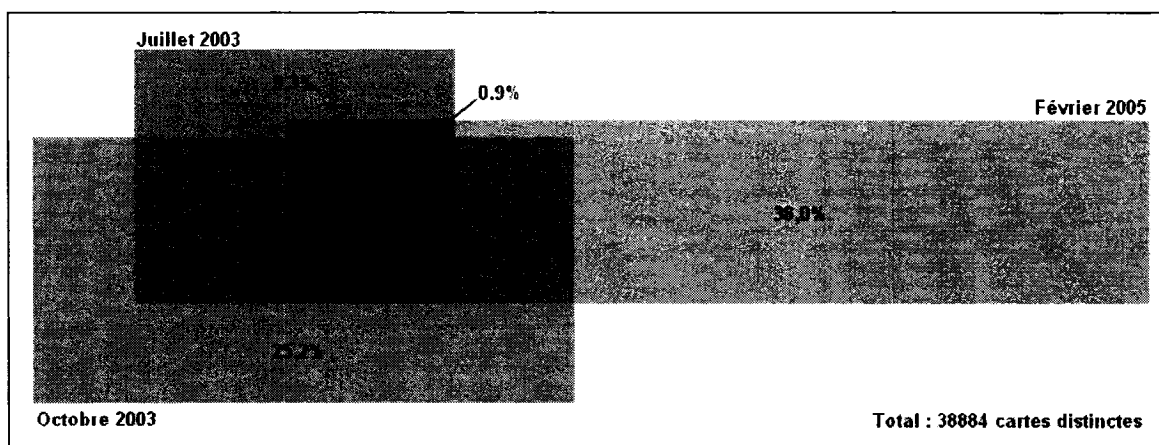


Figure 4.5 – Répartition des cartes selon appartenance aux ensembles de données

Nous reviendrons plus tard à ces aspects alors que des comparaisons d'attributs impliquant plusieurs mois seront présentées.

4.2.2 Démographie des usagers

Étant donné que plus d'une carte peut se voir attribuer un même arrêt-domicile, on peut observer leur répartition sous cette perspective. Le rattachement des cartes à leur arrêt-domicile constitue un élément intéressant dans la mesure où l'on veut connaître la spatialisation des usagers selon leur lieu de résidence.

Taille des arrêts-domicile

À la manière d'une enquête origine destination où l'on évalue la taille des ménages, on peut donc apprécier la « taille » des arrêts-domicile. Près de 6 arrêts du réseau sur 10 constituent des arrêts-domicile. C'est-à-dire qu'au moins une carte leur est rattachée. Une répartition du nombre de cartes par arrêt-domicile pour le mois de février 2005 est présentée à la Figure 4.6. On y apprécie que plus de la moitié (52 %) de ces arrêts-domicile ont une « population » de dix cartes et moins. Les arrêts-domicile ne comportant qu'une seule carte comptent pour 13 %.

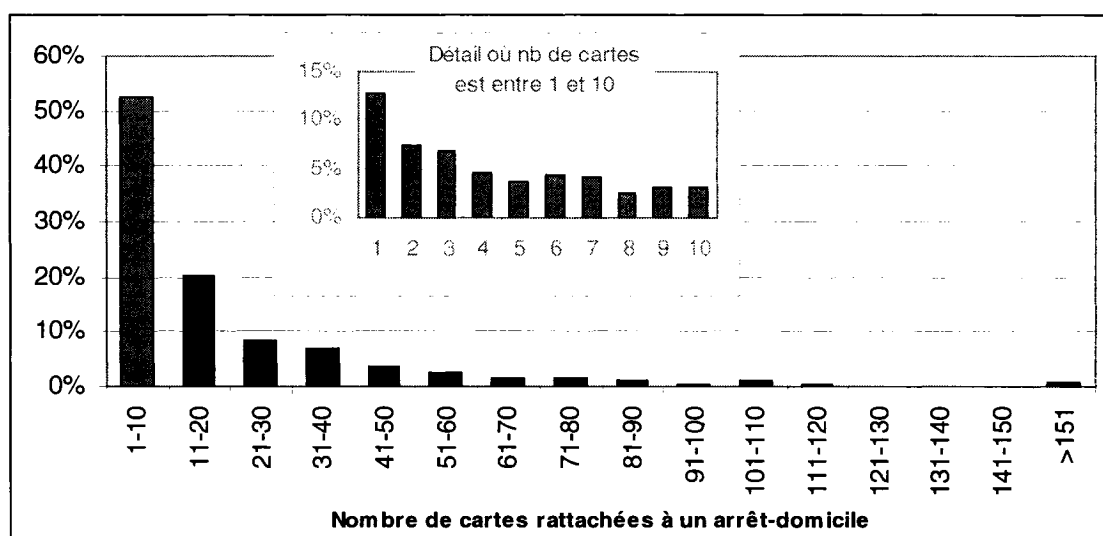


Figure 4.6 – Répartition des arrêts-domicile selon leur nombre de cartes d'utilisateur rattachées (février 2005)

À l'opposé, on retrouve neuf arrêts-domicile auxquels sont rattachées plus de 150 cartes. Ces cas pourraient s'expliquer par une très forte densité résidentielle à proximité ou bien la présence d'un stationnement incitatif.

Répartition de la clientèle

En effectuant une opération afin d'agréger par zones, on peut avoir une bonne idée de répartition spatiale des usagers du système. La Figure 4.7 illustre cette démarche.

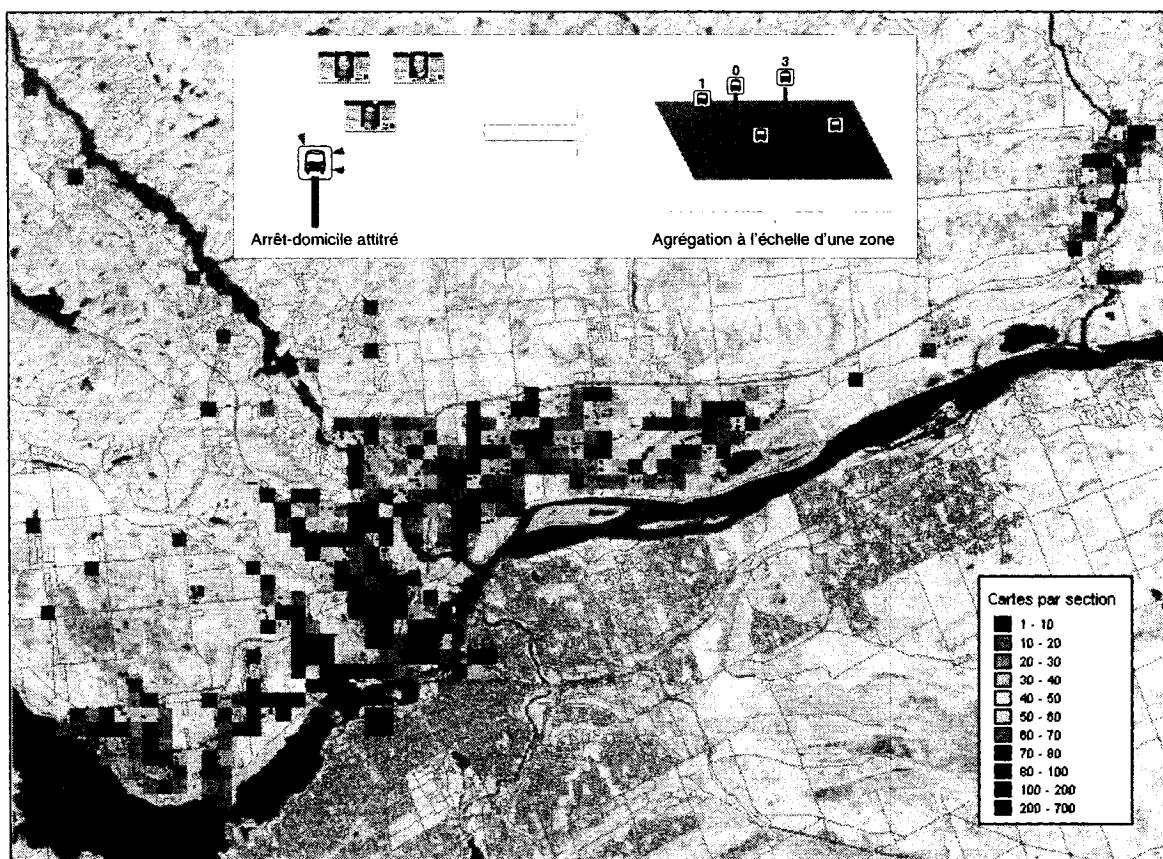


Figure 4.7 – Agrégation spatiale du nombre de cartes d'utilisateur par zone de 0,25 km² (février 2005)

Ayant en notre possession plusieurs ensembles de données, il est possible de mesurer certaines variations de cette démographie dérivée. Le paragraphe suivant s'attarde sur un sous-groupe particulier.

Variation d'un sous-groupe démographique

Les usagers utilisant un titre étudiant sont beaucoup moins nombreux en été qu'en automne. Dans les faits, c'est avec une proportion de plus d'un sur trois

que cette différence s'exprime entre les mois de juillet et d'octobre 2003. La Figure 4.8 illustre où les cartes étudiantes se trouvent à apparaître entre ces deux mois. Deux modes de visualisation sont présents car d'une part, la grille d'analyse permet de mieux visualiser où se retrouvent les faibles écarts négatifs mais la représentation surfacique (à partir des mêmes zones d'analyse) n'est pas limitée par le nombre de classes, ce qui permet une visualisation moins épurée.

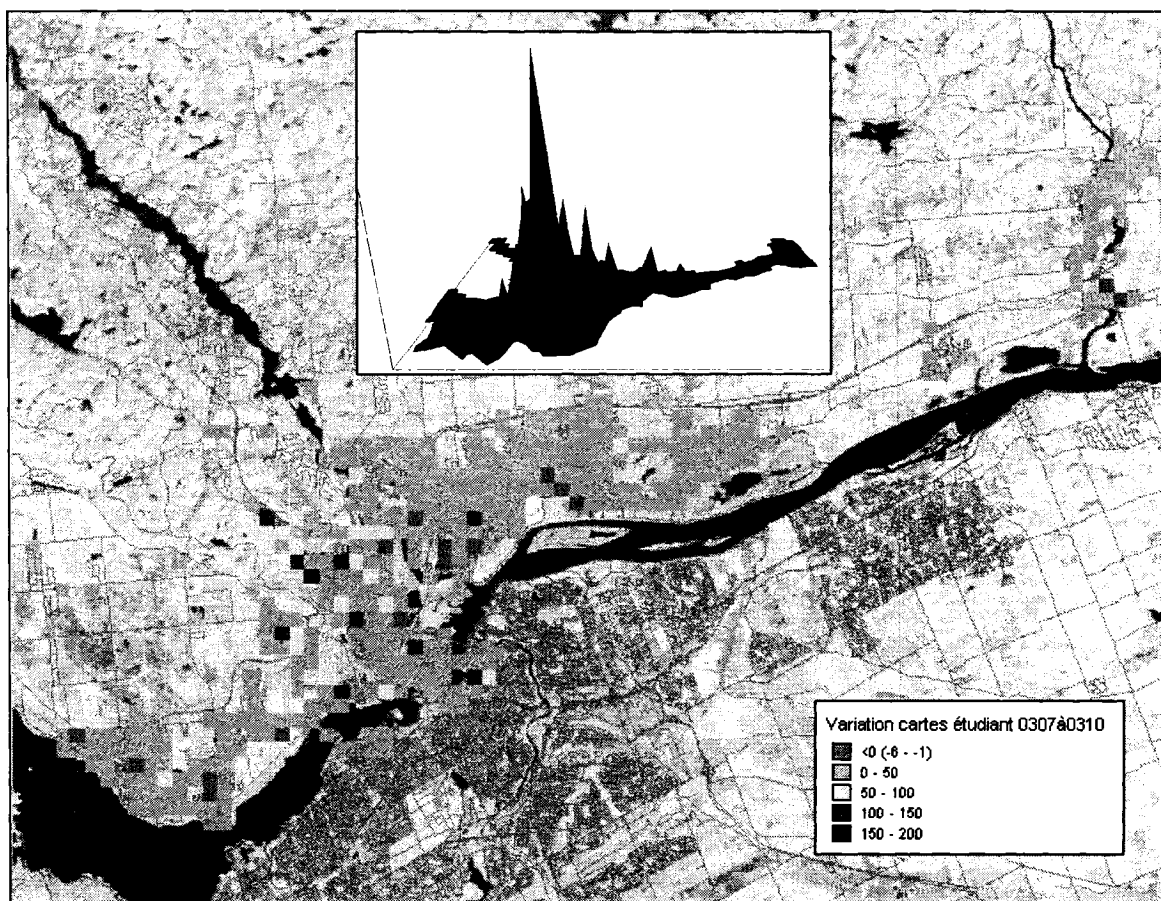


Figure 4.8 – Variation du nombre de cartes de type étudiant entre le mois de juillet 2003 et octobre 2003, par zone de 0,25km² (Représentation surfacique et par grille)

Dans les deux cas, on remarque que les accroissements les plus importants sont concentrés dans la partie centrale (arrondissement Hull) et semblent représenter plusieurs établissements scolaires. La zone où l'augmentation est la plus importante (+189 cartes) est occasionnée par l'arrêt libellé « Collège de

l'Outaouais / Stationnement». Le collège étant un important pôle étudiant et comprend des résidences étudiantes.

4.2.3 Taux de déplacement

Le taux de déplacements/jour est un indicateur régulièrement dérivé des enquêtes ménage. Il est important de rappeler qu'ici, cet indicateur ne s'appuie que sur les déplacements en transport en commun effectués à l'aide d'une carte à puce et que les usagers utilisant le paiement comptant sont exclus du calcul.

La Figure 4.9 exprime les taux globaux⁸ de déplacement par jour (de semaine) à travers les trois ensembles de données disponibles pour chacun des groupes démographiques. On remarque que les taux des titres aînés sont significativement inférieurs aux taux des autres titres, sauf pour le mois d'été où il est rejoint par le taux des titres étudiants. Cela est vraisemblablement attribuable au changement important des rythmes d'activités des étudiants, moins nombreux les mois d'été tel qu'exprimé à la section précédente.

⁸ Par opposition, un taux spécifique se calcule en divisant le nombre de déplacements non pas par le nombre total de jours mais par le nombre d'occurrences clients-jour (Chapleau, 1996).

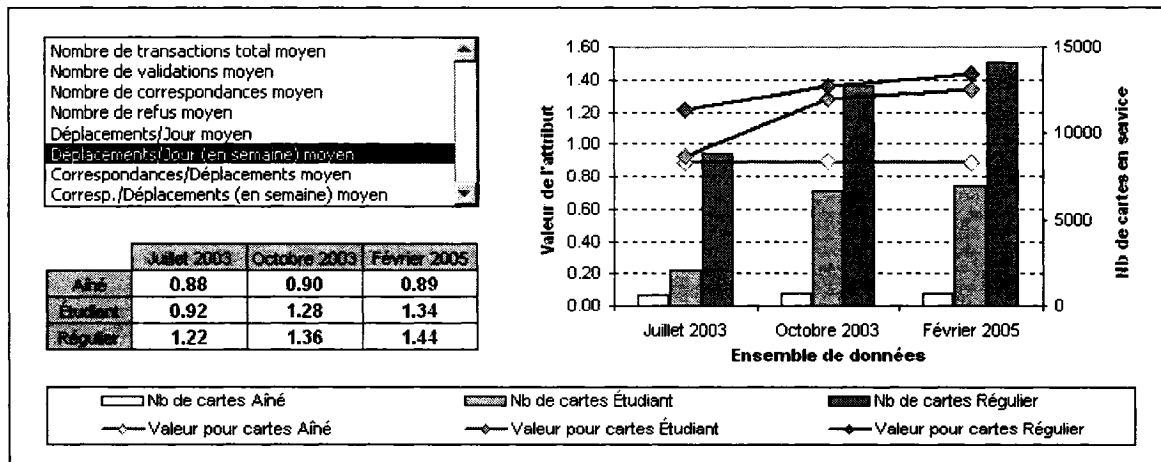


Figure 4.9 - Évolution du taux moyen de déplacements/jour de semaine, selon groupe démographique (juillet 2003, octobre 2003 et février 2005)

Il est intéressant de noter la similitude des variations des taux de déplacement avec le nombre de cartes en activité durant ces mois. Alors que le nombre de cartes « aîné » ne varie presque pas (environ $\pm 10\%$), le nombre de cartes « étudiant » en service l'été diminue du double (environ -70%) comparé à la diminution estivale des titres « réguliers » (environ -36%). En testant la similitude par une régression linéaire entre les paires de valeurs, on obtient un R^2 de $88,3\%$. Il atteint $99,9\%$ en retirant les paires concernant le type de carte étudiant.

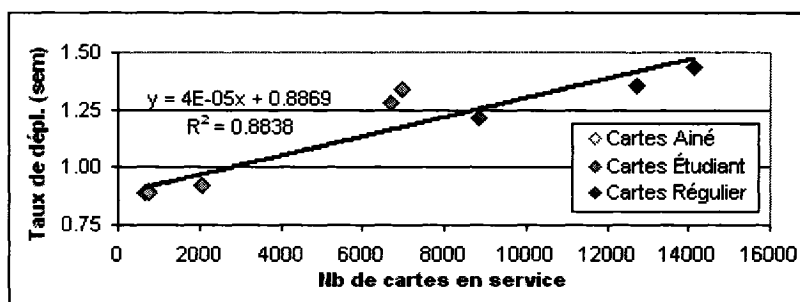


Figure 4.10 – Régression linéaire entre les taux de déplacement et le nombre de cartes en service (selon type de carte et mois)

Selon la perspective de l'éloignement par rapport au centre, le taux de déplacement semble légèrement fléchir avec la distance. La Figure 4.11 illustre

ce constat. (L'éloignement radial n'est pas affiché au-delà de 15 km étant donné certaines inconstances dû au faible nombre de données)

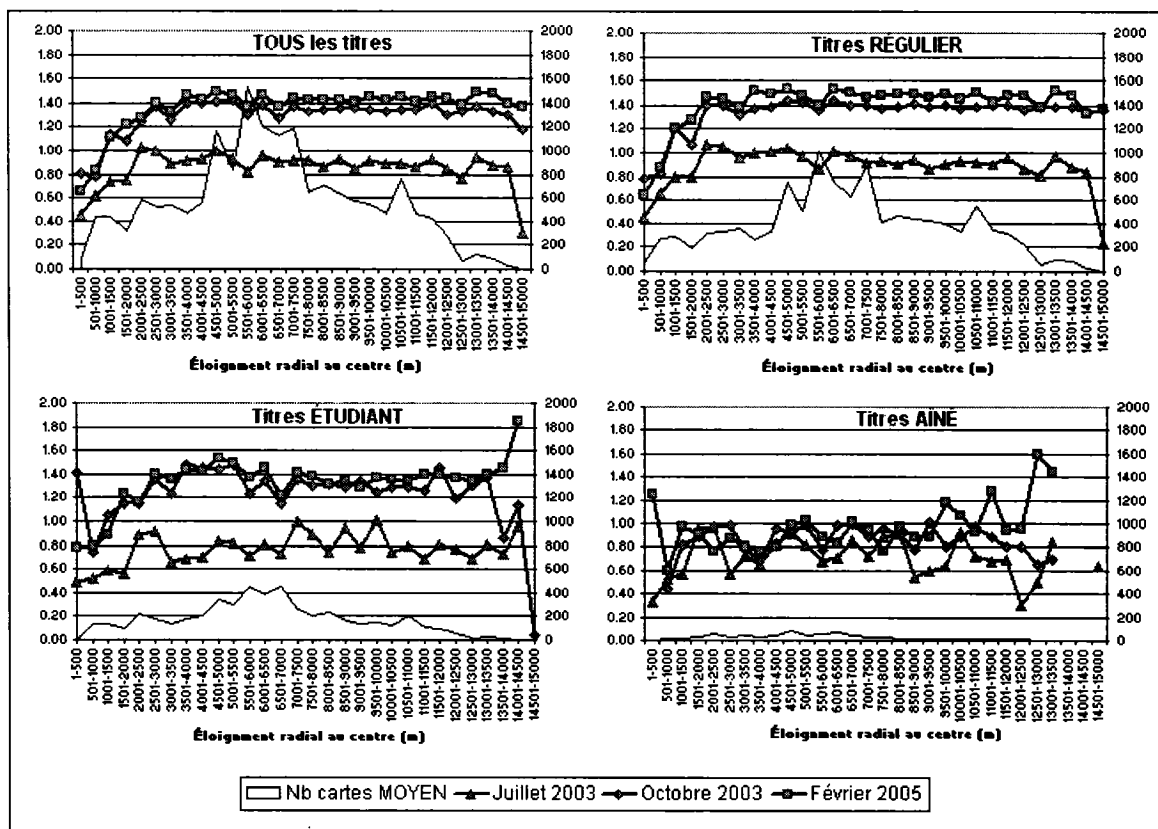


Figure 4.11 – Taux de déplacement et nombre de cartes (moyen) selon éloignement au centre, par ensemble de données et par groupe démographique

Autres échantillons

Bagchi (2003) calcule ses taux de déplacement en utilisant les déplacements non-liés et sans discrimination du type de jour. En appliquant la même méthodologie, on arrive toutefois à des résultats forts différents (Tableau 4.3) mais qui sont néanmoins tous compris à l'intérieur de la fourchette de valeurs obtenues par Bagchi.

Tableau 4.3 – Comparaison entre taux de déplacement (non-liés) de Bagchi (2003) et STO (juillet, octobre 2003 et février 2005)

Catégorie de titre	STO Juillet 2003 (31 jours)	STO Octobre 2003 (31 jours)	STO Février 2005 (28 jours)	Échantillon Bradford (Bagchi)	Échantillon Southport (Bagchi)	Période de l'échantillon de Bagchi
Aîné	1.16	1.20	1.11	0.59	0.40	28 jours
				0.47	0.39	36 jours
Étudiant	1.00	1.28	1.30	n/a	n/a	28 jours
				n/a	n/a	36 jours
Régulier	1.09	1.21	1.23	2.10	n/a	28 jours
				2.11	n/a	36 jours

4.2.4 Taux de correspondances/déplacement

Le ratio de correspondances par déplacement tente d'exprimer le degré de complexité des déplacements réalisés par la clientèle. La Figure 4.12 illustre la variation de cet attribut selon le groupe démographique et l'ensemble de données étudié.

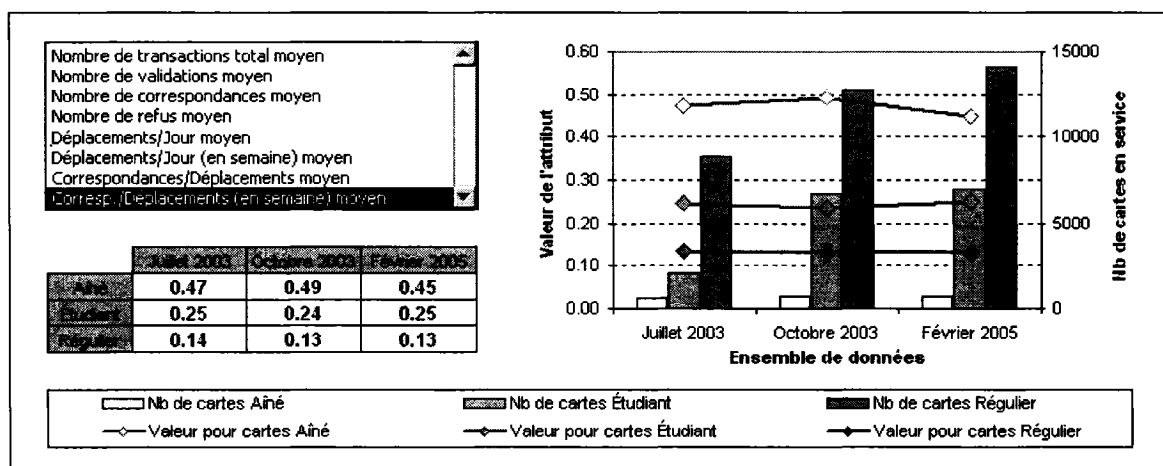


Figure 4.12 – Évolution du taux de correspondances/déplacement en semaine, selon groupe démographique (juillet 2003, octobre 2003 et février 2005)

Alors qu'il ne varie presque pas selon le mois étudié, on peut apprécier la différence marquée entre les types de clientèle. On confirme la plus grande tolérance aux correspondances des clientèles étudiante et aînée.

4.2.5 Taux d'embarquements/déplacement lié

Bagchi (2003) exprime le degré de complexité des déplacements différemment en calculant plutôt le ratio de transactions par déplacement. Son appellation, « linked trip rate », fait appel à la notion de déplacement lié, qu'il définit comme étant deux ou plusieurs embarquements sur des véhicules différents par un individu lors d'un même jour à l'intérieur d'un certain laps de temps.

Bien que Bagchi établisse ce laps de temps comme flexible, dans le cas des deux échantillons concernés, cette durée est de 30 minutes. En ce qui nous concerne, la durée maximale entre deux transactions est la même que celle utilisée par le système de validation de la STO, soit 120 minutes. Malgré cette différence, les résultats découlant de l'application de cette procédure sont présentés au Tableau 4.4.

Tableau 4.4 – Comparaison entre taux de transactions par déplacement (lié) de Bagchi (2003) et STO (juillet, octobre 2003 et février 2005)

Catégorie de titre	STO Juillet 2003 (31 jours)	STO Octobre 2003 (31 jours)	STO Février 2005 (28 jours)	Échantillon Bradford (Bagchi)	Échantillon Southport (Bagchi)
Aîné	1.51	1.49	1.43	1.08	1.04
Étudiant	1.18	1.14	1.14	n/a	n/a
Régulier	1.22	1.19	1.18	1.25	n/a

Bien que très similaires pour la clientèle régulière, l'écart significatif en matière de résultats des titres aînés pourrait s'expliquer par la différence méthodologique évoquée plus haut, la clientèle aînée effectuant davantage de correspondances.

4.2.6 Jours d'activité

On a vu à la section 2.1 que la clientèle utilisant la carte à puce possède un abonnement qui couvre (au moins) un mois. On peut s'intéresser à l'usage que fait la clientèle de ces titres de transport illimités en relation avec le nombre de jours dans lesquels ils ont fait au moins un déplacement. La Figure 4.13 montre

pour chaque mois étudié, la distribution des pourcentages de cartes en activité selon le nombre de jours (de semaine et total) dans le mois.

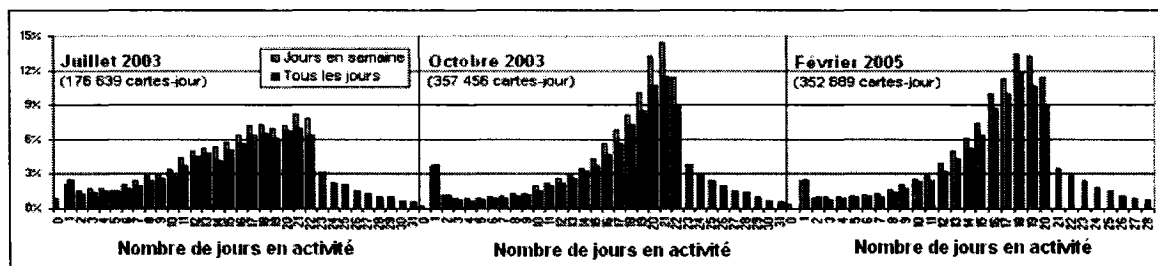
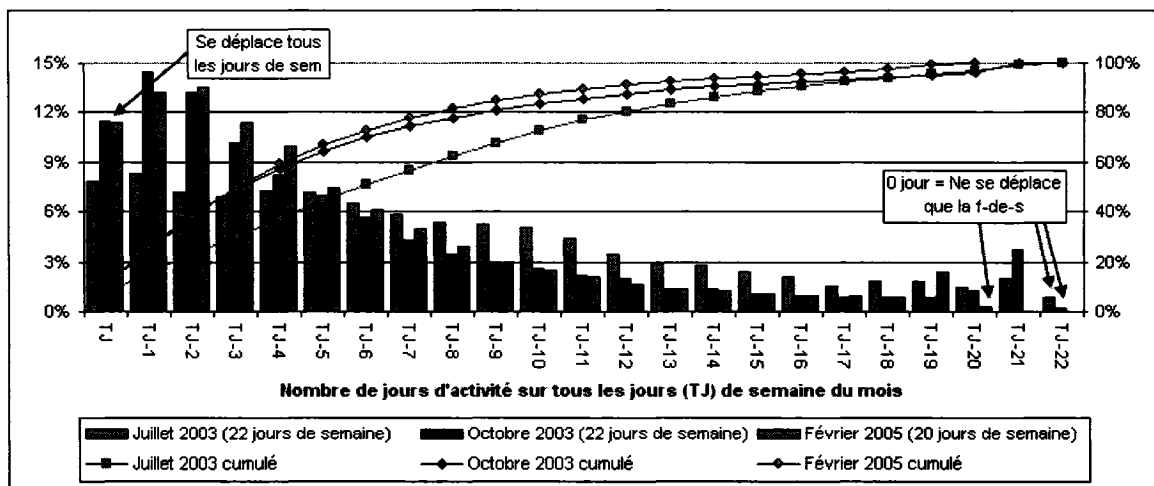


Figure 4.13 – Distribution des pourcentages de cartes en activité selon nombre de jours dans le mois (juillet, octobre 2003 et février 2005)

Activité à tous les jours de semaine

La Figure 4.14 compare les trois mois simultanément selon la perspective d'une activité à tous les jours de semaine. Ainsi pour juillet 2003, octobre 2003 et février 2005, on a respectivement 8,6 %, 11,5 % et 11,6 % des cartes qui ont été utilisées à chaque jour de semaine du mois⁹.



⁹ Les lundi 1^{er} juillet (Fête du Canada) et lundi 13 octobre 2003 (Action de grâce) étant des jours fériés, ils n'ont pas été classés comme jours de semaine.

Figure 4.14 – Distribution du pourcentage de cartes selon nombre de jours d'activité en semaine (juillet, octobre 2003 et février 2005)

La distribution est plus étalée le mois d'été alors que pour octobre et février, on observe une distribution plus concentrée autour d'un maximum de jours dans les mois. Ainsi, pour ces deux mois, près de la moitié des cartes ont eu moins de quatre jours (de semaine) sans utilisation. Le Tableau 4.5 présente une analyse détaillée des jours d'activité pour une semaine.

Tableau 4.5 – Distribution ordonnée et cumulée des combinaisons de jours d'activité pour une semaine (lundi 21 au dimanche 27 février 2005)

Combinaison des jours de la semaine du 21 fév 2005		%	% cumulé
1111100	les cinq jours en semaine	51.8%	51.8%
1111000	quatre jours en semaine (pas le vendredi)	11.0%	62.8%
0111100	quatre jours en semaine (pas le lundi)	7.1%	69.9%
1111111	les sept jours	6.6%	76.5%
1111110	six jours (pas le dimanche)	5.9%	82.4%
1101100	quatre jours en semaine (pas le mercredi)	3.9%	86.3%
1111101	six jours (pas le samedi)	3.8%	90.1%
1110100	quatre jours en semaine (pas le jeudi)	3.5%	93.6%
1011100	quatre jours en semaine (pas le mardi)	3.5%	97.0%
1110000	trois jours en semaine (pas le jeudi et vendredi)	3.0%	100.0%

Unique jour d'activité mensuel

Fait intéressant, pour chacun des trois mois de données, le nombre de cartes qui n'ont qu'un seul jour d'activité échappe à la tendance décroissante et affiche un pourcentage inattendu¹⁰.

Lorsqu'on observe les titres de la clientèle pour lesquels on n'a qu'un jour d'activité dans le mois (Tableau 4.6), on pourrait s'attendre à ce que les usagers qui profitent le moins de leur accès illimité soient principalement des employés

¹⁰ De plus, si notre définition de déplacement avait inclut les transactions refusées, le nombre de cartes n'ayant qu'un jour d'activité aurait doublé pour le mois de juillet 2003. Le reste des résultats n'aurait pas changé significativement. Rappelons que les transactions refusées de juillet 2003 comptent pour 2.4% du total. (0.0006% pour octobre 2003 et aucune pour juillet 2005)

de la STO (qui ont probablement un abonnement gratuit). On constate plutôt que les employés de la société de transport ne comptent seulement pour 27,8 %, 8,1 % et 17,2 % des cartes dans cette situation.

Tableau 4.6 – « Top 10 » des titres les plus utilisés pour les cartes avec un seul jour d'activité mensuelle (juillet 2003, octobre 2003 et février 2005)

Juillet 2003			Octobre 2003			Février 2005		
Titre	Nb	%	Titre	Nb	%	Titre	Nb	%
ADULTE REGULIER	97	35.5%	ADULTE REGULIER	223	29.5%	ADULTE REGULIER	188	34.8%
EMPLOYE STO	76	27.8%	ETUDIANT REGULIER 2002	199	26.3%	EMPLOYE STO	93	17.2%
ETUDIANT REGULIER 2002	35	12.8%	ETUDIANT REGULIER 2003	122	16.1%	ETUDIANT REGULIER 2004	67	12.4%
ADULTE EXPRESS	18	6.6%	EMPLOYE STO	61	8.1%	INTEGRE SCOLAIRE 2004 (5)	33	6.1%
CF REGULIER	14	5.1%	ADULTE EXPRESS	41	5.4%	ADULTE EXPRESS	33	6.1%
AINE	12	4.4%	AINE	21	2.8%	AINE	31	5.7%
ETUDIANT ETE REGULIER	9	3.3%	ETUDIANT INTEGRE 2002	20	2.6%	CF REGULIER	16	3.0%
CF EXPRESS	5	1.8%	ETUDIANT ETE REGULIER	15	2.0%	ETUDIANT REGULIER 2003	14	2.6%
ADULTE INTERZONE	4	1.5%	ETUDIANT INTEGRE 2003	14	1.9%	ADULTE INTERZONE	13	2.4%
ETUDIANT EXPRESS 2002	2	0.7%	CF REGULIER	14	1.9%	ARCHIVE	12	2.2%

Afin d'expliquer cette situation, il est possible que certains usagers aient à changer de carte au cours du mois. Vérification faite, un pourcentage important (entre 25 % et 39 %) des uniques jours d'activité sont effectués le premier jour du mois, tel qu'illustré à la Figure 4.15. En revanche, on ne remarque aucune relation particulière reliée au titre utilisé.

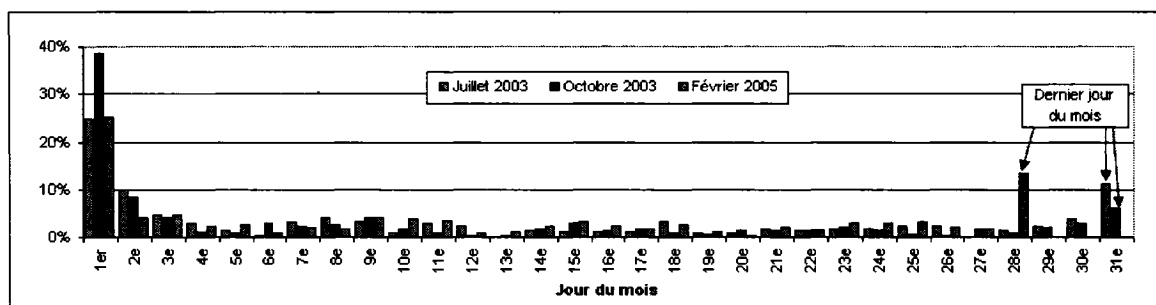


Figure 4.15 – Distribution des pourcentages de cartes selon le jour où a eu lieu l'unique jour d'activité dans le mois (juillet, octobre 2003 et février 2005)

Bien qu'il soit possible que des chauffeurs soient indulgents envers les usagers le premier du mois, on note aussi des valeurs significativement importantes le dernier jour du mois. Le Tableau 4.7 ventile les transactions uniques du mois d'octobre 2003 et de février 2005 en relation avec le type de message d'erreur (juillet 2003 n'a pas ces messages). On remarque que la distribution du type de message diffère beaucoup selon que l'on regarde le premier jour, le dernier jour

ou les autres jours. Ainsi, l'augmentation significative du nombre de transactions uniques le dernier jour du mois est presque exclusivement causée par des transactions avec message « carte invalide ».

Tableau 4.7 – Répartition des messages d'erreur pour les transactions uniques des mois d'octobre 2003 et février 2005

#Err	Description message	Octobre 2003			Février 2005		
		1er jour (mar)	Autres jours	Dernier jour (ven)	1er jour (mar)	Autres jours	Dernier jour (ven)
0	Aucun message	53%	23%	34%	86%	59%	24%
20	Déjà valide						
21	Carte refusée LI						
22	Carte refusée LN		1%			2%	
23	Carte invalide		2%	38%		5%	61%
26	Supplément exigé				2%		
26	Carte expirée	4%	46%	22%		23%	13%
27	Titre invalide	43%	28%	6%	2%	9%	3%
28	Titre invalide						
29	Ligne invalide						
30	Réponse invalide						
ÉCHANTILLON		189	340	32	81	244	38

Notons toutefois que cet échantillon est relativement faible et que le titre « employé STO » n'obtient jamais de message d'erreur.

4.3 Variabilité sur échantillon constant

Cette section s'intéresse à la clientèle active dans les trois ensembles de données. À la base, le questionnement principal repose sur la variabilité de l'arrêt-domicile mais sera élargi au groupe démographique ainsi qu'à certains attributs de mobilité. Nous avons vu en début de section 4.2 que 10,2 % des cartes inventoriées ont été utilisées à l'intérieur des trois mois de données, ce qui correspond à près de 4000 cartes.

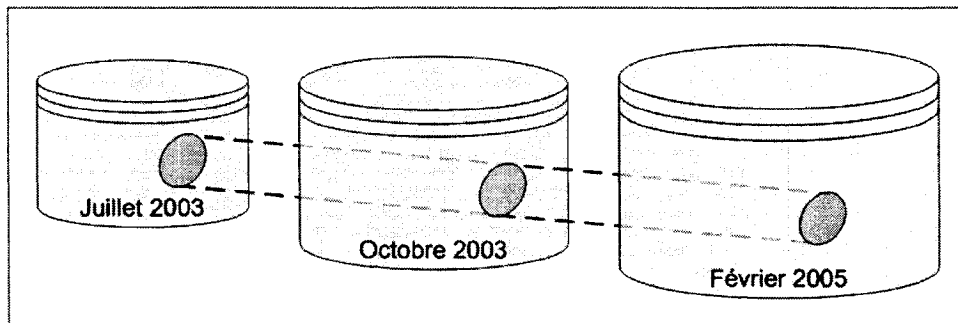


Figure 4.16 – Échantillon comprenant les mêmes éléments dans les trois ensembles de données

4.3.1 Représentativité de l'échantillon

Afin de connaître les biais éventuels de ce genre d'analyse, il importe de se questionner sur la représentativité des sous-groupes impliqués dans l'échantillon analysé. En l'absence de données socio démographiques sur la clientèle de la STO, nous devons nous rabattre sur l'unique caractère démographique à notre disposition, soit le groupe démographique le plus fréquent. La Figure 4.17 exprime la sous-représentation de la clientèle étudiante qui ne constitue ici que 1 % de l'échantillon alors qu'elle varie entre 18 % et 33 % pour les autres ensembles de données.

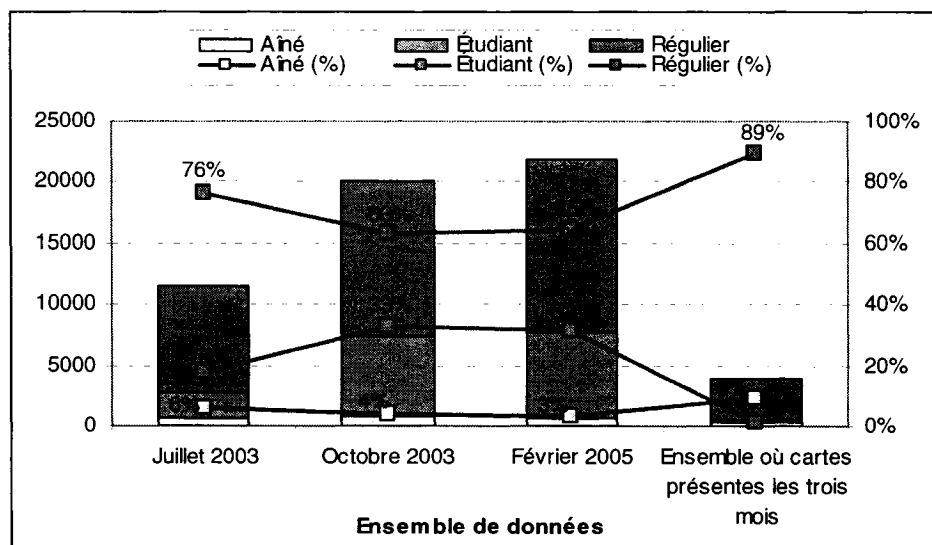


Figure 4.17 – Répartition des groupes démographiques des cartes selon l'ensemble étudié (juillet, octobre 2003 et février 2005)

Cette circonstance semble pouvoir s'expliquer par la spécificité de l'abonnement étudiant. En effet, les étudiants de moins de 21 ans doivent renouveler leur carte à puce annuellement avant chaque rentrée automnale afin de profiter du tarif réduit (Société de transport de l'Outaouais, 2006). On imagine qu'à ce renouvellement, la photo sur la carte est remplacée de même que la carte physique, ayant pour conséquence l'émission d'un nouveau numéro de carte.

Tel que mentionné à la section 2.1, il existe aussi un forfait renouvelable à chaque semestre (forfait Cam-Puce) donnant aux adultes inscrits aux études à temps complet accès au tarif réduit. Vraisemblablement, le 1 % de cartes à tarif étudiant de l'échantillon représenterait ainsi 48 adultes ayant profité de ce forfait d'une façon ininterrompue (pas de renouvellement de carte) sur les 20 mois compris entre juillet 2003 et février 2005.

4.3.2 Variabilité du titre et de l'arrêt-domicile

Le groupe démographique (régulier, étudiant ou aîné) demeure le même pour 98,5 % des cas mais pour ce qui est de l'arrêt-domicile, ce taux tombe à 52 %

des cartes qui ne voient pas leur arrêt-domicile demeurer constant sur les trois périodes.

Qu'est-ce qui explique un pourcentage aussi important (48 %) des cartes qui ont des arrêts-domicile dissemblables? Lorsqu'on observe la distance maximum entre les arrêts-domicile différents des trois ensembles (Figure 4.18), on se rassure un peu en constatant qu'une importante partie (36 %) est en deçà de 100 m, 36 % se situe entre 100 m et 1000 m, puis que seulement 28 % des distances maximales dépassent le kilomètre. C'est-à-dire que pour les clients qui modifient leur arrêt matinal, une majorité opte pour un nouvel arrêt relativement proche de l'ancien.

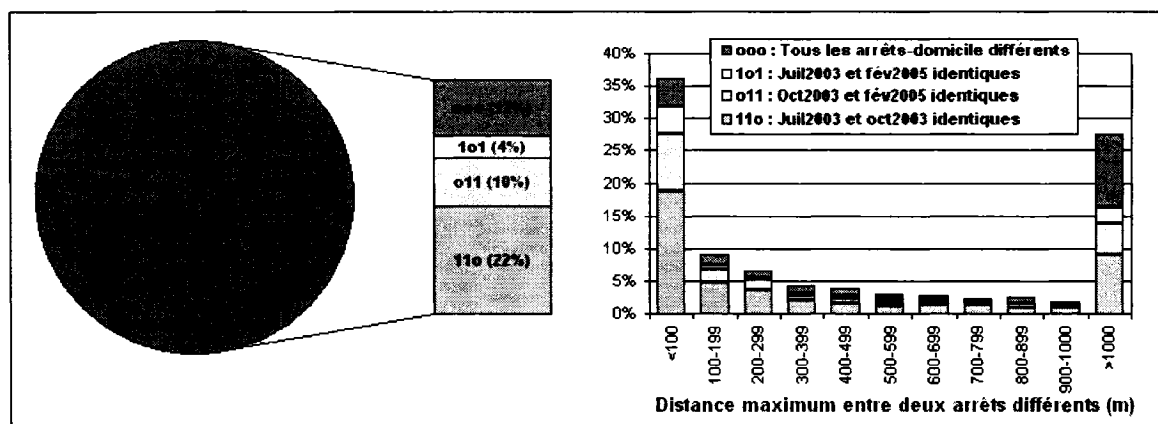


Figure 4.18 – Constance de l'arrêt-domicile et distribution selon distance maximum pour les arrêts-domicile non identiques (3981 cartes apparaissant dans mois de juillet 2003, octobre 2003 et février 2005)

Bien qu'un décalage temporel relativement important sépare les premiers ensembles de données (juillet et octobre 2003) avec le dernier (février 2005), il est possible qu'une partie de ces cas ne s'expliquent par un changement des profils d'activité, d'un changement de mode (i.e. vélo et marche plus fréquents en été) ou d'une modification du réseau (arrêt déplacé/ retiré).

4.3.3 Taux de déplacement

Parmi les variations intéressantes induites des caractéristiques de mobilité préalablement calculées, on note la variation du nombre de déplacements/jour (en semaine) où on constate l'utilisation inférieure lors du mois d'été. (Tableau 4.8)

Tableau 4.8 – Taux de déplacement de l'échantillon constant selon mois et groupe démographique¹¹

Titres	Juillet 2003	Octobre 2003	Février 2005
Aîné	0.96	0.99	0.89
Étudiant	1.44	1.51	1.31
Régulier	1.28	1.49	1.52
Total	1.26	1.44	1.46

Par contre, ce taux ne varie pas beaucoup selon la distance au centre (Figure 4.19). Notons qu'au-delà de 15 km, c'est le faible échantillon qui induit les variations importantes des taux.

¹¹ Étant donné que 98.5% des cartes de l'échantillon ont une catégorie de titre invariable, ce tableau utilise une ventilation par catégorie de titre d'un mois en particulier. (février 2005)

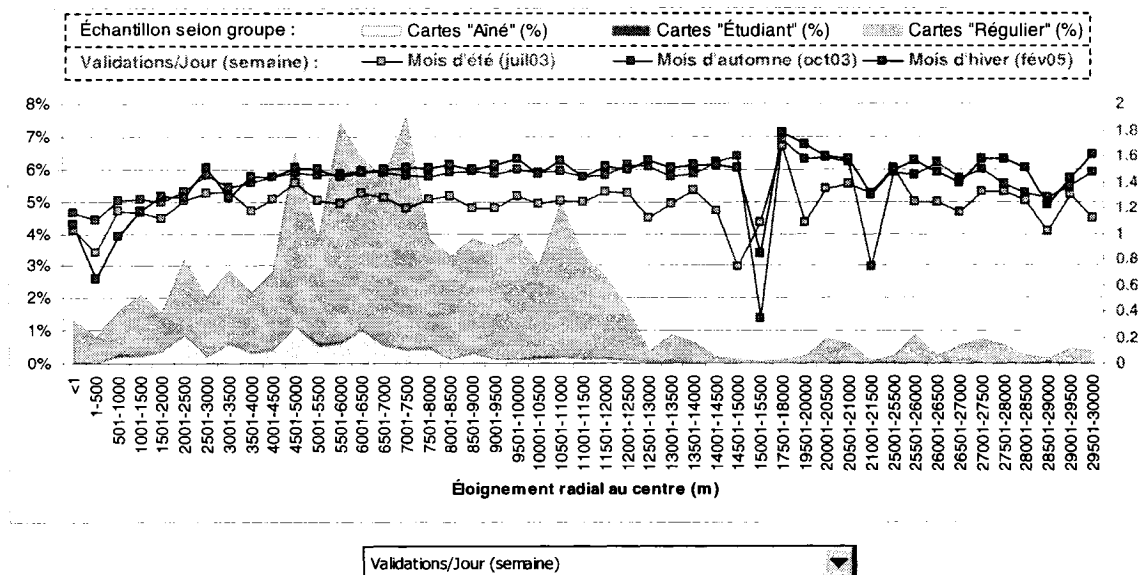


Figure 4.19 – Variation du ratio de validations simples par jour de semaine, distribué selon l'éloignement radial au centre. Arrière-plan : distribution radiale du nombre de cartes

4.3.4 Taux de correspondances/déplacement

On peut aussi apprécier les taux de correspondances par déplacement (en semaine) rapportés au Tableau 4.9.

Tableau 4.9 – Taux de correspondances/déplacement de l'échantillon constant selon mois et groupe démographique¹²

Titres	Juillet 2003	Octobre 2003	Février 2005
Aîné	0.50	0.53	0.47
Étudiant	0.22	0.23	0.19
Régulier	0.11	0.11	0.11
Total	0.15	0.15	0.15

Cette fois, les taux ne varient pas selon la période étudiée mais ont plutôt tendance à diminuer avec la distance au centre. (Figure 4.20)

¹² Idem

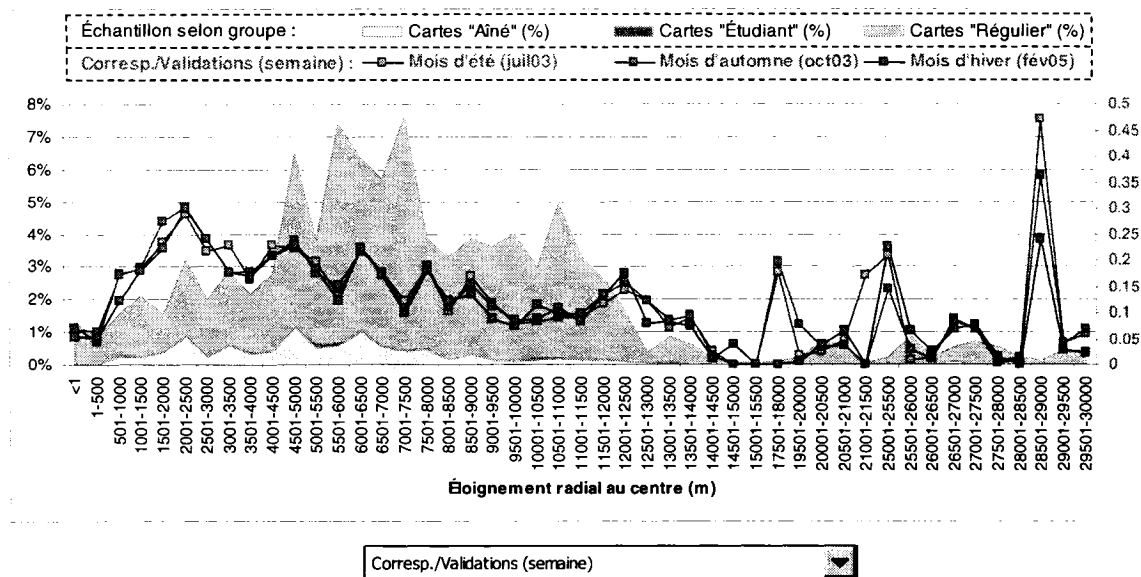


Figure 4.20 – Variation du taux de correspondance/déplacement en semaine distribué selon l'éloignement radial au centre. Arrière-plan : distribution du nombre de cartes

Ici aussi, le faible échantillon explique la grande variabilité au-delà de 15 km du centre.

4.3.5 Analyse transversale vs longitudinale

Dans le paragraphe où on compare les taux de déplacement entre les différents mois, on observait une différence importante entre les mois d'été et d'automne mais cette différence est-elle significative? Avec un échantillon où on suit le même groupe d'individus, la réponse à cette question dépend de la méthode utilisée.

Yee et Niemeier (1996) ont comparé la méthode classique d'enquêtes transversales répétées dans le temps (Repeated Cross-Section Surveys) où on analyse une variation globale, à la méthode longitudinale où les variations sont mesurées à l'échelle individuelle. Dans les deux cas, on estime si une variation est significative à l'aide de l'erreur standard et d'un test Student (test T).

Alors que la méthode transversale d'analyse de variation par comparaison agrégée étudie la variation de la moyenne, l'analyse longitudinale observe plutôt

la moyenne des variations individuelles. Dans le premier cas, l'observation d'une population à un instant précis peut faire en sorte que des écarts opposés se neutralisent lorsqu'on évalue la moyenne. En revanche, l'analyse longitudinale prend en considération toutes les variations particulières alors que l'on suit les mêmes éléments d'une population.

Lorsque ces méthodes sont appliquées aux taux de déplacement de notre échantillon constant (Figure 4.21), on observe une non-concordance des résultats pour la variation entre octobre 2003 et février 2005. C'est-à-dire que la faible augmentation entre ces deux mois n'est pas considérée significative en agrégeant les variations par mois mais devient significative lorsque ce sont les variations individuelles qui sont comptabilisées.

De0307A0310	De0310A0502	De0310A0502	
1.26	1.44	1.46	Moy. du mois
0.00864	0.00732	0.00818	Err. stand.

De0307A0310	De0310A0502	
0.19	0.02	Variation de la moyenne
0.01132	0.01097	Err. stand. de la variation
16.368	1.413	Valeur du test T
1.645	1.645	Test T à 95% (K=taille)
significatif	non signif.	

De0307A0310	De0310A0502	
0.19	0.02	Moyenne des variations
0.00867	0.00746	Err. stand. de la moyenne
21.365	2.078	Valeur du test T
1.645	1.645	Test T à 95% (K=taille)
significatif	significatif	

Analyse de variation par comparaison agrégée

Analyse de variation par comparaison désagrégée

Figure 4.21 – Méthode des sondages répétés vs analyse longitudinale. Application aux taux de déplacement des 3981 cartes de l'échantillon. Distribution T à 95 % (degrés de liberté ∞)

La conclusion obtenue (la variation entre octobre 2003 et février 2005 devient significative) est identique lorsqu'on utilise les données sur les taux de correspondances par déplacement en semaine.

CONCLUSION

Bien que l'implantation de systèmes carte à puce soit généralement motivée par une réduction de la fraude et/ou la modernisation des systèmes de perception, l'utilisation des données opérationnelles a posteriori constitue une manne informationnelle considérable. Avant de tenter d'en tirer profit, ce mémoire a voulu valider certains éléments de cohérence de l'information en appliquant aux données, des logiques spatio-temporelles représentant des contraintes physiques et opérationnelles auxquelles sont soumis les objets de transport impliqués.

On passe ainsi au tamis une information brute où plusieurs anomalies sont ressorties. Parmi celles-ci, on retrouve :

- des enregistrements manquants;
- des tracés et/ou arrêts invalides;
- des erreurs d'assignation;
- des erreurs dans l'arrêt dérivé;
- des désynchronismes de l'horloge interne de l'ordinateur de bord;
- des situations suspectes où la vitesse dérivée d'un véhicule est anormalement élevée;
- des situations suspectes où un chauffeur est simultanément dans deux véhicules;
- des situations suspectes où deux véhicules effectuent le même départ.

Au-delà de certaines données suspectes ou anomalies évidentes, l'application de ces mécanismes de validation a permis de mettre en lumière certaines conjonctures qui auraient pu passer inaperçues. En outre, on constate que les

erreurs d'assignation constituent la principale problématique relative à la qualité des données étudiées. Si bien que pour les mois de juillet 2003, octobre 2003 et février 2005, ces erreurs affectent respectivement approximativement 15 %, 14 % et 8 % des transactions.

La deuxième partie du mémoire s'attarde sur la description et l'application de méthodologies où des manipulations et traitements tentent de bonifier nos processus de connaissance via, notamment, une caractérisation de l'offre de service et de la variabilité de la demande.

Ainsi, c'est dans une perspective de modélisation réseau que l'analyse des correspondances réalisées par la clientèle vient aider à la codification et à la simplification du réseau. L'agrégation en nœuds des arrêts de proximité et des arrêts où une correspondance a déjà été effectuée (et où elle sera dorénavant permise) contribue à simplifier un réseau avec lequel on pourra éventuellement calibrer un modèle d'affectation, dans le but d'optimiser ce réseau. En appliquant des logiques spatio-temporelles aux segments de transfert entre les arrêts de correspondance, cette analyse aura aussi permis de dégager certaines anomalies, majoritairement engendrées par des enregistrements pourvus d'assignations erronées.

Ensuite, l'application d'une méthodologie visant la déduction du service planifié (d'un jour moyen de semaine), à l'aide de la régularité des tournées, contribue à bâtir un outil de référence pouvant notamment être utilisé afin de corriger ces assignations erronées. Le fait que les attributs déduits comportent un champ associé agissant comme mesure de leur qualité, peut contribuer à nuancer les imputations potentielles.

Puis, l'application d'une méthodologie visant l'analyse de la demande à travers les profils de la clientèle a permis de caractériser chaque carte/usager selon des attributs ontologiques (groupe démographique du titre, arrêt-domicile, distance au centre, etc.) et de mobilité (taux de déplacement, jours d'activité, etc.). S'en

suit des analyses selon diverses perspectives où l'on remarque que certains indicateurs dérivés s'avèrent comparables à des résultats similaires, provenant de la littérature. Finalement, l'analyse de type longitudinale sur un échantillon constant a notamment montré une bonne cohérence de l'arrêt-domicile dérivé, même quand celui-ci n'est pas identique sur les trois mois.

Pistes de recherche future

"The travel behaviour researcher views GPS data like a child facing a candy store" (Wolf et al, 2004). La même chose pourrait être affirmée à propos des données provenant de systèmes carte à puce tant leur potentiel d'exploration est riche et varié. En effet, un des défis relevés dans cette recherche fut d'en limiter l'exploration à certains aspects. D'autres travaux en cours devraient poursuivre cette exploration de manière plus approfondie.

Ainsi, les pistes de recherches évoquées ici mettent plutôt l'accent sur le potentiel de liaison avec d'autres sources de données pouvant bonifier l'utilisation de l'information provenant de systèmes carte à puce. Par exemple, des données de comptes à bord (manuels ou automatiques), peuvent notamment contribuer à calibrer un modèle de dérivation des points de descente des usagers carte à puce.

De manière similaire, une liaison avec des données provenant d'enquêtes origine destination pourrait permettre d'accroître l'information déduite des données carte à puce en intégrant notamment les éléments de déplacements TC, les lieux de résidence, les lieux d'emploi, les générateurs de déplacement, etc. Déjà accompli avec des données GPS (Wolf et al, 2001), l'information provenant de cartes de zonage peut conduire à imputer des motifs aux déplacements. Même des données de nature météorologique peuvent être pertinentes afin de confirmer leur impact sur le service rendu (exemple avec GPS dans Hofmann et O'Mahony, 2005b) ou sur le plan du comportement des usagers.

Enfin, les données opérationnelles carte à puce, tout comme ces autres sources de données, font partie d'un éventail de contributions, qui par leur combinaison et leur croisement, peuvent contribuer à l'élargissement de la gamme d'outils disponibles ayant comme finalité l'aide à la planification.

RÉFÉRENCES

- BAGCHI, M; WHITE, P.R., 2005, The potential of public transport smart card data, Transport Policy 12, Vol 5, p464-474.
- BAGCHI, M., 2003, Use of smart card data from bus systems for travel behaviour analysis, and implications for marketing, Thèse de doctorat, University of Westminster (non publiée), 269p.
- BAGCHI, Mousum; WHITE, Peter R., 2004, What role for smart-card data from bus systems?, Proceedings of the Institution of Civil Engineers: Municipal Engineer, v 157, n 1, p39-46.
- BARRY, James J.; NEWHOUSER, Robert; RAHBEE, Adam; SAYEDA, Shermeen , 2002, Origin and destination estimation in New York City with automated fare system data, Transportation Research Record, n 1817, p183-187.
- BERGERON, Daniel; GAGNÉ, Sébastien, 2005, Survol des potentialités informationnelles et analytiques d'un système GPS de localisation en temps réel, 40e Congrès annuel de l'association québécoise du transport et des routes, Québec, 7p.
- BOYLE, Daniel K, 1998, Passenger counting technologies and procedures, TCRP synthesis #29, National Academy Press, 50p.
- BRYDIA, Robert E.; TURNER, Shawn M.; EISELE, William L.; LIU, Jyh C., 1998, Development of intelligent transportation system data management, Transportation Research Record, n 1625, p124-130.

- BULLOCK, Philip; JIANG, Qingjian; STOPHER, Peter R., 2005, Using GPS Technology to Measure On-Time Running of Scheduled Bus Services, Journal of Public Transportation, Vol. 8, No. 1, p21-40.
- CHAPLEAU, Robert, 1999, Analyse totalement désagrégée de la mobilité des travailleurs montréalais, Routes et transports, Montréal, vol. 28, no. 2, p9-20.
- CHAPLEAU, Robert, 2003, Measuring the internal quality of a CATI travel household survey, in Stopher, Peter, Jones, Peter (2003). Transport Survey Quality and Innovation, Kruger, South Africa, Pergamon, p69-87.
- CHAPLEAU, Robert; ALLARD, Bruno; TRÉPANIÉ, Martin; MORENCY, Catherine, 2000, Les sites Web transport dans la Grande région de Montréal : problématique et orientations informationnelles, 35e congrès de l'Association québécoise du transport et des routes, Québec, 26p.
- CHAPLEAU, Robert; MORENCY, Catherine, 2002, Pour une confusion interrompue à propos des déplacements urbains, 37e congrès de l'Association québécoise du transport et des routes, Québec, 18p.
- CHAPLEAU, Robert; MORENCY, Catherine, 2001, Quelques enjeux-questions sur l'information qui assiste la prise de décision en transport urbain, 36e congrès annuel de l'Association québécoise du transport et des routes, Laval, 20p.
- CHUNG, Eui-Hwan ; SHALABY, Amer, 2005, A Trip Reconstruction Tool for GPS-based Personal Travel Surveys, Transportation Planning and Technology, v. 28 , n. 5, p381.
- CHURCH, Richard L; NORONHA, Val; LEI, Ting; CORRIGAN, Wils; BURBIDGE, Shaunna; MARSTON, Jim, 2005, Spatial and Temporal Utility Modeling to

Increase Transit Ridership - Final Report, Vehicule Intelligence & Transportation Analysis Laboratory, University of California, 50p.

FURTH, Peter G., 2000, Data analysis bus planning and monitoring, TCRP Synthesis 34, National Academy Press, 62p.

FURTH, Peter G.; MULLER, Theo H.J.; STRATHMAN, James G.; HEMILY, Brendon , 2004, Designing automated vehicle location systems for archived data analysis, Transportation Research Record, n 1887, p62-70.

GREAVES, Stephen, 2004, GIS and the Collection of Travel Survey Sata, P. Stopher, K. Button, K. Haynes and D. Hensher (eds.) Handbook of Transport 5: Transport Geography and Spatial Systems, Pergamon/Elsevier Science, Chap.21.

HAMMERLE, Meghan; HAYNES, Michael; McNEIL, Sue, 2005, Use of automatic vehicle location and passenger count data to evaluate bus operations - Experience of the Chicago transit authority, Illinois, Transportation Research Record, n 1903, p27-34.

HARVEY, Andrew S., 2004, TimeUse and Activity Systems, P. Stopher, K. Button, K. Haynes and D. Hensher (eds.) Handbook of Transport 5: Transport Geography and Spatial Systems, Pergamon/Elsevier Science, Chap.35.

HOFMANN, M.; O'MAHONY, M., 2005a, Transfer journey identification and analyses from electronic fare collection data, Proceedings of Intelligent Transportation Systems 2005, p34-39.

HOFMANN, M.; O'MAHONY, M., 2005b, The impact of adverse weather conditions on urban bus performance measures, Proceedings of Intelligent Transportation Systems 2005, p84-89.

- KITAMURA, Ryuichi, 2004, Spatial Processes, P. Stopher, K. Button, K. Haynes and D. Hensher (eds.) Handbook of Transport 5: Transport Geography and Spatial Systems, Pergamon/Elsevier Science, Chap.29.
- KWAN, Mei-Po, 2000, Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set, Transportation Research Part C: Emerging Technologies, v 8, n 1, p185-203.
- MARTIN, H.A.; STEANE, A.H.; MAUCERI, V.C., 1992, The use of automatic passenger counting data to verify schedule adherence Computer-Aided Transit Scheduling, Proceedings of the Fifth International Workshop on Computer-Aided Scheduling of Public Transport, p245-58.
- MILLER, Harvey J., 2005, What about people in geographic information science?, Fisher and D. Unwin (eds.), Re-Presenting Geographic Information Systems, John Wiley, p215-242.
- MILLER, Harvey J., 2004, Activities in Space and Time, P. Stopher, K. Button, K. Haynes and D. Hensher (eds.) Handbook of Transport 5: Transport Geography and Spatial Systems, Pergamon/Elsevier Science, Chap.36.
- MILLER, Harvey J.; SHAW, Shih-Lung, 2001, Geographic Information Systems for Transportation: Principles and Applications, Oxford University Press, 458p.
- MORENCY, Catherine; CHAPLEAU, Robert, 2004, De petits points lourds de connaissance: essais méthodologiques sur les données, 39e Congrès annuel de l'association québécoise du transport et des routes, Québec, 24p.
- MORENCY, Catherine, 2004, Contributions à la modélisation totalement désagrégée des interactions entre mobilité urbaine et dynamiques spatiales,

Thèse de doctorat en génie civil, géologique et des mines, Ecole polytechnique de Montréal, 487p.

MORENCY, Catherine; TRÉPANIÉ, Martin; AGARD, Bruno, 2006, Contributions des données de cartes à puces à la compréhension des rythmes d'activités en transport, 41e congrès de l'Association québécoise du transport et des routes, Laval, 19p.

MURAKAMI, E; WAGNER, D.P., 1999, Can using global positioning system (GPS) improve trip reporting?, Transportation Research Part C: Emerging Technologies, v 7, n 2-3, p149-165.

NI, Daiheng; LEONARD II, John D.; GUIN, Angshuman; FENG, Chunxia , 2005, Multiple imputation scheme for overcoming the missing values and variability issues in ITS data, Journal of Transportation Engineering, v 131, n 12, p931-938.

OKAMURA, T.; FUJIWARA, A.; ZHANG, J., 2004, Finding behavioral rules of urban public transport passengers by using boarding records of integrated stored fare card system, 10th World Conference on Transport Research, Istanbul, Turkey, July 4-8, 11p.

O'SULLIVAN, D ; MORRISON, A ; SHEARER, J, 2000, Using desktop GIS for the investigation of accessibility by public transport: An isochrone approach, International Journal of Geographical Information Science, v. 14 , n. 1, p85.

PENDYALA, Ram M., 2005, Collection and analysis of GPS-based travel data for understanding and modeling activity-travel patterns in time and space, Présentation et article présentés à la conférence: GPS Tracking and Time-Geography Applications for Activity Modeling and Microsimulation, Santa Barbara, CA, 4p.

- QUIROGA, Cesar A., 1997, An Integrated GPS-GIS Methodology for Performing Travel Time Studies, Louisiana State University Dissertation, 171p.
- QUIROGA, Cesar A., 1999, Accuracy of linearly referenced data by using geographic information systems, Transportation Research Record, n 1660, p100-107.
- QUIROGA, Cesar A.; BULLOCK, Darcy, 1999, Travel time information using global positioning system and dynamic segmentation techniques, Transportation Research Record, n 1660, p48-57.
- Secrétariat à l'enquête Origine-Destination, 2005, Enquête origine-destination 2003, Agence métropolitaine de Transport, Montréal, 171p.
- SMITH, Brian ; VENKATANARAYANA, Ramkumar, 2005, Realizing the Promise of Intelligent Transportation Systems (ITS) Data Archives, Journal of Intelligent Transportation Systems, v. 9 , n. 4, p75.
- Société de Transport de l'Outaouais, 2006, Titres de transport & tarifs 2006, Société de Transport de l'Outaouais, Dépliant.
- Société de Transport de l'Outaouais, 2004, Rapport Annuel 2004, Société de Transport de l'Outaouais, 41p.
- STOPHER, Peter R., 1975, Urban transportation modeling and planning, Lexington, Mass.: Lexington Books, 345p.
- STOPHER, Peter R ; BULLOCK, P ; JIANG, Q, 2003, Visualising trips and travel characteristics from GPS data, Road and Transport Research, v. 12 , n. 2, p3.
- TCRP, 2003, Transit Capacity and Quality of Service Manual, Second Edition, TCRP Report 100, Transportation Research Board, 572p.

- TRANCHANT, Nicolas, 2005, Modèle de dérivation des déplacements en transport collectif à partir de données de cartes à puce, Mémoire de maîtrise en génie industriel, Ecole polytechnique de Montréal, 209p.
- TRÉPANIÉ, Martin; CHAPLEAU, Robert; TRANCHANT, Nicolas, 2005, Cartes à puces en transport en commun : une exploitation informationnelle pour fins de planification, 40e congrès de l'Association québécoise du transport et des routes, Laval, 14p.
- WOLF, J.; SCHONFELDER, S.; SAMAGA, U.; OLIVEIRA, M.; AXHAUSEN, K.W., 2004, Eighty weeks of global positioning system traces: Approaches to enriching trip information, Transportation Research Record, n 1870, p46-54.
- WOLF, Jean, 2004, Applications Of New Technologies In Travel Surveys, International Conference on Transport Survey Quality and Innovation, Costa Rica, 14p.
- WOLF, Jean; GUENSLER, Randall; BACHMAN, William, 2001, Elimination of the travel diary: Experiment to derive trip purpose from Global Positioning System travel data, Transportation Research Record, n 1768, p125-134.
- YALAMANCHILI, Lalit; PENDYALA, Ram M.; PRABAHARAN, N.; CHAKRAVARTHY, Pramodh, 1999, Analysis of global positioning system-based data collection methods for capturing multistop trip-chaining behavior, Transportation Research Record, n 1660, p58-65.
- YEE, Julie L.; NIEMEIER, Debbie, 1996, Advantages and Disadvantages: Longitudinal vs. Repeated Cross-Section Surveys, Project Battelle 94-16, FHWA, HPM-40, 7p.

ANNEXE A – Échantillons des tables de données

NumDéplacement	NbrTransactions	NumCarte	NumTitre	NumJour	DateDéplacement	HeureDéplacement	TypeTransaction	NumTrace	SensLigne	NumArrêt	NumVoiture	HeureLigne	NumBus	NumChauffeur
10414872	34	48626123	2	9	01-Jul-03	1213	1	61	0	3651	120	1203	9415	3209
10414873	29	3795188427	23	9	01-Jul-03	1256	1	260	0	3420	120	1235	9415	3209
10414874	154	2732285132	26	9	01-Jul-03	1333	9	439	0	2213	120	1311	9415	3209
10414875	209	580188619	1	9	01-Jul-03	1333	1	439	0	2213	120	1311	9415	3209
10414876	30	311952577	1	9	01-Jul-03	1334	1	439	0	2219	120	1311	9415	3209
10414877	249	43517121	1	9	01-Jul-03	1334	1	439	0	2219	120	1311	9415	3209
10414878	57	2465593803	27	9	01-Jul-03	1338	1	439	0	2742	120	1311	9415	3209
10414879	214	853604555	1	9	01-Jul-03	1345	1	439	0	2002	120	1311	9415	3209
10414880	23	2728427457	23	9	01-Jul-03	1346	1	439	0	2006	120	1311	9415	3209
10414881	8	847610049	23	9	01-Jul-03	1348	1	439	0	2026	120	1311	9415	3209

Table des embarquements

NumArrêt	LIBEL_ARRÊT	COORD X	COORD Y
1000	FRONT/CORMIER	18433093	5028191
1001	FRONT/DE LA TERRASSE-EARDLEY	18432986	5028635
1002	FRONT/CORMIER	18433105	5028185
1003	FRONT/DE LA TERRASSE-EARDLEY	18432975	5028568
1004	FRONT/PEARSON	18433083	5028046
1005	FRONT/DE LA TERRASSE-EARDLEY	18433080	5028429
1006	FRONT/PEARSON	18433071	5028067
1007	MCCONNELL-LARAMÉE/WILFRID LAVI	18433563	5028805
1008	FRONT/D'AIGUILLON	18433054	5027779
1009	MCCONNELL-LARAMÉE E	18433491	5028795

Table des arrêts

NumLigne	NumTrace	SansLigne	NumArrêt	OrdreArrêts	CodetLieu	Distance
5	5	1	5008	0	RKE	0
5	5	1	5010	1	RKE	130
5	5	1	5014	2	RKE	402
5	5	1	5022	3	RKE	867
5	5	1	5030	4	RKE	1217
5	5	1	5034	5	RKE	1433
5	5	1	5042	6	RKE	1733
5	5	1	5050	7	RKE	2083
5	5	1	6040	8	DHUL	2810
5	5	1	2612	9	HVL	3747

Table des arrêts-ligne

ANNEXE B – Requêtes SQL : Application des logiques

Une carte ne devrait pas être validée deux fois en même temps (ou à l'intérieur d'un certain intervalle)

Un système de validation devrait empêcher une carte d'être validée plus d'une fois à l'intérieur d'un même intervalle. Une requête groupera les transactions par carte-date-heure et ne se préoccupera que des trios qui apparaissent plus d'une fois.

```
SELECT Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement,
Count(Deplacements.HeureDeplacement) AS CountOfHeureDeplacement, Deplacements.TypeTransaction
FROM Deplacements
GROUP BY Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement,
Deplacements.TypeTransaction
HAVING (((Count(Deplacements.HeureDeplacement))>1));
```

Une carte ne devrait pas être validée au même moment dans deux bus distincts

La requête, très semblable à la précédente, regroupe par carte-date-heure et affiche le premier et dernier bus du groupement, toujours sans se préoccuper des occurrences qui n'apparaissent qu'une fois.

```
SELECT Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement,
Count(Deplacements.HeureDeplacement) AS CountOfHeureDeplacement, First(Deplacements.NumBus) AS
FirstOfNumBus, Last(Deplacements.NumBus) AS LastOfNumBus
FROM Deplacements
GROUP BY Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement
HAVING (((Count(Deplacements.HeureDeplacement))>1) AND ((First(Deplacements.NumBus))<>Last([numbus])));
```

Une carte ne devrait pas être validée au même moment à deux arrêts distincts

Encore une fois, on regroupe par carte-date-jour puis on réclame qu'on ait des arrêts différents entre la première et la dernière transaction. Aussi inclus directement dans la requête, les coordonnées géographiques liées aux arrêts afin de calculer la distance euclidienne qui les séparent.

```
SELECT Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement,
First(Deplacements.NumArret) AS FirstOfNumArret, Last(Deplacements.NumArret) AS LastOfNumArret,
First(Arrets.COORD_X) AS FirstOfCOORD_X, First(Arrets.COORD_Y) AS FirstOfCOORD_Y, Last(Arrets.COORD_X) AS
LastOfCOORD_X, Last(Arrets.COORD_Y) AS LastOfCOORD_Y
FROM Deplacements INNER JOIN Arrets ON Deplacements.NumArret = Arrets.NumArret
GROUP BY Deplacements.NumCarte, Deplacements.DateDeplacement, Deplacements.HeureDeplacement
```

HAVING (((First(Deplacements.NumArret))<>Last([deplacements].[numarret])));

Une carte devrait voir ses correspondances être accomplies à l'intérieur d'un certain laps de temps, après une transaction préalable

On veut analyser l'écart temporel entre une validation standard et une correspondance subséquente. Pour ce faire il est important de bien définir ce qu'est qu'une correspondance. On arrive facilement à identifier une correspondance par le type de transaction (type 3). Ce qui est plus délicat, c'est de retrouver la transaction précédente de la même carte. On se restreint évidemment à la même journée et on prend en compte le numéro de la transaction qui devrait être d'une unité inférieure à celle de la correspondance.

Dû à des considérations techniques, nous ne nous sommes attardé qu'aux transactions d'un jour moyen de semaine (2 juillet 2003 – 14 283 transactions, 2075 correspondances). Celles-ci ont d'abord été triées par le numéro de carte, puis par l'heure de transaction et enfin par le numéro de transaction. Le fait de trier par l'heure de transaction avant le numéro de transaction est attribuable à deux causes :

- il arrive fréquemment que le numéro de transaction ne s'incrémente pas;
- le numéro de transaction suit un cycle de 256 valeurs. Ainsi, le numéro suivant 255 est 0.

Puis de façon linéaire, un script s'est attardé aux transactions de type « correspondance » puis (pour la même carte) a comparé l'heure de la transaction de correspondance avec l'heure de la transaction précédente.

Une carte devrait avoir ses numéros de transaction séquentiels et continus

Nous sommes intéressés à observer le comportement de ces numéros de transaction puisqu'ils pourraient mener à confirmer ou infirmer l'assumption de données manquantes.

Avant de pouvoir comparer les numéros de transaction, il importe de préparer la table de transactions afin de parvenir à des résultats justes. Ainsi, la table doit être triée selon les critères ordonnés suivants :

➔ Date ➔ Numéro de la carte ➔ Heure de la transaction ➔ Numéro de la transaction

On a déjà expliqué précédemment pourquoi il était important de trier par l'heure de transaction avant le numéro de transaction. Un script VBA parcourt alors la table et note les variations suspectes (différent de 1) des numéros de transaction.

Une carte devrait avoir ses numéros de transaction quotidiens chronologiques

La vérification de cette logique n'est pas directe puisque on ne peut pas se fier au numéro de transaction pour établir la séquence des chaînes de déplacement d'un individu. En effet, à l'intérieur d'une même journée, un numéro de transaction peut se répéter ou « reculer » à 2. On aura donc recours à un script qui établit la véritable séquence des transactions d'un usager par jour.

Un véhicule ne devrait pas se déplacer au-delà d'une certaine vitesse

On désire valider la vitesse entre deux transactions du même bus. On prend un échantillon d'une journée de semaine (mardi 2 juillet, 14 283 transactions) comprenant le numéro de bus, l'heure de transaction et les coordonnées X et Y de l'arrêt associé à la transaction. Ayant les coordonnées dans le système UTM, le calcul de la distance parcourue par le temps écoulé est plutôt direct. Pour chaque transaction, on ne calculera la vitesse avec la transaction précédente seulement si :

- le véhicule est le même,
- l'écart temporel n'est pas nul,
- l'écart temporel est inférieur à un certain seuil,

- aucune des quatre coordonnées n'est 0. (erreur de repérage d'arrêt)

On utilise un écart temporel minimal dans le but d'éliminer les vitesses entre deux transactions où il y aurait eu un battement, une pause ou une perturbation de service. Cela nous évite de prendre en considération un changement de départ ou de chauffeur. Dans ces cas, il serait question de vitesses anormalement basses alors qu'ici, ce qui nous intéresse est plutôt d'observer ce qui se passe dans le cas de vitesses anormalement élevées.

Un véhicule ne devrait pas rester trop longtemps au même arrêt

Afin de déterminer les erreurs potentielles où un véhicule s'arrête trop longtemps, une requête regroupera les transactions par jour-bus-trace-sens-départ-arrêt (2 juillet 2003 uniquement) et on s'intéressera par la suite à la différence temporelle entre la première et la dernière transaction de chaque regroupement. Aussi, afin d'alléger la taille des résultats on ne retient que les regroupements où les deux heures sont différentes.

```
SELECT Deplacements.DateDeplacement, Deplacements.NumBus, Deplacements.NumTrace,
Deplacements.SensLigne, Deplacements.HeureLigne, Deplacements.NumArret, Min(Deplacements.HeureDeplacement)
AS MinOfHeureDeplacement, Max(Deplacements.HeureDeplacement) AS MaxOfHeureDeplacement
FROM Deplacements
GROUP BY Deplacements.DateDeplacement, Deplacements.NumBus, Deplacements.NumTrace,
Deplacements.SensLigne, Deplacements.HeureLigne, Deplacements.NumArret
HAVING (((Deplacements.DateDeplacement)=#7/2/2003#) AND ((Deplacements.NumTrace)<>900) AND
((Min(Deplacements.HeureDeplacement))<>Max([deplacements].[heureDeplacement])));
```

Un véhicule ne devrait pas être à deux arrêts différents au même moment

Une variante de la requête précédente est utilisée.

Un véhicule devrait voir ses transactions quotidiennes survenir chronologiquement

Une combinaison de script VBA et d'Excel est utilisée. Le script trie tous les enregistrements par date, numéro de carte et séquence de transaction puis les parcourt en repérant les transactions consécutives d'un usager-jour qui ne sont pas chronologiques.

Un chauffeur ne devrait pas être dans deux bus en même temps (ou à l'intérieur d'un certain intervalle)

Avec une requête qui groupe cette fois les transactions par date-heure-chauffeur et qui compare le premier bus du groupement avec le dernier, on arrive facilement à identifier les cas problématiques.

```
SELECT Deplacements.DateDeplacement, Deplacements.HeureDeplacement, Deplacements.NumChauffeur,
First(Deplacements.NumBus) AS FirstOfNumBus, Last(Deplacements.NumBus) AS LastOfNumBus,
Count(Deplacements.NumChauffeur) AS CountOfNumChauffeur
FROM Deplacements
GROUP BY Deplacements.DateDeplacement, Deplacements.HeureDeplacement, Deplacements.NumChauffeur
HAVING (((First(Deplacements.NumBus))<>Last([numbus])));
```

Un chauffeur ne devrait pas avoir un quart de travail dérivé par tournée (amplitude) ou par jour (pièce de travail) dépassant un certain laps de temps

On aimerait détecter les erreurs potentielles à travers les périodes de présence des chauffeurs. On utilise une requête qui groupe les transactions par chauffeur-jour et qui compare l'heure de la première transaction avec la dernière, l'écart sera appelé ici amplitude dérivée.

```
SELECT Deplacements.DateDeplacement, Deplacements.NumChauffeur, Min(Deplacements.HeureDeplacement) AS
MinOfHeureDeplacement, Max(Deplacements.HeureDeplacement) AS MaxOfHeureDeplacement
FROM Deplacements
GROUP BY Deplacements.DateDeplacement, Deplacements.NumChauffeur;
```

On peut observer comment ces amplitudes sont structurées lorsqu'on inclut le bus utilisé, ce qui équivaut à parler ici de pièce de travail. On regroupe donc les transactions par chauffeur-jour-bus puis on compare l'heure de la première et dernière transaction, d'où l'appellation pièce de travail dérivée.

```
SELECT Deplacements.DateDeplacement, Deplacements.NumBus, Deplacements.NumChauffeur,
Min(Deplacements.HeureDeplacement) AS MinOfHeureDeplacement, Max(Deplacements.HeureDeplacement) AS
MaxOfHeureDeplacement
FROM Deplacements
GROUP BY Deplacements.DateDeplacement, Deplacements.NumBus, Deplacements.NumChauffeur;
```

Un départ ne devrait être assigné qu'à un seul véhicule

On regroupe par date-tracé-sens-départ avec les valeurs des premiers/derniers NumBus et NumChauffeur afin de voir quand ceux-ci ne sont identiques.

```
SELECT Deplacements.DateDeplacement, Deplacements.NumTrace, Deplacements.SensLigne,
```

```
Deplacements.HeureLigne, Min(Deplacements.HeureDeplacement) AS MinOfHeureDeplacement,  
Max(Deplacements.HeureDeplacement) AS MaxOfHeureDeplacement, First(Deplacements.NumChauffeur) AS  
FirstOfNumChauffeur, Last(Deplacements.NumChauffeur) AS LastOfNumChauffeur, First(Deplacements.NumBus) AS  
FirstOfNumBus, Last(Deplacements.NumBus) AS LastOfNumBus INTO Doublons_de_departs_dans_un_jour4  
FROM Deplacements  
GROUP BY Deplacements.DateDeplacement, Deplacements.NumTrace, Deplacements.SensLigne,  
Deplacements.HeureLigne  
HAVING (((Deplacements.NumTrace)<>900) AND ((Deplacements.HeureLigne)<>"0000"));
```

ANNEXE C – Script évaluant les numéros de séquence des transactions d'un usager-jour

```

Sub Sequence_NoTransactions()
'Détermine les séquences d'une carte-jour et l'écrit dans un nouveau champ

Dim arrNoTr(40), arrSeq(40) As Integer
Dim rsCartes As Recordset
Set rsCartes = CurrentDb.OpenRecordset("SELECT NumCarte FROM Deplacements GROUP BY NumCarte",
    dbOpenForwardOnly)
While Not rsCartes.EOF
'POUR CHAQUE NUMÉRO DE CARTE DE LA BD
    cetteCarte = rsCartes!NumCarte
    NbrCarte = NbrCarte + 1
    CurrentDb.Properties!AppTitle = "Carte : " & NbrCarte    ' Change title bar
    Application.RefreshTitleBar                               ' Update title bar on screen

    For intJour = 1 To 31
    'POUR CHAQUE JOUR DE JUILLET

        dtJour = "7/" & intJour & "/2003"
        sqlParCarte = "SELECT NbrTransactions, SeqTransactions" & _
            " FROM Deplacements" & _
            " WHERE NumCarte = " & cetteCarte & _
            " AND DateDeplacement = #" & dtJour & "#" & _
            " ORDER BY NbrTransactions, HeureDeplacement"
        Set rsParCarte = CurrentDb.OpenRecordset(sqlParCarte)

        If rsParCarte.EOF = False Then
            Erase arrNoTr
            Erase arrSeq
            With rsParCarte

                'EMMAGASINE LES NUMÉROS DE TRANSACTION
                i = 0
                While .EOF = False
                    i = i + 1
                    arrNoTr(i) = !NbrTransactions
                    .MoveNext
                Wend
                arrNoTr(i + 1) = 999

                'ÉTABLIT LA SÉQUENCE
                Call TrouveSequence(arrNoTr, arrSeq)

                'ÉCRIT LES NUMÉROS DE SÉQUENCE À LA BD
                noSeq = 0
                .MoveFirst
                DAO.DBEngine.BeginTrans
                While .EOF = False
                    noSeq = noSeq + 1
                    .Edit
                    !SeqTransactions = arrSeq(noSeq)
                    .Update
                    .MoveNext
                Wend
                DAO.DBEngine.CommitTrans
                .Close

            End With
            Set rsParCarte = Nothing
        End If
    Next intJour
End While
End Sub

```

```

        'PROCHAIN JOUR
        Next intJour

        'PROCHAIN NUMÉRO DE CARTE
        rsCartes.MoveNext
    Wend

    rsCartes.Close
End Sub

```

Function TrouveSequence(arrSource As Variant, ByRef arrSeq As Variant)
 'Étant donné un tableau comportant les numéros de transactions, établit les séquences

```

    i = 1
    While arrSource(ind + 1) <> 999
        ind = ind + 1
        NoTr = arrSource(ind)
        If ind = 1 Then
            DiffNoTr = 1
        Else
            DiffNoTr = NoTr - NoTrPR
        End If

        Select Case DiffNoTr
            Case 0
                'Même s'il n'y a pas de variation, on incrémente
                arrSeq(ind) = ind
            Case Is > 3
                'C'est une réinitialisation du NoTransaction. On repart à 1 et
                'on incrémentera à chaque tour, toutes les valeurs le précédant
                arrSeq(ind) = 1
                saut = ind
            Case Else '(1, 2 ou 3)
                'Variation "normale", on incrémente de l'ordre de la variation
                If ind = 1 Then
                    arrSeq(ind) = 1
                Else
                    arrSeq(ind) = arrSeq(ind - 1) + DiffNoTr
                End If
            End Select

        'Au besoin, incrémente les valeurs précédant la valeur qui a réinitialisé
        If saut <> 0 Then
            For v = 1 To saut - 1
                arrSeq(v) = arrSeq(v) + 1
            Next v
        End If

        NoTrPR = NoTr
    Wend

    arrSeq(ind + 1) = 999
End Function

```

ANNEXE D – Script de caractérisation et de validation des correspondances

```

Sub Determination_des_noeuds_de_corresp()
#####
'### Pour chaque transaction de correspondance, ###
'### retrouve la validation simple la précédant, détermine l'arrêt de descente ###
'### et évalue la validité de la correspondance ###
#####

'### DÉCLARATION ET INITIALISATION ###
titre = Application.Caption
Debut = Now
PathDB = "D:\\"
NameDB = "BD STO - Juillet 2003.mdb"
tableSTO = "Deplacements"
Dim ArrTransfer As Integer, DistTransfer As Integer, Err As String
Dim TraceSens1 As String, ArrValid As Integer, ArrCorr As Integer
Dim connSTO As New ADODB.Connection
Dim rsToutCorr As New ADODB.Recordset
Dim rsValidat As New ADODB.Recordset
connSTO.ConnectionString = "Provider=Microsoft.Jet.OLEDB.4.0;Data Source= " & _
    PathDB & NameDB & ";Mode= Share Deny None"
connSTO.Open

'### SÉLECTIONNE TOUTES LES CORRESPONDANCES RÉALISÉES ###
strsql = "SELECT NumCarte, DateDeplacement, SeqTransactions, NumTrace,
    SensLigne, HeureLigne, NumArret, HeureDeplacement" & _
    " FROM " & tableSTO & _
    " WHERE TypeTransaction = 3"
rsToutCorr.Open strsql, connSTO, adOpenForwardOnly, adLockReadOnly
rsToutCorr.MoveFirst

'### POUR CHAQUE CORRESPONDANCE ###
While rsToutCorr.EOF = False

    intCarte = rsToutCorr!NumCarte
    dateDate = rsToutCorr!DateDeplacement
    TraceSens2 = rsToutCorr!Numtrace & " _" & rsToutCorr!SensLigne
    HrLigne2 = rsToutCorr!HeureLigne
    ArrCorr = rsToutCorr!NumArret
    Err = ""

    '### TROUVE LA VALIDATION PRÉCÉDENTE (MÊME USAGER-JOUR) ###
    strsql = "SELECT SeqTransactions, NumTrace, SensLigne, HeureLigne, NumArret, HeureDeplacement" & _
        " FROM " & tableSTO & _
        " WHERE DateDeplacement = #" & dateDate & " #" & _
        " AND NumCarte = " & intCarte & _
        " ORDER BY SeqTransactions DESC"
    rsValidat.Open strsql, connSTO, adOpenForwardOnly, adLockReadOnly
    rsValidat.MoveFirst
    While rsValidat!SeqTransactions <> rsToutCorr!SeqTransactions
        rsValidat.MoveNext
    Wend
    rsValidat.MoveNext

    If rsValidat.EOF Then
        '### ERREUR: Pas d'arrêt de validation pour cette correspondance ###
        TraceSens1 = ""
        HrLigne1 = ""
        ArrValid = 0
        Err = Err & "B"
    End If
End While
End Sub

```

```

Else
    TraceSens1 = rsValidat!Numtrace & " " & rsValidat!SensLigne
    HrLigne1 = rsValidat!HeureLigne
    ArrValid = rsValidat!NumArret

    If (rsValidat!Numtrace >= 500) Or (ArrValid <= 2) _
    Or (rsToutCorr!Numtrace >= 500) Or (ArrCorr <= 2) Then
        '### ERREUR: Invalidité d'un arrêt (1,2) ou tracé-sens (900,500) ###
        ArrTransfer = 0
        DistTransfer = 0
        Err = Err & "C"
    Else
        '### TROUVE L'ARRET DE DESCENTE LE PLUS PROBABLE ###
        Call TrouveArrTransfer(TraceSens1, ArrValid, ArrTransfer, ArrCorr, DistTransfer, Err)
        If DistTransfer > 1000 Then
            '### ERREUR: Distance de marche > 1km ###
            Err = Err & "E"
        End If

        If DiffTemps(rsValidat!HeureDeplacement, rsToutCorr!HeureDeplacement) > 119 Then
            '### ERREUR: Laps avant correspondance >= 2hrs ###
            Err = Err & "F"
        End If
    End If

End If

'### ÉCRIT LES RÉSULTATS ###
NbResultats = NbResultats + 1
If Err = "" Then Err = "-"
With ActiveWorkbook.Sheets("Résultats")
    .Range("A" & NbResultats + 1) = NbResultats
    .Range("B" & NbResultats + 1) = ArrValid
    .Range("C" & NbResultats + 1) = TraceSens1
    .Range("D" & NbResultats + 1) = HrLigne1
    .Range("E" & NbResultats + 1) = ArrTransfer    "Dérivé"
    .Range("F" & NbResultats + 1) = ArrCorr
    .Range("G" & NbResultats + 1) = TraceSens2
    .Range("H" & NbResultats + 1) = HrLigne2
    .Range("I" & NbResultats + 1) = DistTransfer    "Dérivé"
    .Range("J" & NbResultats + 1) = Err            "Dérivé"
End With
ArrValid = 0: TraceSens1 = "": HrLigne1 = "": ArrTransfer = 0: ArrCorr = 0: TraceSens2 = "": HrLigne2 = "":
DistTransfer = 0: Err = ""
rsValidat.Close
Set rsValidat = Nothing
If NbResultats Mod 100 = 0 Then
    Application.Caption = titre & " (Progrès: " & NbResultats & ")"
End If

'### PASSE À LA PROCHAINE CORRESPONDANCE ###
rsToutCorr.MoveNext
Wend

rsToutCorr.Close
MsgBox Format(Now - Debut, "hh:mm:ss")
End Sub

```

```

Function DiffTemps(t1 As String, t2 As String) As Integer
'### Retourne l'intervalle temporel (nb min) entre deux heures
    Temps1 = 60 * Left(t1, 2) + Right(t1, 2)
    Temps2 = 60 * Left(t2, 2) + Right(t2, 2)
    DiffTemps = Temps2 - Temps1
End Function

```

```

Public Function TrouveArrTransfer(TraceSensValid As String, _
    ArrValid As Integer, _
    ByRef ArrTransfer As Variant, _
    ArrCorresp As Integer, _
    ByRef DistTransfer As Variant, _
    ByRef Err As Variant)

'#####
'### Ayant (le tracé-sens et l'arrêt de validation) ###
'### ET (le tracé-sens et l'arrêt de correspondance), ###
'### trouve l'arrêt de descente le plus probable (+proche) ###
'#####

ActiveWorkbook.Sheets("Calculs").Range("B:H").ClearContents
With ActiveWorkbook.Sheets("RéfArrêtsLigne")

    '### Écrit les données du 2e arrêt (correspondance) ###
    Set c = .Range("C:C").Find(ArrCorresp, LookIn:=xlValues)
    If c Is Nothing Then
        Err = Err & "G" 'ERREUR: L'arrêt de correspondance est inconnu à la table des arrêts-ligne
    Else
        CoordX = .Range("D" & c.Row): CoordY = .Range("E" & c.Row)
    End If
    ActiveWorkbook.Sheets("Calculs").Range("F1") = ArrCorresp
    ActiveWorkbook.Sheets("Calculs").Range("G1") = CoordX
    ActiveWorkbook.Sheets("Calculs").Range("H1") = CoordY
    j = 1

    '### Cherche tous les arrêts-ligne subséquents au 1er arrêt (validation) ###
    Set c = .Range("A:A").Find(TraceSensValid, LookIn:=xlValues)
    If c Is Nothing Then
        Err = Err & "H" 'ERREUR: Le tracé-sens est inconnu à la table des arrêts-ligne
    Else
        i = c.Row
        trouve = False
        While .Range("A" & i).Value = TraceSensValid
            If .Range("C" & i) = ArrValid Then
                trouve = True 'Ne s'occupe qu'à partir du moment où on trouve l'arrêt dans la séquence de l'arrêt-ligne
            End If
            If trouve = True Then
                '### Écrit les infos de ces arrêts dans Excel qui déterminera le minimum ###
                ActiveWorkbook.Sheets("Calculs").Range("B" & j) = .Range("C" & i)
                ActiveWorkbook.Sheets("Calculs").Range("C" & j) = .Range("D" & i)
                ActiveWorkbook.Sheets("Calculs").Range("D" & j) = .Range("E" & i)
                j = j + 1
            End If
            i = i + 1
        Wend

        '### Résultat : Plus petite distance + Arrêt ###
        If j = 1 Then
            DistTransfer = 0: ArrTransfer = 0
            Err = Err & "A" 'ERREUR: Arrêt de validation inconnu au tracé-sens
        Else
            DistTransfer = ActiveWorkbook.Sheets("Calculs").Range("J3") 'Distance euclidienne calculée par Excel
            ArrTransfer = ActiveWorkbook.Sheets("Calculs").Range("K3") 'Arrêt où dist. est minimale (déterminé par Excel)
            If j = 2 Then Err = Err & "D" 'ERREUR: Arrêt de validation = Dernier arrêt du tracé-sens
        End If

    End If
End With

End Function

```

ANNEXE E – Script de caractérisation des tracés-sens-départ depuis les embarquements réalisés

```

Sub DeterminationService()
'#####
'### Script traversant la table des tracés-sens-départ (TSD) et   ###
'### qui les caractérise à l'aide de la table des embarquements   ###
'#####

WriteToErrorLog ("### DÉBUT ###")
Dim arrDepIPrecTSD(40), arrDepISuivTSD(40), arrChauffeur(40) As String
Dim arrDernDepI(40), occurrence, intNbItem As Integer
Dim strDepart As String
Dim strItem As String
Dim DureeMoy1 As Double
Dim DureeMoy2 As Double
Dim nbMoy2 As Integer
Dim rsTousLesTSD, rsCeTSD, rsDEPL As Recordset
Set rsTousLesTSD = CurrentDb.OpenRecordset("SELECT * FROM TSD ORDER BY TSD_ID")
rsTousLesTSD.FillCache
rsTousLesTSD.MoveFirst

While Not rsTousLesTSD.EOF
'### POUR CHAQUE TRACÉ-SENS-DEPART DÉTECTÉ (TABLE "TSD") ###

    CurrentDb.Properties!AppTitle = "Progrès : " & rsTousLesTSD!TSD_ID ' Change title bar
    Application.RefreshTitleBar ' Update title bar on screen

    rsTousLesTSD.Edit
    intTrace = rsTousLesTSD!Trace
    intSens = rsTousLesTSD!Sens
    strDepart = rsTousLesTSD!Depart

    '### TROUVE TOUS LES OCCURANCES (EN SEM) OÙ APPARAÎT CE TRACÉ-SENS-DEPART ###
    strSQL_PourCeTSD = "SELECT NumTrace, SensLigne, HeureLigne, DateDeplacement, NumJour, " & _
        "FIRST(NumDeplacement) AS FirstNumDepl, " & _
        "FIRST(TourneeID) AS FirstTournee, " & _
        "FIRST(NumChauffeur) AS FirstChauffeur, " & _
        "LAST(NumDeplacement) AS LastNumDepl, " & _
        "LAST(TourneeID) AS LastTournee, " & _
        "LAST(NumChauffeur) AS LastChauffeur, " & _
        "LAST(HeureDeplacement) AS LastHrDepl " & _
        "FROM Deplacements " & _
        "GROUP BY NumTrace, SensLigne, HeureLigne, DateDeplacement, NumJour " & _
        "HAVING (NumTrace=" & intTrace & ") " & _
        "AND (SensLigne=" & intSens & ") " & _
        "AND (HeureLigne=" & strDepart & ") " & _
        "AND (NumJour < 6)"
    Set rsCeTSD = CurrentDb.OpenRecordset(strSQL_PourCeTSD)
    rsCeTSD.MoveLast
    nbCeTSD = rsCeTSD.AbsolutePosition + 1
    rsCeTSD.MoveFirst

    For occurrence = 1 To nbCeTSD
'### POUR CHAQUE OCCURANCE OÙ APPARAÎT CE TRACÉ-SENS-DEPART ###

        DateTSD = rsCeTSD!DateDeplacement
        FirstNumDepl = rsCeTSD!FirstNumDepl
        FirstTournee = rsCeTSD!FirstTournee
        LastNumDepl = rsCeTSD!LastNumDepl
        LastTournee = rsCeTSD!LastTournee
        If FirstTournee <> LastTournee Then

```

```

    '### Vérifie que le départ n'est que sur une tournée
    rsTousLesTSD!Suspect = 1
End If

'### TROUVE TOUS LES EMBARQUEMENTS DE CE BUS POUR CETTE TOURNÉE ###
strSQL_DateBus = "SELECT * FROM Deplacements " & _
    "WHERE DateDeplacement = #" & DateTSD & "#" & _
    "AND TourneeID = " & FirstTournee
Set rsDEPL = CurrentDb.OpenRecordset(strSQL_DateBus)
rsDEPL.MoveFirst

'### TROUVE LE DÉPLACEMENT PRÉCÉDENT CE TRACÉ-SENS-DÉPART ###
rsDEPL.FindFirst "NumDeplacement = " & FirstNumDepl
rsDEPL.MovePrevious
If rsDEPL.BOF Then
    arrDeplPrecTSD(occurance) = "PremierTSD" 'Pas de TSD précédent
Else
    TSD = rsDEPL!NumTrace & "_" & rsDEPL!SensLigne & "_" & rsDEPL!HeureLigne
    arrDeplPrecTSD(occurance) = TSD 'Capture le TSD précédent
End If

'### MODIFIE LA REQUÊTE SI IL Y A DEUX BUS EN PRÉSENCE ###
If FirstTournee <> LastTournee Then
    rsDEPL.Close
    strSQL_DateBus = "SELECT * FROM Deplacements " & _
        "WHERE DateDeplacement = #" & DateTSD & "#" & _
        "AND TourneeID = " & LastTournee
    Set rsDEPL = CurrentDb.OpenRecordset(strSQL_DateBus)
End If

'### TROUVE LE DÉPLACEMENT SUIVANT CE TRACÉ-SENS-DÉPART ###
rsDEPL.FindFirst "NumDeplacement = " & LastNumDepl
rsDEPL.MoveNext
If rsDEPL.EOF Then
    arrDeplSuivTSD(occurance) = "DernierTSD" 'Pas de TSD suivant
Else
    TSD = rsDEPL!NumTrace & "_" & rsDEPL!SensLigne & "_" & rsDEPL!HeureLigne
    arrDeplSuivTSD(occurance) = TSD 'Capture le TSD suivant
End If
rsDEPL.Close

'Capture le numéro de chauffeur
If rsCeTSD!FirstChauffeur <> rsCeTSD!LastChauffeur Then
    WriteToErrorLog ("Tournee " & FirstTournee & vbTab & DateTSD & vbTab & _
        "Nb de chauff pour cette tournée >1 (chang de chauff)")
    End If
    arrChauffeur(occurance) = rsCeTSD!FirstChauffeur
'Capture la durée calculée entre l'heure de départ planifiée et la dernière montée
If strDepart <> "0000" Then
    arrDernDepl(occurance) = IntervalleEnMinutes(strDepart, rsCeTSD!LastHrDepl)
Else
    arrDernDepl(occurance) = 0
End If

rsCeTSD.MoveNext
Next occurrence
'### PASSE À LA PROCHAINE OCCURANCE (JOUR) OÙ APPARAÎT CE TRACÉ-SENS-DEPART ###

arrDeplPrecTSD(occurance) = "fin" 'Ajoute un "bookmark de fin" pour l'array
arrDeplSuivTSD(occurance) = "fin"
arrChauffeur(occurance) = "fin"
arrDernDepl(occurance) = 32000

'### ÉVALUE ET ÉCRIT LES INDICATEURS À LA TABLE TSD ###
'Chauffeur de ce TSD
Call TrouveLePlusFrequent(arrChauffeur, stritem, intNbItem)

```

```

        rsTousLesTSD!Chauffeur = strItem
        rsTousLesTSD!ChSurMois = intNbItem
    'Durée
    Call Moyenne1et2(arrDernDepl, DureeMoy1, DureeMoy2, nbMoy2)
    rsTousLesTSD!DureeMoy1 = DureeMoy1
    rsTousLesTSD!DureeMoy2 = DureeMoy2
    rsTousLesTSD!nbMoy2 = nbMoy2
    DureeMoy1 = 0
    DureeMoy2 = 0
    nbMoy2 = 0
    'TSD précédent
    Call TrouveLePlusFrequent(arrDeplPrecTSD, strItem, intNbItem)
    rsTousLesTSD!PR_TSD = strItem
    rsTousLesTSD!PR_T_SurNb = intNbItem
    'Index du TSD précédent
    If Right(strItem, 3) <> "TSD" Then
        rsTousLesTSD!PR_Index = TrouverIndexTSD(strItem)
    Else
        rsTousLesTSD!SU_Index = 0 'Premier ou dernier TSD
    End If
    'TSD suivant
    Call TrouveLePlusFrequent(arrDeplSuivTSD, strItem, intNbItem)
    rsTousLesTSD!SU_TSD = strItem
    rsTousLesTSD!SU_T_SurNb = intNbItem
    'Index du TSD suivant
    If Right(strItem, 3) <> "TSD" Then
        rsTousLesTSD!SU_Index = TrouverIndexTSD(strItem)
    Else
        rsTousLesTSD!SU_Index = 0 'Premier ou dernier TSD
    End If
    rsTousLesTSD.Update

    'Efface les arrays
    Erase arrDeplPrecTSD
    Erase arrDeplSuivTSD
    Erase arrChauffeur
    Erase arrDernDepl

    rsTousLesTSD.MoveNext
Wend
'### PASSE A UN AUTRE TRACÉ-SENS-DÉPART ###

rsTousLesTSD.Close
WriteToErrorLog ("### FIN ###")
End Sub

```

```

'          #####
'          ##### SOUS-FUNCTIONS #####
'          #####

```

```

Sub WriteToErrorLog(Texte As String)
'### Écrit dans le fichier ErrorLog qui se trouve dans le même dossier que la BD ###

    Open CurrentProject.Path & "ErrorLog.txt" For Append As #1
    Print #1, Now & vbTab & Texte
    Close #1

End Sub

```

```

Function TrouveLePlusFrequent(arrSource As Variant, ByRef strPlusGrand As String, ByRef intPlusGrand As Integer)
'### Prends un array et renvoie l'item le plus fréquent avec le nb d'occurences ###

    Dim arrResult(), arrResult2() As Variant

```

```

ReDim arrResult1(0), arrResult2(0)
intDiffStr = 0
While arrSource(i) <> "fin"
    strCurrent = arrSource(i)

    '### Recherche la valeur dans celles déjà trouvées ###
    For intLookup = 0 To intDiffStr
        If strCurrent = arrResult1(intLookup) Then
            arrResult2(intLookup) = arrResult2(intLookup) + 1 'Incrémente le nb d'occ de l'item
            DejaLa = 1
        End If
    Next intLookup

    '### Cas où on aurait pas trouvé l'item ###
    If DejaLa = 0 Then
        arrResult1(intDiffStr) = strCurrent 'Ajoute l'item
        arrResult2(intDiffStr) = 1 'Mets 1 comme nb d'occ
        intDiffStr = intDiffStr + 1 'Incrémente le nb de différentes occurrences
        ReDim Preserve arrResult1(intDiffStr) 'Incrémente la dimension du array des items
        ReDim Preserve arrResult2(intDiffStr) 'Incrémente la dimension du array des valeurs
    End If

    DejaLa = 0
    i = i + 1
Wend

'### Recherche de l'item avec le plus d'occurrences ###
For tt = 0 To intDiffStr
    If arrResult2(tt) > seuil Then
        strPlusGrand = arrResult1(tt)
        intPlusGrand = arrResult2(tt)
        seuil = intPlusGrand
    End If
    tous = tous & " | " & arrResult1(tt) & "-" & arrResult2(tt)
Next tt

'MsgBox tous
End Function

```

```

Function IntervalleEnMinutes(t1 As String, t2 As String) As Integer
'On assume que les heures peuvent dépasser 2400

    h1 = Left(t1, 2): h2 = Left(t2, 2): m1 = Right(t1, 2): m2 = Right(t2, 2)
    IntervalleEnMinutes = (h2 - h1) * 60 + (m2 - m1)
    If IntervalleEnMinutes < 0 Then
        IntervalleEnMinutes = -IntervalleEnMinutes
    End If
End Function

```

End Function

```

Function Moyenne1et2(arrSource As Variant, ByRef Moyenne1 As Double, ByRef Moyenne2 As Double, ByRef nbMoy2 As Integer)
'### Calcule la moyenne et l'écart-type, puis recalcule la moyenne ###
'### sans prendre en compte les valeurs au-delà de cet écart-type. ###

```

```

    Dim EcartType As Double
    Dim intCurrent, intSumSquares, nb, i As Integer
    i = 1
    While arrSource(i) <> 32000
        testlog = testlog & vbLf & arrSource(i)
        intCurrent = intCurrent + arrSource(i)
        intSumSquares = intSumSquares + (arrSource(i) ^ 2)
        nb = nb + 1
    End While

```

```

    i = i + 1
Wend
Moyenne1 = (intCurrent / nb)
If i <> 2 Then EcartType = Format(Sqr(Abs((intSumSquares - ((i - 1) * (Moyenne1 ^ 2))) / (i - 2))), "0.00")
'Source: http://en.wikipedia.org/wiki/Standard\_deviation

'### Refait le calcul de la moyenne sans les valeurs ±10 ###
i = 1
nb = 0
intCurrent = 0
While arrSource(i) <> 32000
    If (Moyenne1 + EcartType - arrSource(i)) >= 0 Then
        intCurrent = intCurrent + arrSource(i)
        nbMoy2 = nbMoy2 + 1
    End If
    i = i + 1
Wend
Moyenne1 = Format(Moyenne1, "0.00")
Moyenne2 = Format((intCurrent / nbMoy2), "0.00")

End Function

```

```

Function TrouverIndexTSD(strTSD As String)
'### Retrouve l'index d'un tracé-sens-départ (ex : "37_0_1535") dans la table des TSD ###

Dim ceTSD As Recordset
premUS = InStr(1, strTSD, "_")
Trace = Left(strTSD, premUS - 1)
Sens = Mid(strTSD, premUS + 1, 1)
Depart = Right(strTSD, 4)

Set ceTSD = CurrentDb.OpenRecordset("SELECT TSD_ID FROM TSD WHERE Trace = " & Trace & _
                                     "AND Sens = " & Sens & _
                                     "AND Depart = " & Depart & "")

If Not ceTSD.EOF Then TrouverIndexTSD = ceTSD!TSD_ID

End Function

```

ANNEXE F – Script de caractérisation des profils d'utilisateur

```

Sub ProfilDesUsagers()
'#####
'### Analyse et caractérise chaque usager d'une table ###
'### comprenant préalablement les numéros de carte ###
'#####

'### MOIS ÉTUDIÉ ###
pathBD = "D:\donnees_cartes_a_puces_STO\"
nomBD = "BD1 STO - Juillet 2003.mdb": moisBD = "0307": NbJour = 31: NbJourSem = 22

Dim wrkJet As Workspace
Set wrkJet = CreateWorkspace("", "admin", "", dbUseJet)
Set TransactionsDb = wrkJet.OpenDatabase(pathBD & nomBD, True)
Dim arrTitre(400), arrArret(400) As Variant
Dim strPlusFrequent As String
Dim intPlusFrequent As Integer
Dim rsProfilsUsagers, rsUsagerJour As Recordset
strSQL = "SELECT * FROM SG_ProfilsUsagers WHERE `"` & moisBD & "`" = 1"
Set rsProfilsUsagers = CurrentDb.OpenRecordset(strSQL, dbOpenDynaset)

While rsProfilsUsagers.EOF = False
'#####
'### POUR CHAQUE USAGER... ###
'#####

    NbrCarte = NbrCarte + 1
    If NbrCarte Mod 10 = 0 Then
        CurrentDb.Properties!AppTitle = "Carte : " & NbrCarte ' Change title bar
        Application.RefreshTitleBar ' Update title bar on screen
    End If

    '### Réinitialise les arrays et variables utilisées pour un usager ###
    Erase arrTitre: Erase arrArret
    i = 0: NbTransactions = 0: NbValid = 0: NbCorresp = 0: NbRefus = 0: NbValidJSem = 0: NbCorrespJSem = 0:
    NbArretPlusFrequent = 0: ArretPlusFrequent = "": RatioPlusFrequent = 0: NbArret = 0: RatioArretPlusFrequent = 0:
    RatioTitrePlusFrequent = 0: intPlusFrequent = 0: strPlusFrequent = ""
    multiTitre = False

    '### Requête les transactions de cet usager ###
    strSQL = "SELECT TypeTransaction, NumArret, NumJour, NumTitre, SeqTransactions" & _
        " FROM Deplacements" & _
        " WHERE NumCarte = " & rsProfilsUsagers!NoCarte
    Set rsUsagerJour = TransactionsDb.OpenRecordset(strSQL, dbOpenForwardOnly, dbReadOnly)
    While rsUsagerJour.EOF = False

        '#####
        '### Passe à travers les transactions de cette usager ###
        '#####

        '### Titre de transport ###
        arrTitre(i) = rsUsagerJour!NumTitre
        If (i > 0) Then
            If (arrTitre(i - 1) <> arrTitre(i)) Then multiTitre = True 'Détermine quand cas où 1 carte = 2 types de titres
        End If

        '### Arrêt lorsque 1ere transaction de la journée ###
        If (rsUsagerJour!SeqTransactions = 1) Then
            arrArret(NbArret) = rsUsagerJour!NumArret
            NbArret = NbArret + 1
        End If
    End While
End While

```

```

'### Type de validations (correspondance, simple ou refus) ###
If (rsUsagerJour!TypeTransaction = 1) Then
    NbValid = NbValid + 1
ElseIf (rsUsagerJour!TypeTransaction = 3) Then
    NbCorresp = NbCorresp + 1
ElseIf (rsUsagerJour!TypeTransaction = 9) Then
    NbRefus = NbRefus + 1
End If

'### Type de validations (correspondance ou simple) en jour de semaine ###
If (rsUsagerJour!NumJour < 6) And (rsUsagerJour!TypeTransaction = 1) Then
    NbValidJSem = NbValidJSem + 1
ElseIf (rsUsagerJour!NumJour < 6) And (rsUsagerJour!TypeTransaction = 3) Then
    NbCorrespJSem = NbCorrespJSem + 1
End If

i = i + 1
rsUsagerJour.MoveNext

Wend
arrTitre(i) = "fin"
arrArret(NbArret) = "fin"

'#####
'### Analyse les infos collectées sur l'usager ###
'#####

'Arrêt-domicile
Call TrouveLePlusFrequent(arrArret, strPlusFrequent, intPlusFrequent)
ArretPlusFrequent = strPlusFrequent
If NbArret <> 0 Then
    RatioArretPlusFrequent = Round(intPlusFrequent / NbArret * 100, 0)
Else: Debug.Print "NbArret = 0 pour carte: " & rsProfilsUsagers!NoCarte
End If

'Titre de transport
If multiTitre = True Then
    Call TrouveLePlusFrequent(arrTitre, strPlusFrequent, intPlusFrequent)
    TitrePlusFrequent = strPlusFrequent
    RatioTitrePlusFrequent = Round(intPlusFrequent / i * 100, 0)
Else
    TitrePlusFrequent = arrTitre(0)
    RatioTitrePlusFrequent = 100
End If

'#####
'### Écrit à BD les résultats relatifs à l'usager ###
'#####

rsProfilsUsagers.Edit
'Remplit les champs du mois analysé
rsProfilsUsagers.Fields("ArretDomicile" & moisBD) = ArretPlusFrequent
rsProfilsUsagers.Fields("RatioArret" & moisBD) = RatioArretPlusFrequent
rsProfilsUsagers.Fields("Eloignement" & moisBD) = DistanceAuCentre(ArretPlusFrequent)
rsProfilsUsagers.Fields("Groupe" & moisBD) = Groupe(TitrePlusFrequent) 'Appelle fonction "Groupe"
rsProfilsUsagers.Fields("RatioGroupe" & moisBD) = RatioTitrePlusFrequent
rsProfilsUsagers.Fields("NbTrans" & moisBD) = i
rsProfilsUsagers.Fields("NbValid" & moisBD) = NbValid
rsProfilsUsagers.Fields("NbCorr" & moisBD) = NbCorresp
rsProfilsUsagers.Fields("NbRefus" & moisBD) = NbRefus
rsProfilsUsagers.Fields("DeplJour" & moisBD) = NbValid / NbJour
rsProfilsUsagers.Fields("DeplJSem" & moisBD) = NbValidJSem / NbJourSem
If NbValid <> 0 Then
    rsProfilsUsagers.Fields("RatioCV" & moisBD) = NbCorresp / NbValid
Else
    rsProfilsUsagers.Fields("RatioCV" & moisBD) = 0

```

```

End If
If NbValidJSem <> 0 Then
    rsProfilsUsagers.Fields("RatioJSemCV" & moisBD) = NbCorrespJSem / NbValidJSem
Else
    rsProfilsUsagers.Fields("RatioJSemCV" & moisBD) = 0
End If

rsProfilsUsagers.Update
Set rsUsagersJour = Nothing

'#####
'### PROCHAIN USAGER ###
'#####
rsProfilsUsagers.MoveNext
Wend
End Sub

```

```

Function TrouveLePlusFrequent(arrSource As Variant, ByRef strPlusGrand As String, ByRef intPlusGrand As Integer)
'### Prends un array et renvoie l'item le plus fréquent avec le nb d'occurrences ###

```

```

Dim arrResult(), arrResult2() As Variant
ReDim arrResult1(0), arrResult2(0)
intDiffStr = 0
While arrSource(i) <> "fin"
    strCurrent = arrSource(i)

    '### Recherche la valeur dans celles déjà trouvées ###
    For intLookup = 0 To intDiffStr
        If strCurrent = arrResult1(intLookup) Then
            arrResult2(intLookup) = arrResult2(intLookup) + 1 'Incrémente le nb d'occ de l'item
            DejaLa = 1
        End If
    Next intLookup

    '### Cas où on aurait pas trouvé l'item ###
    If DejaLa = 0 Then
        arrResult1(intDiffStr) = strCurrent 'Ajoute l'item
        arrResult2(intDiffStr) = 1 'Mets 1 comme nb d'occ
        intDiffStr = intDiffStr + 1 'Incrémente le nb de différentes occurrences
        ReDim Preserve arrResult1(intDiffStr) 'Incrémente la dimension du array des items
        ReDim Preserve arrResult2(intDiffStr) 'Incrémente la dimension du array des valeurs
    End If

    DejaLa = 0
    i = i + 1
Wend

'### Recherche de l'item avec le plus d'occurrences ###
For tt = 0 To intDiffStr
    If arrResult2(tt) > seuil Then
        strPlusGrand = arrResult1(tt)
        intPlusGrand = arrResult2(tt)
        seuil = intPlusGrand
    End If
    tous = tous & " | " & arrResult1(tt) & "-" & arrResult2(tt)
Next tt

```

```

End Function

```

```

Function Groupe(ByVal NoTitre As Integer) As String
'### Catégorise le titre de transport comme faisant ###
'### partie d'un des groupes Régulier, Étudiant ou Aîné ###

```

```

Dim arrR As Variant '## RÉGULIER ##
arrR = Array(1, 2, 3, 9, 10, 15, 16, 17, 42, 43)
Dim arrE As Variant '## ÉTUDIANT ##
arrE = Array(4, 5, 6, 7, 11, 12, 13, 14, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39,
40, 41, 44, 45, 46, 47, 48)
Dim arrA As Variant '## AÎNÉ ##
arrA = Array(8, 18)

For i = 0 To UBound(arrR)
    If arrR(i) = NoTitre Then Groupe = "r": Exit Function
Next i
For i = 0 To UBound(arrE)
    If arrE(i) = NoTitre Then Groupe = "e": Exit Function
Next i
For i = 0 To UBound(arrA)
    If arrA(i) = NoTitre Then Groupe = "a": Exit Function
Next i
Groupe = "Z" 'ERR: Le NoTitre n'est pas répertorié

End Function

```

```

Function DistanceAuCentre(ByVal ArretDomicile As String) As Double
'Retourne la distance (en mètres) entre une paire de coordonnées et celles de l'arrêt utilisé comme centre

If ArretDomicile <> "1" And ArretDomicile <> "2" Then
    Dim rsCoordArretDomicile As Recordset
    strSQL = "SELECT COORD_X, COORD_Y FROM Arrets WHERE NumArret=" & ArretDomicile
    Set rsCoordArretDomicile = CurrentDb.OpenRecordset(strSQL, dbOpenForwardOnly, dbReadOnly)
    If rsCoordArretDomicile.EOF = True Then
        DistanceAuCentre = 9999 'ERR: Arrêt non répertorié
        Exit Function
    End If
    X0 = 445007
    Y0 = 5030010
    X1 = rsCoordArretDomicile!COORD_X - 18000000
    Y1 = rsCoordArretDomicile!COORD_Y
    DistanceAuCentre = Sqr((X0 - X1) ^ 2 + (Y0 - Y1) ^ 2)
Else
    DistanceAuCentre = 0
End If

End Function

```