

Contributions des données de cartes à puces à la compréhension des rythmes d'activités en transport

Catherine Morency (cmorency@polymtl.ca), Ing., Ph.D., professeure adjointe¹

Martin Trépanier (mtrepanier@polymtl.ca), Ing., Ph.D., professeur agrégé^{1,2}

Bruno Agard (bagard@polymtl.ca), Ph.D., professeur adjoint²

¹ Groupe MADITUC en planification des transports urbains

² Groupe Polygistique

Département des génies civil, géologique et des mines

Département de mathématiques et génie industriel

École Polytechnique
C.P. 6079, succursale Centre-Ville
Montréal, Québec
H3C3A7



Association Québécoise du transport et des routes
9 au 11 avril 2006
Québec

Contributions des données de cartes à puces à la compréhension des rythmes d'activités en transport

Catherine Morency, Martin Trépanier et Bruno Agard

Résumé

Les systèmes de paiement par cartes à puces sont des outils dont la finalité fondamentale est de gérer la perception des frais inhérents à l'utilisation des services de transport en commun métropolitains. En effet, plusieurs bénéfices opérationnels peuvent être attendus de tels systèmes, tels que la validation des compatibilités entre titre et zone tarifaire, la limitation des risques d'usage frauduleux des titres de transport ainsi que la facilitation de l'usage combiné de services offerts par des autorités organisatrices de transport différentes.

Ces systèmes peuvent toutefois avoir d'autres retombées de nature analytique. En effet, les transactions effectuées par les usagers en différents points du réseau de transport en commun sont enregistrées dans une base de données qui peut être par la suite exploitée à des fins d'observation de la mobilité sur le réseau.

L'objectif de cet article est de discuter des différentes sources de données permettant d'observer les comportements individuels de mobilité et d'identifier les contributions potentielles particulières des données issues des systèmes de perception par cartes à puces.

L'article s'appuie sur des données provenant du système de paiement par cartes à puces de la Société de transport de l'Outaouais (STO). La STO est le leader québécois en matière de cartes à puces. Les données sont analysées afin de développer des indicateurs d'activité et de régularité pertinents, le tout basé sur la littérature en la matière. Ces indicateurs permettent une classification sommaire des usagers, dans le respect de la confidentialité d'usage : en aucun temps, ces données ne peuvent être associées à des noms de personnes.

Ces données ne viennent cependant pas remplacer les autres méthodes d'acquisition de données actuellement utilisées par différents agents institutionnels. Elles peuvent, au mieux, contribuer à une meilleure compréhension de différents aspects très spécifiques de la mobilité du sous-ensemble de la population faisant usage du réseau de transport en commun. L'article discute de cette complémentarité éventuelle.

Contributions des données de cartes à puces à la compréhension des rythmes d'activités en transport

Catherine Morency, Martin Trépanier et Bruno Agard

1. Introduction

Les systèmes de paiement par cartes à puces sont des outils dont la finalité fondamentale est de gérer la perception des frais inhérents à l'utilisation des services de transport en commun. En effet, différents bénéfices opérationnels sont attendus de tels systèmes : validation des compatibilités entre titre et zone tarifaire, limitation des risques d'usage frauduleux des titres de transport ou facilitation de l'usage combiné de services offerts par des autorités organisatrices de transport différentes (concept d'intégration tarifaire).

Ces systèmes peuvent toutefois avoir d'autres retombées de nature analytique. En effet, les transactions effectuées par les usagers en différents points du réseau de transport en commun sont enregistrées dans une base de données qui peut, par la suite, être exploitée à des fins d'observation des comportements de mobilité des usagers sur le réseau.

L'objectif de cet article est de discuter des potentialités analytiques des données de cartes à puces en faisant ressortir les particularités de ces ensembles par rapport aux différentes sources de données permettant d'observer les comportements individuels de mobilité.

Dans cette perspective, l'article aborde les thématiques suivantes :

- Détails méthodologiques et potentialités d'analyse des différents ensembles de données et de microdonnées permettant d'observer les comportements individuels de mobilité;
- Particularités informationnelles de données de transactions provenant des systèmes de perception par cartes à puces;
- Expérimentations analytiques visant à mesurer la variabilité comportementale des usagers du réseau de transport notamment :
 - Structure hebdomadaire d'activités et régularité sur une période de 10 mois;
 - Régularité quotidienne des comportements individuels des usagers sur une période de 277 jours et classification des jours selon les comportements observés;
 - Variabilité spatio-temporelle de la mobilité quotidienne.

2. Les systèmes informationnelles en transport

Différentes sources de données sont impliquées dans la constitution d'un système informationnel permettant de mesurer les comportements individuels de mobilité. De façon générale, les systèmes d'information rassemblent des données sur le territoire (adresses, codes postaux, intersections, orthophotos), les réseaux de transport (routiers, de transport en commun, camionnage, transport adapté) et la demande.

Les données concernant la demande de transport peuvent être issues de deux types d'opérations : les enquêtes (et recensements) qui exigent une contribution particulière des usagers d'un système et les systèmes d'observation qui permettent de recueillir des données sans aucune intervention de la part des usagers.

2.1 Enquêtes vs systèmes d'observation

Dans la première classe, on retrouve deux principaux types d'enquêtes qui fournissent des détails sur l'utilisation des réseaux de transport en commun :

- Enquêtes ménages régionales Origine-Destination

Les enquêtes Origine-Destination sont réalisées cycliquement dans plusieurs grandes régions du Québec notamment à Montréal, Québec, Gatineau-Ottawa, Sherbrooke et Trois-Rivières. Dans la Grande Région de Montréal, ces enquêtes sont tenues environ tous les cinq ans, depuis 1970, et permettent de recueillir des informations détaillées sur la mobilité quotidienne des résidents. En effet, toutes les personnes d'environ 5% des ménages résidant à l'intérieur du périmètre d'études fournissent des détails spatio-temporels à propos des déplacements qu'ils ont effectués lors d'un jour spécifique d'automne. Ces enquêtes fournissent un portrait ponctuel de la mobilité d'une population résidente. À titre indicatif, l'échantillon complet recueilli pour lors de la plus récente enquête comporte 170 000 personnes appartenant à environ 70 000 ménages.

Ces enquêtes se distinguent notamment par l'ampleur de l'échantillon recueilli, par les détails spatio-temporels recueillis concernant tous les modes de déplacements ainsi que par le fait qu'elles portent sur un seul jour d'observation. La résultante est généralement la constitution d'indicateurs de mobilité portant sur un jour moyen d'automne.

- Enquêtes effectuées à bord des véhicules de transport en commun

Les enquêtes effectuées à bord des véhicules de transport en commun s'intéressent uniquement à la clientèle d'un réseau particulier de transport en commun. Dans la Grande Région de Montréal, elles sont notamment effectuées par la STL et la RTL sur leurs réseaux respectifs d'autobus ainsi que par l'AMT sur les lignes de train de banlieue. Ce type d'enquête permet de recueillir des informations spatio-temporelles sur le déplacement en cours d'exécution par les usagers échantillonnés. Le cumul de ces observations, cumulées sur plusieurs jours d'observation, permet aussi d'estimer l'usage moyen du réseau lors d'un jour moyen.

Il existe aussi différents systèmes d'observation qui, bien que cela ne soit pas nécessairement leur finalité première, permettent d'observer les comportements de différents usagers :

- Les systèmes de comptage automatique de passagers et de localisation spatiale (GPS)

Les systèmes de comptage automatique avec localisation spatiale par SPG (système de positionnement global) permettent d'une part de comptabiliser le nombre de montants et descendants à chaque arrêt du réseau de transport et d'autre part de suivre, en temps réel ou différé, l'adéquation entre les heures de passages annoncées (service planifié) et les heures de passage réelles (service opéré). Tous les usagers faisant usage du réseau sont observés par ce système mais aucun ne peut être suivi dans ses mouvements individuels.

- Les systèmes de perception par cartes à puces

Les systèmes de perception par cartes à puces permettent de contrôler l'accès aux véhicules de transport en commun en effectuant une validation des titres de transport utilisés. Cette validation s'effectue lors de toute montée à bord (amorce d'un déplacement ou correspondance). Il ne s'agit donc pas d'un échantillonnage mais bien du suivi quotidien de toutes les validations effectuées par tous les usagers disposant d'un titre de transport compatible au système de perception.

Sommairement, ces systèmes rassemblent les détails de toutes les transactions d'accès effectuées sur le réseau de transport en commun par les détenteurs d'une carte à puce, notamment les informations spatio-temporelles des transactions.

2.2 Particularités et potentialités générales des données de cartes à puces

Les chercheurs Bagchi et White (2004) ont exposé les principaux intérêts des données issues des systèmes de cartes à puces pour les planificateurs de transport, en comparaison avec les autres technologies existantes de validation des titres, de comptage des passagers, ou encore avec d'autres sources de données :

- un accès à un volume plus large de données sur les trajets individuels;
- la possibilité de relier les transactions à un usager particulier (caractérisé ou non);
- l'accès à des données continues couvrant de plus larges périodes de temps;
- la connaissance d'une grande proportion de sa clientèle

Les données de cartes à puces présentent en effet un intérêt particulier pour le suivi longitudinal de la demande de transport en commun. Bien que les données provenant des enquêtes Origine-Destination et des enquêtes à bord fournissent beaucoup plus de détails concernant la socio-démographie des usagers et la mobilité générale des résidents d'un territoire, elles ne permettent pas d'appréhender la variabilité dans les comportements de mobilité. Les analyses se limitent encore souvent à la constitution d'un jour moyen de déplacement. Les données de cartes à puces offrent toutefois l'opportunité d'observer la saisonnalité des comportements ainsi que la fluctuation, pour différentes périodes, des choix de mobilité de différents usagers. Elles permettent aussi d'envisager le croisement de différentes variables contextuelles (météo, événements, congé scolaire) afin de comprendre leurs effets sur la modification ponctuelle des comportements des usagers.

3. Ensemble de données expérimentales

Les expérimentations analytiques présentées s'appuient sur des données provenant du système de perception par cartes à puces de la Société de transport de l'Outaouais (STO). La STO est, à ce jour, le leader québécois en matière de cartes à puces et contribue à une meilleure compréhension des incidences de ces systèmes en permettant l'usage des données en découlant pour des fins de recherche.

La STO est une société de transport en commun de taille moyenne qui opère environ 200 autobus et qui dessert une population d'environ 240 000 personnes. Le système de cartes à puces y a été implanté en 2001. Actuellement, plus de 80% des usagers de la STO possèdent une carte à puces (puisque environ 15% des usagers proviennent du réseau OC Transpo d'Ottawa et ne possèdent pas de carte à puce; presque tous les usagers québécois de la STO possède une telle carte). Le système implanté à la STO dispose aussi de fonctionnalités de positionnement géographique (système GPS). Ainsi, à chaque transaction d'embarquement, la localisation spatiale de l'arrêt ainsi que le numéro de la ligne empruntée sont enregistrés dans la base de données. L'heure de la transaction est aussi enregistrée. Puisque la STO applique les règles usuelles de respect de la confidentialité des individus, les données de cartes à puces sont totalement anonymes. Aucune information nominale n'est disponible dans cette base de données. Notons que la carte de la STO est individualisée et comporte la photo de l'utilisateur, ce qui assure que la carte ne soit utilisée que par un seul usager.

La Figure 1 présente le diagramme fonctionnel du paiement par cartes à puce à la STO. L'utilisateur obtient sa carte d'un point de vente "émetteur" qui possède l'équipement permettant d'identifier et d'initialiser la carte. Il peut par la suite se rendre dans un point de vente de rechargement afin de faire le paiement du titre mensuel. Le titre présent sur la carte est alors validé jusqu'à une date déterminée. Ces points de vente sont en relation avec le système comptable de la STO pour assurer l'intégrité des transactions financières. Quand l'utilisateur monte à bord d'un autobus, le lecteur valide son titre en fonction du type de titre acheté, de la date et du circuit d'embarquement (titres "express", "interzone", etc.).

Pour assurer la validation, chaque autobus dispose d'un micro-ordinateur sur lequel sont chargées les données sur les tracés et voyages. Ces données proviennent du système opérationnel de la STO, en passant par le serveur central. De façon quotidienne, toutes les montées sont enregistrées par l'ordinateur de bord. Ces données sont téléchargées par infrarouge vers le serveur lorsque le véhicule retourne au garage.

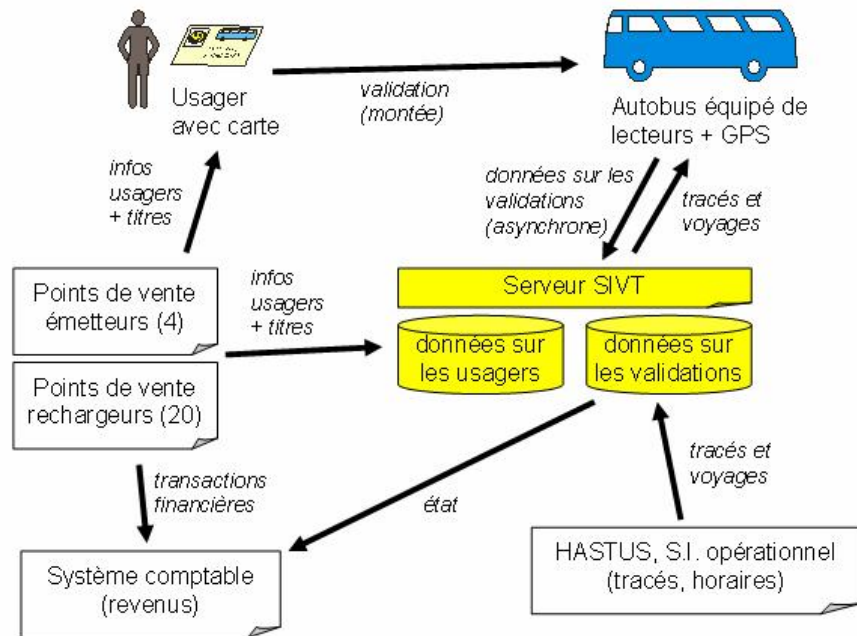


Figure 1. Diagramme fonctionnel du paiement par cartes à puce à la STO (Trépanier, Chapleau et Tranchant, 2005)

La Figure 2 présente le modèle simplifié de la base de données qui enregistre les données d'embarquement.

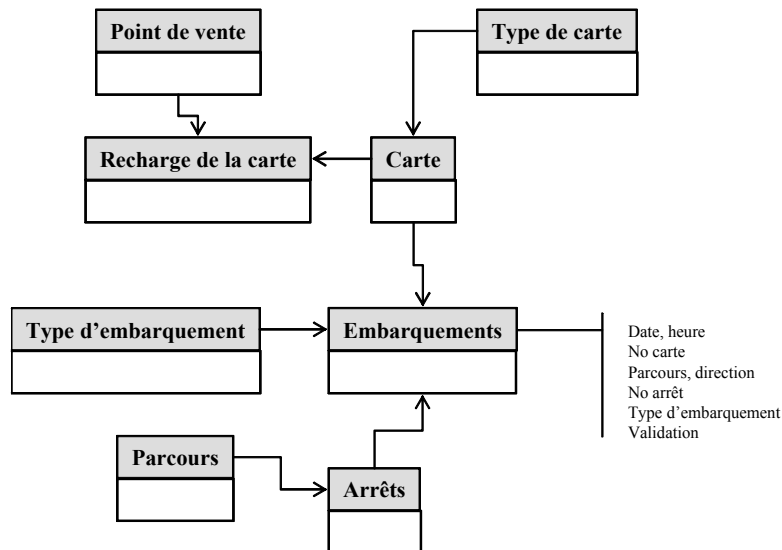


Figure 2. Modèle simplifié de la base de données qui enregistre les transactions d'accès au réseau de transport en commun (Agard, Morency et Trépanier, 2006)

La table principale est celle des embarquements qui contient un enregistrement par validation de carte sur le réseau. Les refus (invalidations de carte) sont aussi enregistrés dans une table et ignorés lors de l'analyse des comportements des usagers. Un lien est fait entre les embarquements et la configuration spatiale du réseau de transport en commun (parcours et arrêts). Ce lien assure la cohérence spatio-temporelle des transactions (heure d'embarquement vs localisation spatiale vs arrêt). Finalement, les informations sur la recharge des cartes sont aussi disponibles.

Ces différents détails permettent de cerner les trous informationnels qui devront être comblés par des procédures intelligentes afin de construire des ensembles de données plus complets en vue d'une

observation continue des comportements des usagers. L'information la plus critique à dériver est sans doute le point de débarquement d'un usager (voir à cet effet Tranchant, 2005). En effet, seules les transactions d'embarquement sur le réseau de transport en commun sont enregistrées : première montée, montée suivant une correspondance. L'étude spatio-temporelle des mouvements d'un usager sur le réseau de transport en commun permet d'identifier d'une part les segments de transport en commun qui doivent être assemblés afin de construire des déplacements (combinaison des embarquements successifs dus aux correspondances) et d'autre part d'identifier un lieu plausible de débarquement étant donnée la structure spatio-temporelle des activités quotidiennes.

4. Expérimentations analytiques

Cet article présente des expérimentations analytiques conduites à partir de différents échantillons de données. Les propriétés de ceux-ci seront spécifiées, au besoin.

4.1 Groupes de comportements homogènes et type d'abonnement

Des techniques d'exploration de données (data mining) ont été expérimentées à l'aide d'un ensemble important de données de cartes à puces (voir Agard, Morency et Trépanier, 2006). L'objectif de cette expérimentation est d'examiner la similarité des comportements des usagers selon le jour et la semaine de déplacement. Les techniques d'exploration de données utilisées sont :

- Le filtrage des données
- La création d'agrégats à l'aide de la méthode des k-moyennes et hiérarchique ascendante
- La caractérisation de groupes

La base de données utilisée pour ces expérimentations rassemble plus de 2 millions d'embarquements faits par 25 452 détenteurs de cartes à puces entre le 10 janvier et le 1^{er} avril 2005. Les transactions ont été groupées en semaines d'observations soit 238 895 usagers-semaines. Chaque enregistrement est caractérisé par 20 variables dichotomiques représentant les cinq jours de la semaine * 4 périodes de la journée (AM : 5h30-8h59, MI : 9h00-15h29, PM : 15h30-17h59 et SO : 18h00 et plus). Le Tableau 1 présente un exemple d'enregistrements de la base de données constituée pour expérimenter les techniques d'exploration de données. Par exemple, les champs J2_AM à J6_SO représentent les 20 périodes temporelles couplant le jour et l'heure de déplacement (2=lundi, ..., 6=vendredi). Ainsi, la carte #12244 est de catégorie "T1" ("adulte"). Cette carte a été validée (i.e. que l'utilisateur a effectué au moins 1 embarquement) le lundi matin (J2_AM) et le lundi après-midi (J2_PM) lors des semaines "S1", "S2" et ainsi de suite. Par ailleurs, on peut supposer que le détenteur de la carte #23231 a fait au moins 1 déplacement à chacune des périodes du lundi de la première semaine.

Tableau 1. Exemple d'enregistrement de la base de données constituée pour l'analyse exploratoire

ID_CARTE	TYPE_CARTE	group	semaine	J2_AM	J2_MI	J2_PM	J2_SO	...	J6_SO
12244	T1	adulte	S1	1	0	1	0
12244	T1	adulte	S2	1	0	1	0
1234312	T2	adulte	S1	1	0	0	1
23231	T34	étudiant	S1	1	1	1	1
...

La première analyse s'appuie sur l'ensemble de la base de données et s'applique à segmenter les enregistrements en agrégats homogènes sur la base des patrons de déplacement observés hebdomadairement. L'objectif est de révéler des groupes naturels d'utilisateurs-semaines. Le nombre d'agrégats n'est pas fixé a priori; la méthode hiérarchique ascendante est utilisée.

La seconde analyse s'intéresse à la composition des agrégats naturels en regard du type de carte (adulte, étudiant, personne âgée). Ceci est fait afin de vérifier si des comportements homogènes sont surtout dus au type d'utilisateur.

La troisième analyse porte sur la variabilité de l'appartenance aux agrégats de comportements homogènes sur les 12 semaines d'observation afin d'observer si un usager est toujours lié au même groupe (stabilité des comportements hebdomadaires). Ceci permet aussi d'identifier les semaines atypiques de déplacements.

4.1.1 Comportement général des usagers

Le traitement de la base de données globale permet de dériver quatre agrégats d'usagers-semaines présentant des comportements hebdomadaires similaires. Deux de ces agrégats (groupe 1 et groupe 2) présentent des patrons facilement interprétables :

- Figure 3: Le premier agrégat (groupe 1), qui rassemble 45.6% des usagers-semaines, est clairement associé aux usagers qui effectuent des déplacements réguliers vers et en partance d'activités contraintes telles que le travail puisqu'ils se déplacent principalement pendant les deux périodes de pointe. En moyenne, 79.4% de ces usagers se déplacent pendant la pointe du matin et 71.0% pendant la pointe du soir lors des jours de semaine. La proportion des usagers se déplaçant pendant les autres périodes est relativement faible autour de 6.4% pendant la journée et 2.6% pendant la soirée.
- Figure 4: Le troisième agrégat (groupe 3), qui représente 14.3% des usagers-semaines, est associé aux usagers qui ont des activités régulières pendant la première partie de la journée. En moyenne, 77.6% des usagers se déplacent pendant la pointe du matin et 74.8% pendant la journée.

Les deux autres agrégats (groupe 2 et groupe 4), qui rassemblent respectivement 14.8% et 25.2% des usagers-semaines, semblent regrouper des usagers avec des rythmes similaires d'immobilité. En fait, aucun patron clair d'activités ne peut être observé pour les usagers du groupe 2 alors que le groupe 4 rassemble des usagers peu mobiles sur le réseau (lors des douze semaines d'observation). Au plus, 36% des usagers-semaines de ce groupe accèdent au réseau de transport en commun pendant la journée.

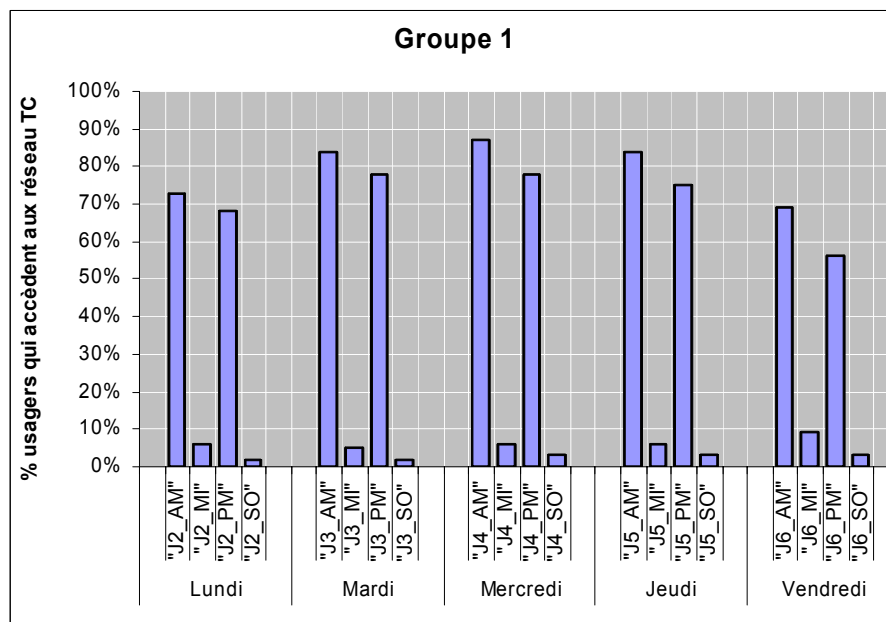


Figure 3. Structure d'accès au réseau TC (embarquements) - Groupe 1

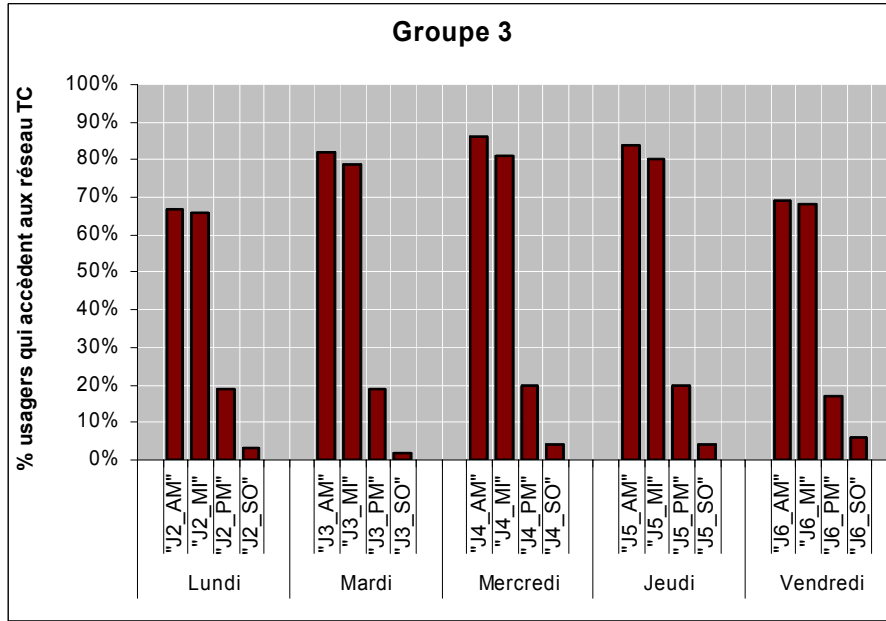


Figure 4. Structure d'accès au réseau TC (embarquements) - Groupe 3

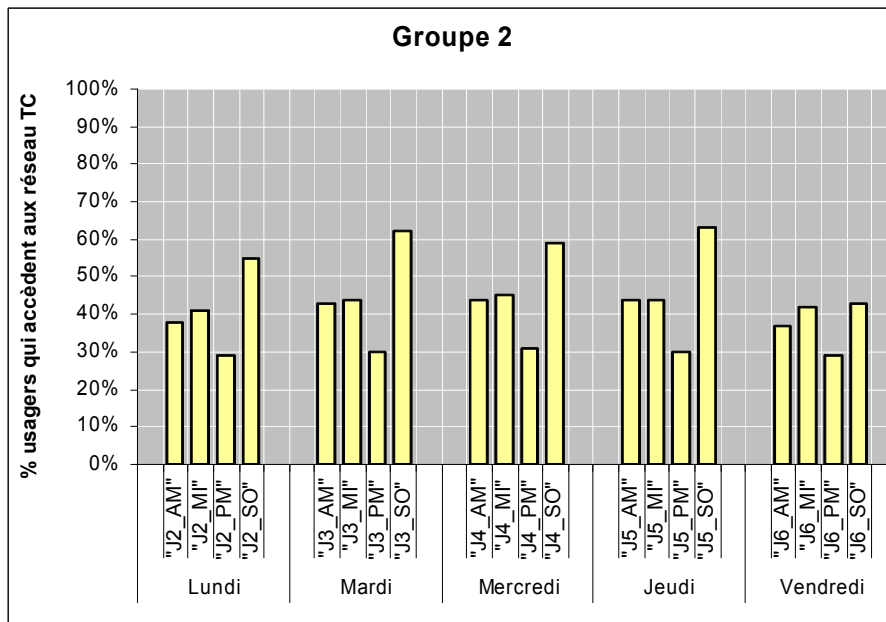


Figure 5. Structure d'accès au réseau TC (embarquements) – Groupe 2

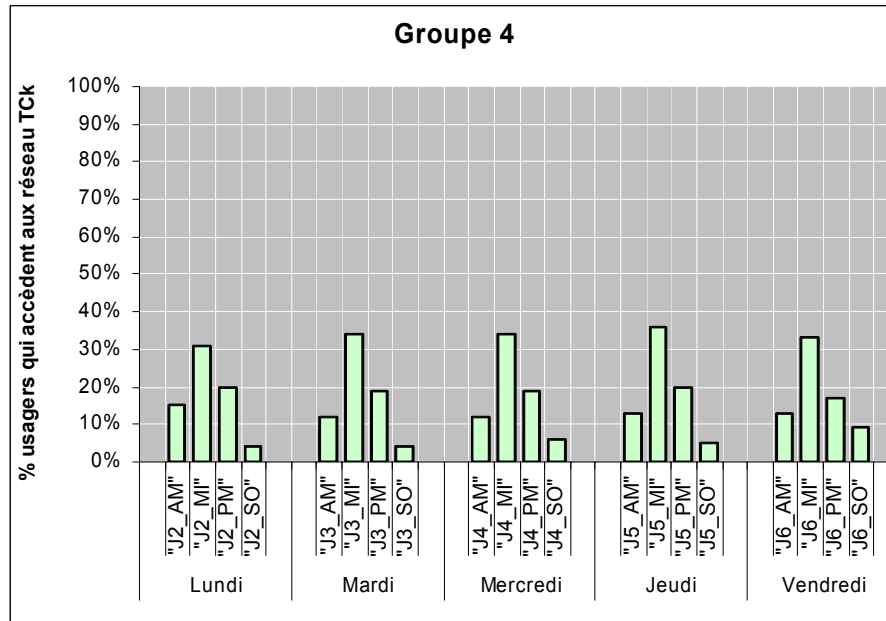


Figure 6. Structure d'accès au réseau TC (embarquements) - Groupe 4

4.1.2 Appartenance des usagers aux groupes d'usagers-semaines

Les tableaux suivants résument la composition des quatre agrégats créés par la méthode d'exploration de données.

D'une part, le Tableau 2 présente la distribution des usagers-semaines dans les quatre groupes selon le type de carte du voyageur. Par exemple, ces données montrent que près de 80% des semaines de déplacements des personnes âgées appartiennent au groupe 4. Ceci n'a rien de surprenant puisque ce groupe a le taux de mobilité le plus faible. Ces données confirment par ailleurs que les mouvements symétriques observés pour les usagers-semaines du groupe 1 sont en partie expliqués par le niveau de contrainte des activités des adultes; ainsi, près de 60% des semaines de déplacements des détenteurs de cartes de type adulte appartiennent à ce groupe. Ces données présentent finalement que les semaines de déplacement des étudiants sont distribuées dans les quatre groupes. Ceci est assez surprenant et pourrait s'expliquer par le fait que certains étudiants reçoivent automatiquement leur laissez-passer sans nécessairement en avoir besoin pour effectuer leurs principaux déplacements.

Tableau 2. Distribution des usagers-semaines en quatre groupes selon le type de carte

Type de carte	Gr1	Gr2	Gr3	Gr4	TOT
Adulte	58.8%	13.9%	9.2%	18.1%	100%
Étudiant	21.0%	17.7%	26.4%	34.8%	100%
Personne âgée	6.2%	6.4%	7.9%	79.5%	100%

D'autre part, le Tableau 3 présente la composition des quatre groupes naturels (agrégats) en termes de type de carte (des voyageurs pour lesquels les semaines de déplacements sont étudiées). Il appert que les détenteurs de cartes de type adulte représentent 85% des usagers-semaines du groupe 1. Ceci explique aussi la régularité des patrons de déplacements observée pour ce groupe. Ces données montrent aussi que les étudiants sont les principaux représentants du groupe 3 qui montre aussi des patrons réguliers d'activités.

Tableau 3. Composition des quatre groupes selon le type de carte

Type de carte	Adulte	Étudiant	Personne âgée	Total
Gr1	85.6%	13.9%	0.5%	100%
Gr2	62.4%	36.1%	1.4%	100%
Gr3	42.7%	55.4%	1.8%	100%
Gr4	47.7%	41.7%	10.6%	100%

Ces tableaux indiquent que certains types de carte sont liés à des comportements réguliers par exemple les personnes âgées (environ 80% de ces usagers appartiennent au groupe 4) mais que d'autres types rassemblent des usagers avec des comportements variables sur les 12 semaines d'observation.

4.1.3 Variabilité pendant les 12 semaines d'observation

L'appartenance des usagers-semaines des détenteurs de cartes étudiantes aux différents groupes pendant les 12 semaines d'observations est présentée à la Figure 7. On y observe rapidement une semaine atypique de déplacement. En effet, la distribution atypique observée à la semaine 10 est due au congé scolaire, période pendant laquelle les étudiants adoptent d'autres rythmes d'activités.

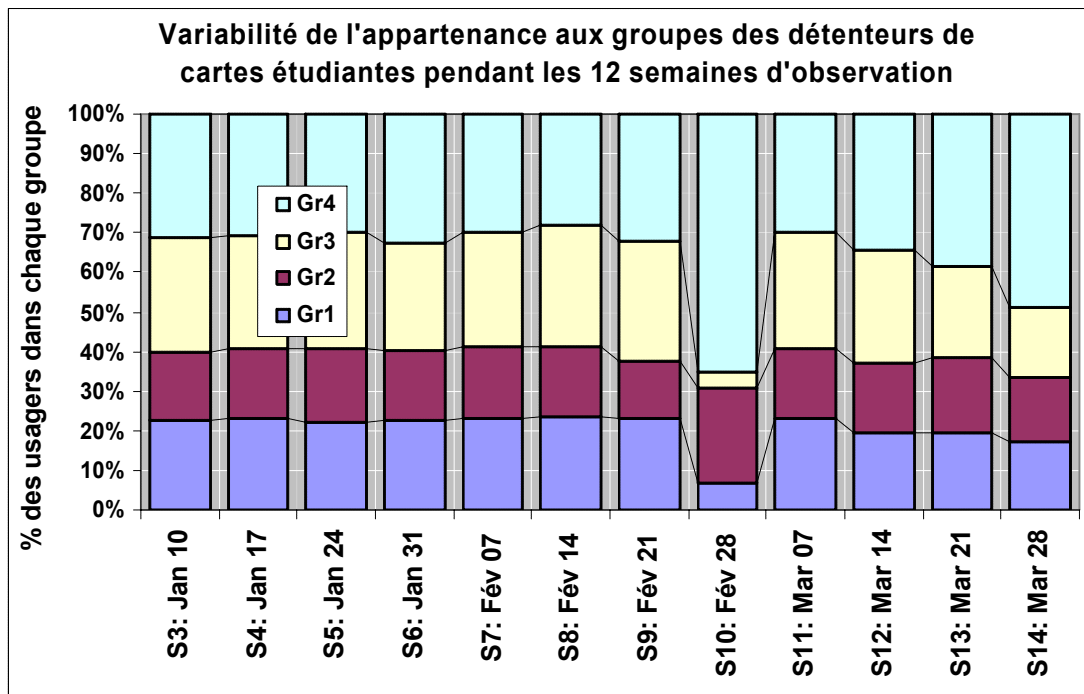


Figure 7. Variabilité de l'appartenance aux groupes des détenteurs de cartes étudiantes pendant les 12 semaines d'observation

Dans la Figure 8, les portions en noir des diagrammes à secteurs indiquent que les usagers appartiennent au même groupe pour toutes les semaines d'observation. La taille du diagramme est proportionnelle au nombre d'usagers. On observe que la majorité des adultes du groupe 1 ont des comportements réguliers. C'est moins le cas des autres groupes. Les étudiants ont des comportements moins réguliers. Ceci pourrait être dû aux horaires irréguliers des étudiants (collège et université) ainsi qu'au fait que les étudiants sont parfois reconduits par leurs parents, n'utilisant donc pas le transport en commun pour ces déplacements. Finalement, les personnes âgées sont concentrées dans le groupe 4, ce groupe étant caractérisé par un niveau très bas de mobilité.

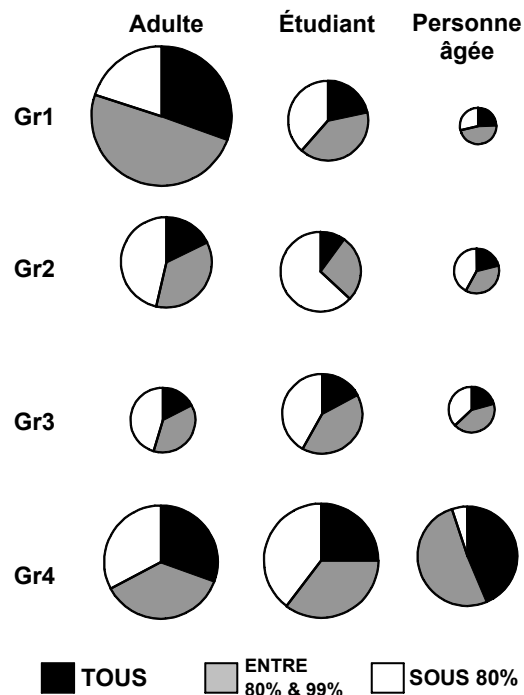


Figure 8. Proportion des usagers selon leur appartenance aux groupes pendant 11 semaines d'observation (semaine # 10 retirée)

4.2 Portrait longitudinal et régularité des comportements quotidiens

D'un point de vue analytique, l'intérêt des données de cartes à puces provient surtout de la répétitivité des observations pour une même unité d'échantillonnage en l'occurrence les usagers du transport en commun. Afin d'explorer les potentialités d'analyse en ce qui concerne la variabilité des comportements, un autre échantillon de données est utilisé. Celui-ci porte sur un maximum de 277 jours consécutifs d'observation soit du 1 janvier au 4 octobre 2005. La base de données résultante compte plus de 6 millions de montées effectuées par 43 248 usagers.

4.2.1 Étude du taux d'utilisation du réseau de transport en commun

Un aperçu plus détaillé de la base de données disponible est fourni par l'étude combinée des taux d'activité sur le réseau ainsi que du nombre de jours d'observation pour les différents usagers. La Figure 9 présente ces deux composantes.

On observe notamment la distribution des usagers selon le taux d'activité sur le réseau à savoir le pourcentage des jours d'observation pour lesquels il y a eu déplacement sur le réseau de transport en commun. En moyenne, un usager se déplace lors de 58% des jours compris dans la plage temporelle d'observation. Environ 14% des usagers présents dans la base de données expérimentale se sont déplacés lors de 60 à 65% des jours d'observation. En ce qui concerne les 7% d'usagers s'étant déplacés dans 95-100% des cas, ils sont dû, à 90%, à des usagers ayant été observés pour 10 jours ou moins.

Les détails concernant les plages d'observation pour les principaux types d'usagers (cumulant environ 80% de tous les détenteurs de cartes à puces observés) sont présentés au Tableau 4. Le fait le plus intéressant est sans doute le nombre moins important de jours d'observation pour les usagers de type étudiant. En effet, il est plausible que ces usagers n'utilisent plus le réseau de transport lorsque leurs activités scolaires sont terminées (période estivale). La proportion des jours observés pour lesquels il y a eu déplacement est comparable pour les différents types d'usagers.

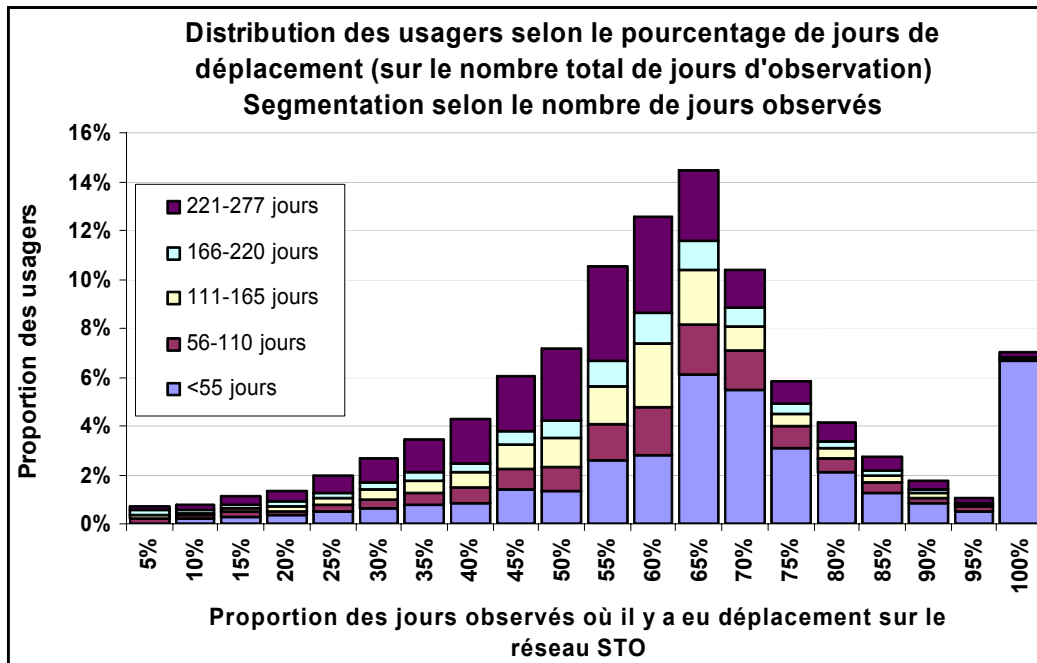


Figure 9. Distribution des détenteurs de cartes à puce selon le nombre de jours d'utilisation du réseau de transport en commun (sur un maximum de 277 jours d'observation) et statistiques sur le nombre moyen de jours de déplacements pour les principaux types d'abonnement

Tableau 4. Nombre moyen de jours d'observation pour les usagers détenteurs des principaux types d'abonnement et proportion moyenne de jours de déplacement

Type	NbMoyJrsObs	%JrsDépl
Adulte Régulier	146	54.6%
Étudiant Régulier	85	50.6%
Adulte Express	161	50.8%
Scolaire	74	54.3%
Adulte fidélité	187	54.7%

4.2.2 Nombre de montées par jour

La Figure 10 présente le nombre moyen de montées par type de jour pour les mêmes classes d'usagers. On observe que ce sont les étudiants réguliers qui font, en moyenne, le plus de montées par jour lors des jours de semaine. Par ailleurs, on observe très clairement une diminution du nombre de montées pour les titres scolaires lors des jours de fin de semaine (étudiant régulier et scolaire). Il semble de plus exister une certaine constance des comportements, en termes de montées, pour les cinq jours de la semaine et ce pour tous les types d'abonnement.

L'étude de la variabilité du nombre de montées, sur les 277 jours d'observation, révèle que ce sont les détenteurs d'un titre d'adulte express qui présentent le plus de régularité lors des jours de semaine (coefficient de variation de moins de 30%). De façon générale, la variabilité du nombre de montées est supérieure lors des jours de fin de semaine (entre 50 et 60% pour les samedis et dimanches) sauf pour les scolaires.

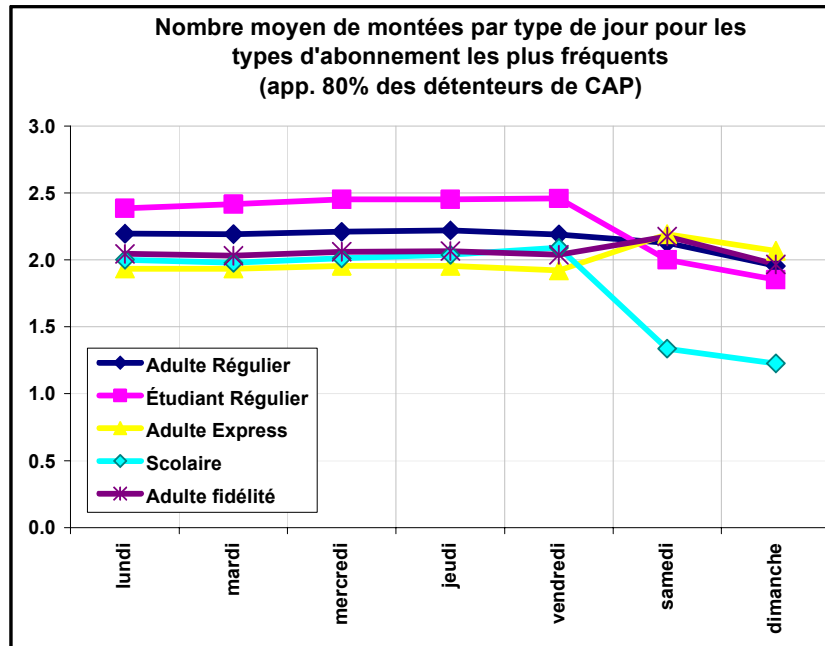


Figure 10. Nombre moyen de montées par type de jour pour les sept types d'abonnement les plus fréquents (qui regroupent près de 80% des usagers observés)

4.2.3 Chronologie de l'utilisation des arrêts

Une appréciation de la régularité d'utilisation des arrêts du réseau peut aussi être construite pour les usagers observés. Le propos consiste à identifier le processus d'acquisition d'arrêts par les usagers ou, en d'autres mots, le rythme temporel d'utilisation d'arrêts différents par les usagers.

La Figure 11 présente le nombre d'arrêts différents utilisés, en montée, par les usagers selon le nombre de jours de déplacement observés sur le réseau. Le taux moyen de renouvellement des arrêts est estimé à 0.25 nouvel arrêt par jour de déplacement sur le réseau. La distribution présentée ci-dessous permet par ailleurs de déceler certains effets de saisonnalité. Sachant que la période d'observation s'amorce en janvier, on peut très bien voir que le taux d'utilisation de nouveaux arrêts se met à croître de façon plus importante suite à 170 jours de déplacement ce qui correspond grosso modo à la période estivale. La variabilité augmente par ailleurs avec le nombre de jours de déplacements observés ce qui pourrait correspondre à l'acquisition de nouveaux comportements par les usagers sur une plage plus longue d'observation.

Ce type de mesure révèle que les comportements des usagers sont assez variables et qu'ils évoluent avec le temps. Les données portant sur seule journée d'observation n'ont pas la capacité de documenter les comportements changeant des usagers. Le portrait construit grâce à des observations répétées sur un même échantillon nuance donc la base usuelle de modélisation qu'est le jour moyen.

À titre comparatif Schoenfelder et Axhausen (2004) ont examiné le nombre de lieux visités sur une longue période d'observation à l'aide de données provenant de 6 bases de données différentes, notamment les données de l'enquête Mobidrive (carnet de déplacements de 6 semaines). En moyenne, ils estiment à environ 0.30 nouveau lieu par jour le taux de renouvellement des lieux d'activités.

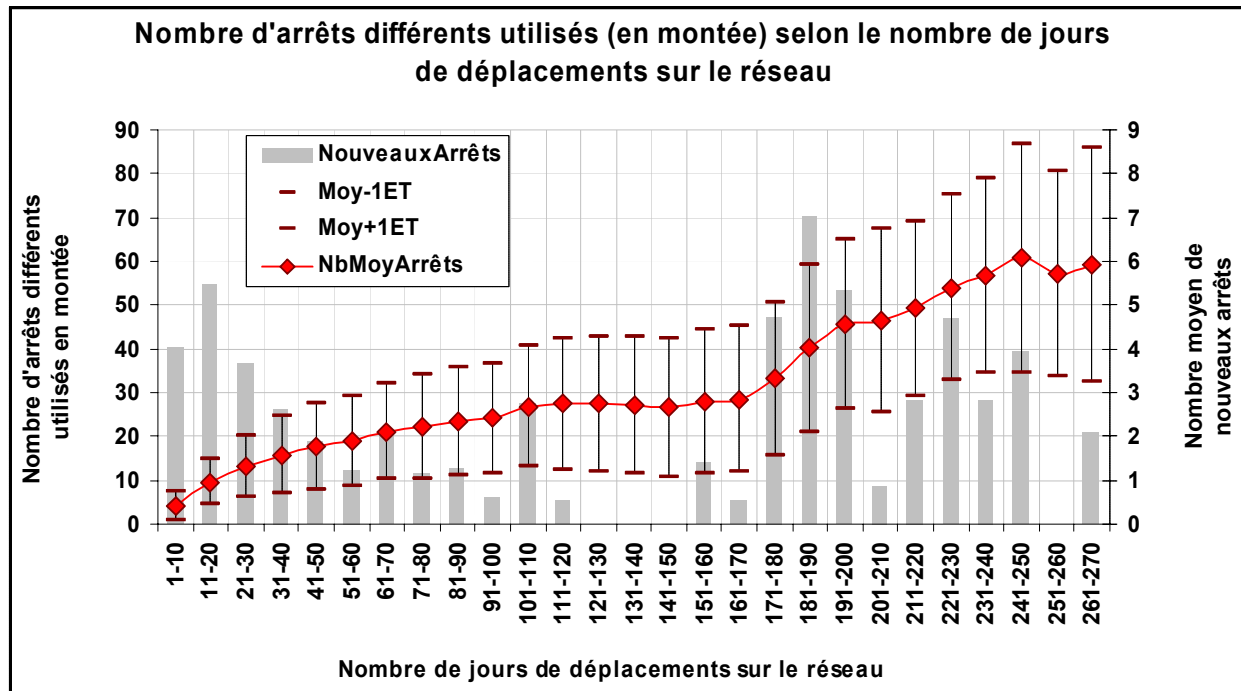


Figure 11. Chronologie d'utilisation d'arrêts différents du réseau de transport en commun sur une plage continue d'observation

4.3 Variabilité spatio-temporelle de déplacements observés du réseau sur une période de 10 mois

L'étude de la variabilité spatio-temporelle des comportements peut aussi s'effectuer au niveau individuel. En effet, la répétitivité des observations, pour une période de 9 mois, fournit des détails très intéressants sur la régularité d'utilisation du réseau de transport en commun.

En guise de démonstration, les transactions associées à deux cartes réseau de transport en commun, un titre aîné et un titre régulier, ont été examinés afin d'observer la variabilité des comportements sous différents aspects. Rappelons qu'en aucun temps nous n'associons ces données sur des cartes à des personnes individuelles.

4.3.1 Variabilité du nombre de montées

Le titre aîné a été associé à 1372 montées sur une période de 277 jours consécutifs d'observation, soit presque 5 montées par jour en moyenne. En fait, le titre a été utilisé le réseau lors de tous les jours d'observation.

Le second cas examiné est un titre régulier pour adulte. Le titre a été associé à 478 montées sur une période de 277 jours consécutifs d'observation. Le titre a été utilisé 274 jours sur 277.

La Figure 12 présente le nombre de montées effectuées à l'aide du titre aîné lors des 277 jours d'observation. Cette représentation permet de déceler quelques jours de comportements particuliers, souvent des vendredis ou des samedis. Le titre régulier reflète un comportement beaucoup plus stable dans le temps en termes de nombre de montées par jour (voir Figure 12).

La variabilité du nombre de montées par type de jour est résumée à la Figure 14 pour le titre aîné et à la Figure 15 pour le titre régulier pour adulte. On remarque que pour le titre aîné, il y a davantage de montées les samedis alors que les dimanches présentent à la fois le nombre moyen de montées le plus faible et la plus grande variabilité en termes de montées par jour. Dans le cas du titre régulier, on observe que tous les samedis observés ont été identiques avec une plage de montée. Du mardi au vendredi, il y a environ 2 montées par jour alors que les lundis sont différents du reste de la semaine. La variabilité, appréciée à l'aide du coefficient de variation, est légèrement supérieure à celle estimée pour le titre aîné.

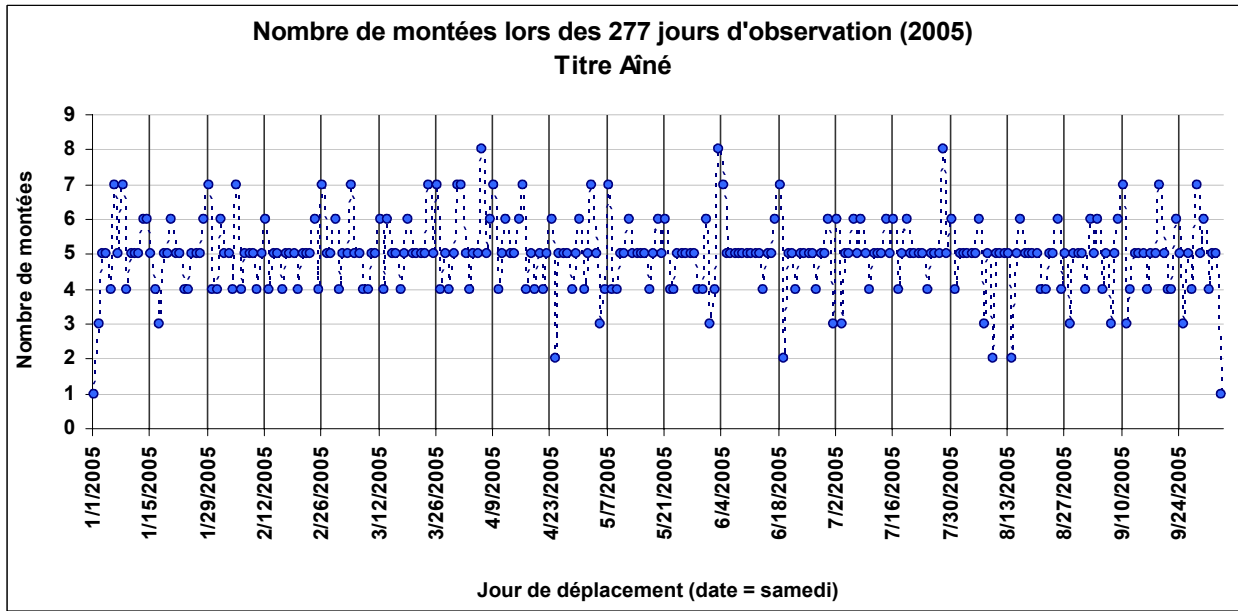


Figure 12. Nombre de montées observées pour un titre aîné pour 277 jours consécutifs d'observation (2005)

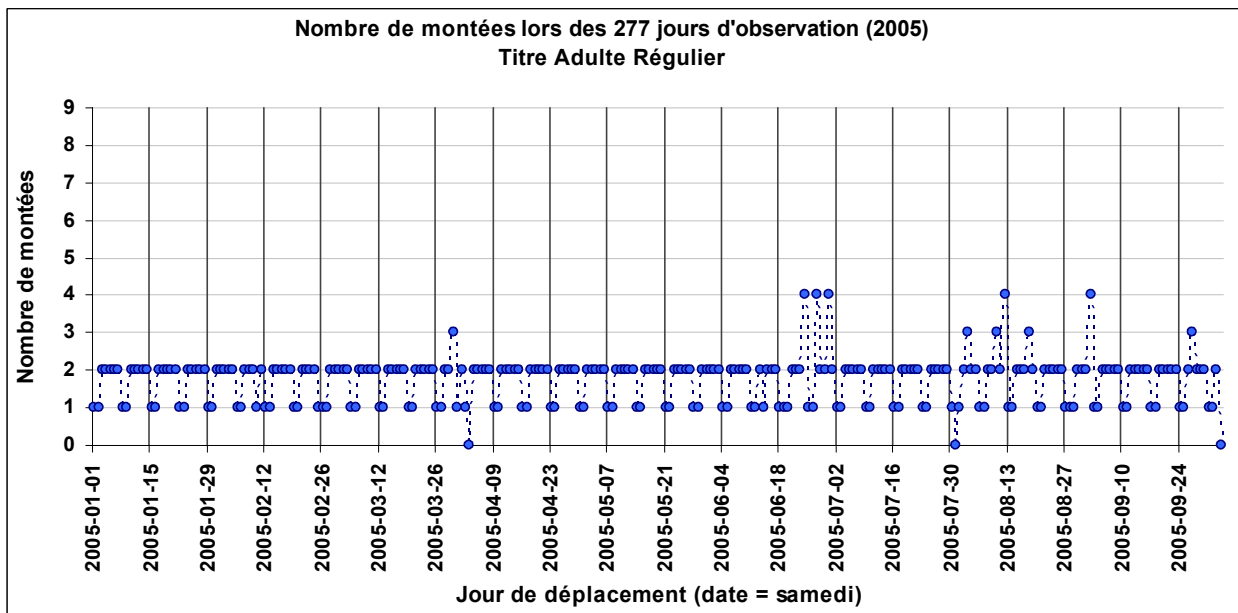


Figure 13. Nombre de montées observées pour un titre adulte régulier pour 277 jours consécutifs d'observation (2005)

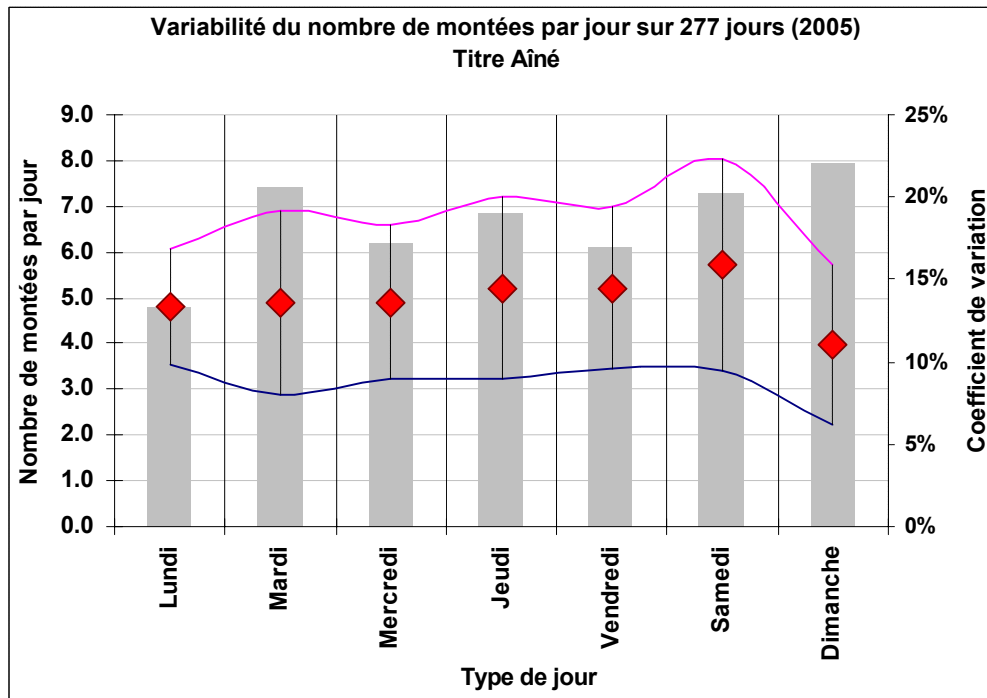


Figure 14. Variabilité du nombre de montées par jour de déplacement pour 277 jours (2005) consécutifs d'observation (moyenne, intervalle de confiance à 95%, coefficient de variation) - titre aîné

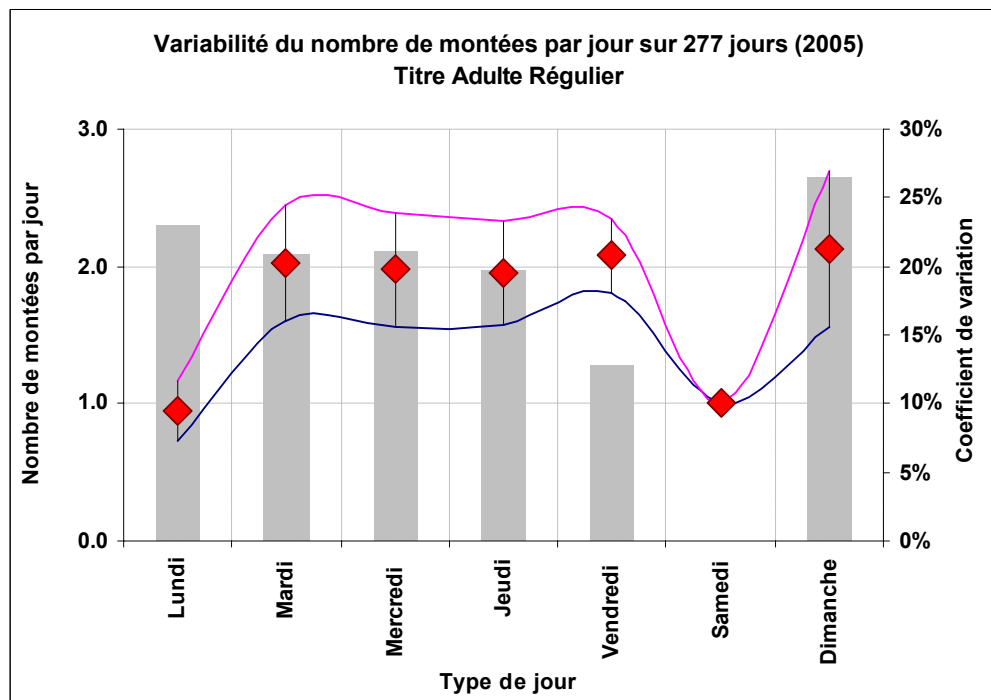


Figure 15. Variabilité du nombre de montées par jour de déplacement pour 277 jours (2005) consécutifs d'observation (moyenne, intervalle de confiance à 95%, coefficient de variation) - titre adulte régulier

4.3.2 Régularité temporelle des déplacements (heures de montée)

Des méthodes d'exploration de données ont à nouveau été utilisées afin de classer les jours d'observation selon les comportements temporels des usagers sur le réseau de transport en commun

(heures pendant lesquelles un usager effectue des montées). Dans cette perspective, un fichier segmentant la mobilité associée à chaque titre en heure a été produit et traité à l'aide de méthodes de filtrage de données et d'agrégation par k-moyennes. L'extrait d'un fichier de mobilité ainsi constitué est présenté au Tableau 5. Pour chaque jour d'observation, une valeur binaire permet d'indiquer si le titre a été associé à une montée lors de cette heure.

Tableau 5. Structure des fichiers produits pour procéder à la classification des jours d'observation selon les comportements temporels associés aux titres

NUMCARTE	TITRE	DATE	JRSEM	H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12	H13	H14
	8	2005-01-01	7	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	8	2005-01-02	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0
	8	2005-01-03	2	0	0	0	0	0	0	0	0	1	1	1	1	0	0
	8	2005-01-04	3	0	0	0	0	0	0	0	0	1	1	0	1	1	0
	8	2005-01-05	4	0	0	0	0	0	0	0	0	1	1	1	0	0	0
	8	2005-01-06	5	0	0	0	0	0	0	0	0	1	1	1	1	0	0
	8	2005-01-07	6	0	0	0	0	0	0	0	0	1	0	1	1	1	0

Le processus de classification a permis de regrouper les 277 jours d'observation du titre aîné en 29 agrégats différents. A première vue, cela témoigne d'une grande variabilité temporelle des comportements. La Figure 16 présente les croisements jour * agrégat pour les 15 agrégats les plus importants (qui rassemblent plus de 85% des jours d'observation). On observe la plus grande régularité les mercredis avec 62% des jours observés appartenant au même agrégat. Par ailleurs, on observe que l'agrégat 1 est le plus important pour les jours de semaine ainsi que le samedi (35.4% de tous les jours d'observation). Cet agrégat représente des jours où des montées s'effectuent aux heures suivantes : 8h-9h, 9h-10h, 10h-11h, 11h-10h et 15h-16h.

Le même processus a regroupé les jours d'observation du titre régulier en 15 agrégats homogènes. En outre, pour chaque type de jour, entre 73% et 95% appartiennent à un seul agrégat. Les comportements apparaissent donc définitivement plus réguliers en terme d'heures de déplacement. Les jours de fin de semaine sont particulièrement réguliers avec 93-95% des jours d'observations étant homogènes. En effet, tel qu'observable à la Figure 17, seulement trois samedis n'appartiennent pas à l'agrégat 12 alors que seulement deux dimanches n'appartiennent pas à l'agrégat 9. L'agrégat 12 correspond à une ou plusieurs montées entre 8h00 et 9h00 alors que l'agrégat 9 correspond à une ou plusieurs montées entre 9h00 et 10h00.

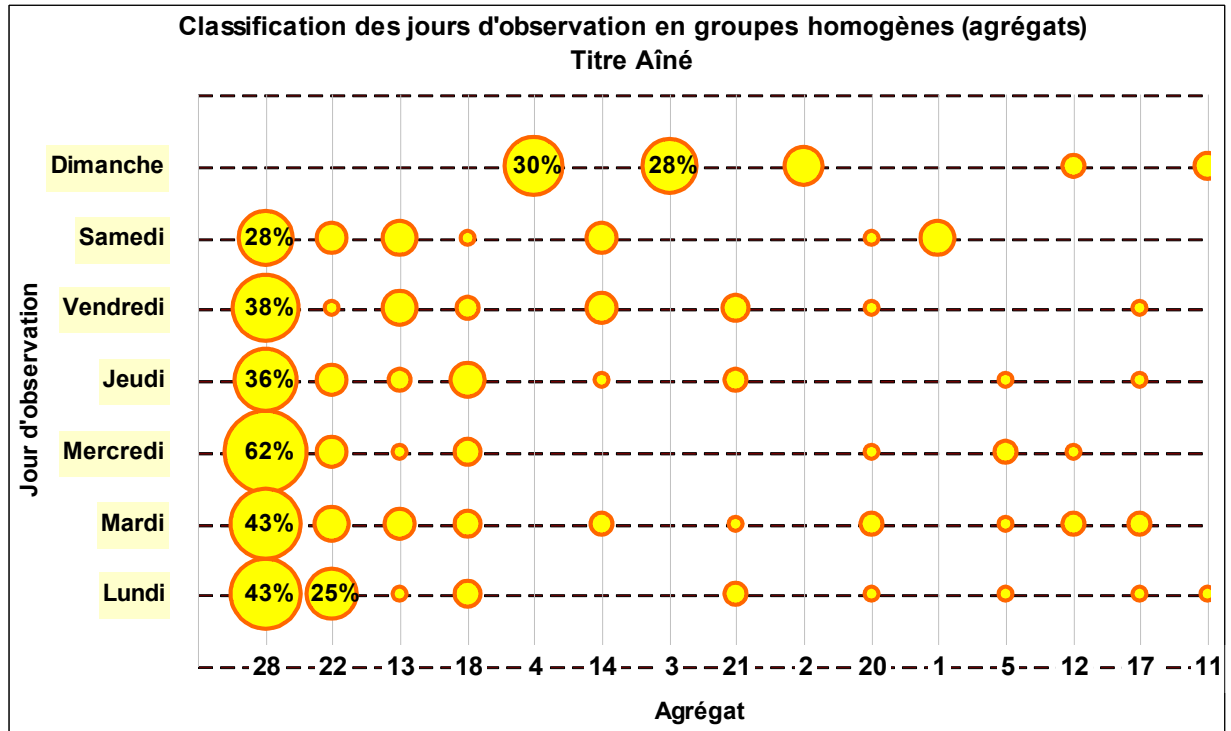


Figure 16. Classification des jours d'observation en groupes homogènes sur la base des heures de montée – titre aîné

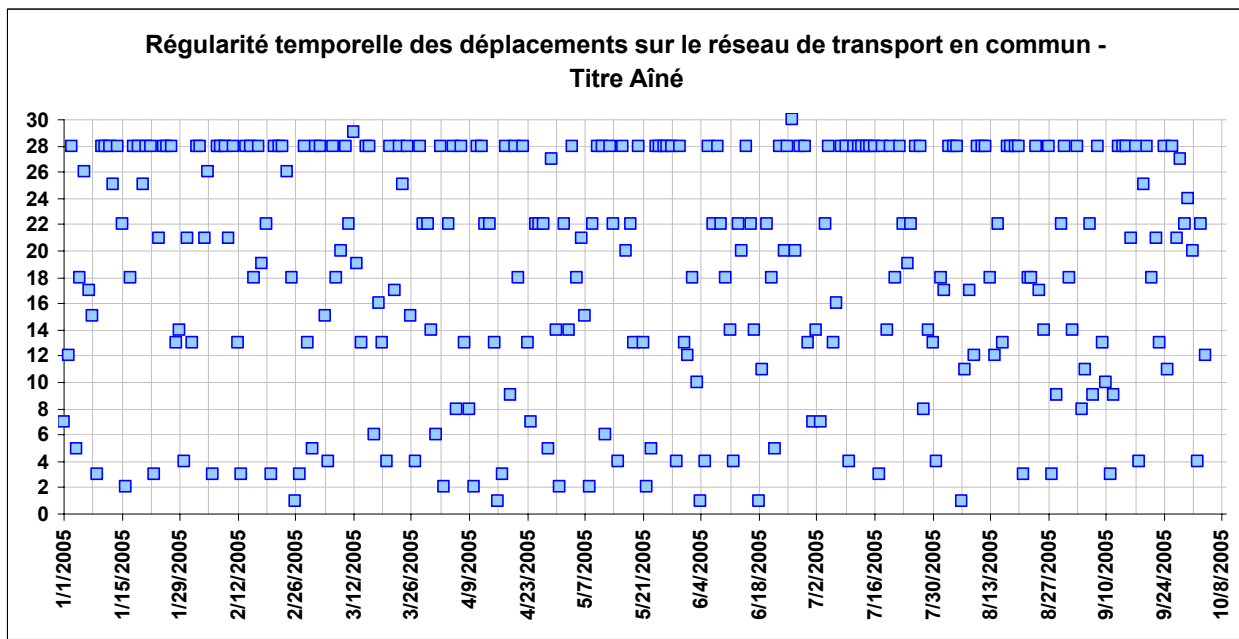


Figure 17. Appartenance des différents jours d'observation aux 29 agrégats formés – titre aîné

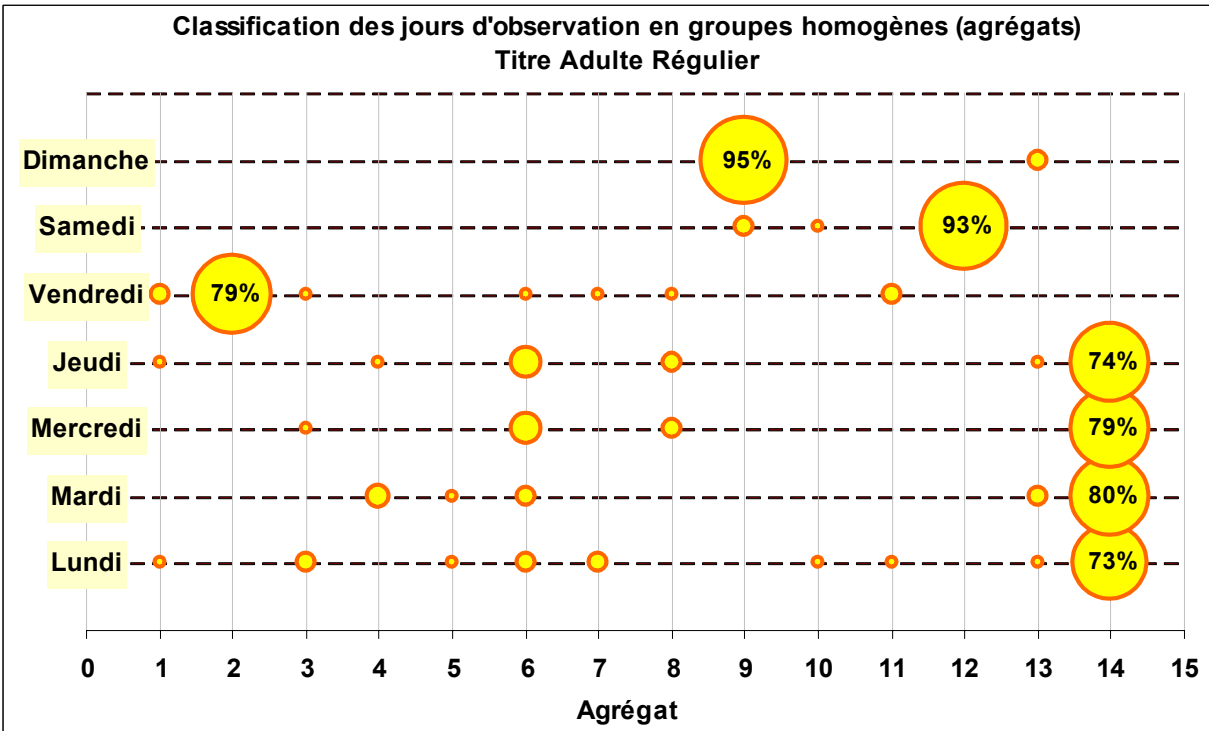


Figure 18. Classification des jours d'observation en groupes homogènes sur la base des heures de montée – titre régulier pour adulte

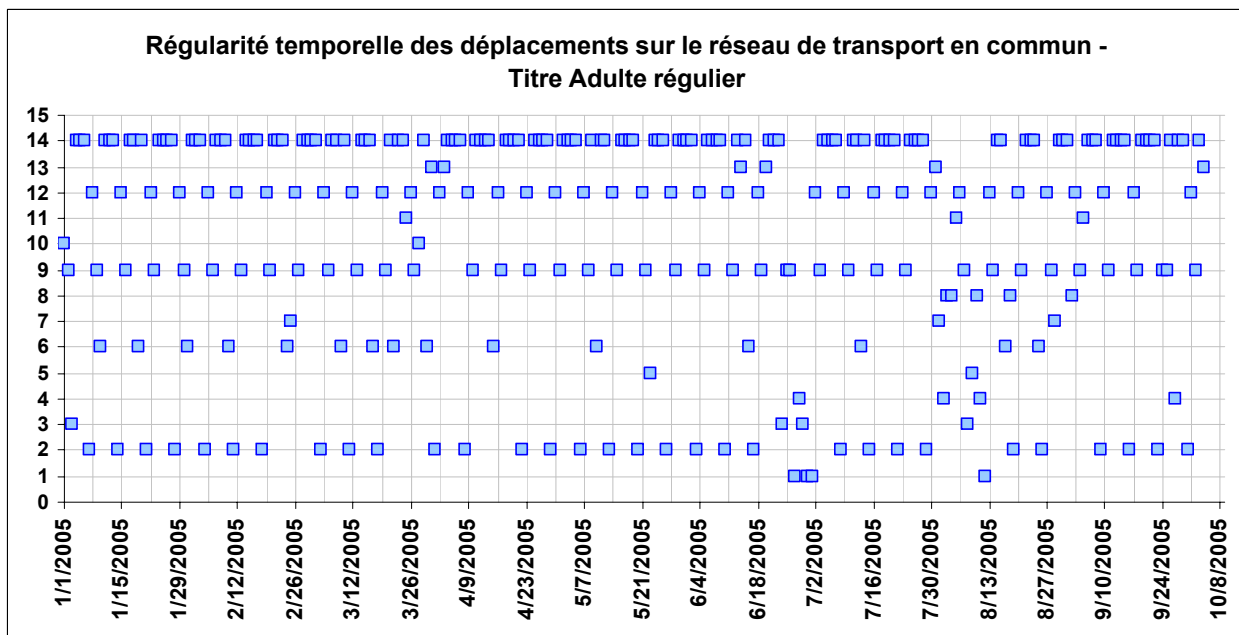


Figure 19. Appartenance des différents jours d'observation aux 15 agrégats formés – titre régulier pour adulte

5. Perspectives

Les données de cartes à puces ont la capacité de révéler différents aspects de régularité et de variabilité des comportements des usagers du service de transport en commun, mesures impossibles à effectuer avec les données classiques d'enquêtes Origine-Destination. Bien évidemment, elles ne peuvent se

substituer à ces enquêtes régionales puisqu'elles ne ciblent que les usagers d'un réseau particulier. Cependant, elles sont très riches en ce qui concerne le suivi longitudinal de l'utilisation de ce service et permettent surtout de nuancer les mesures classiques faites à l'aide d'un jour moyen de déplacement. En effet, certains comportements ont un cycle quotidien mais plusieurs ont un cycle qui s'étale sur plusieurs jours voire plusieurs semaines. La mesure de la régularité et du renouvellement de certaines activités requiert des données recueillies en continu; les données de cartes à puces sont une source intéressante à cet égard.

D'autres expérimentations pourront permettre de multiplier les bénéfices analytiques des données de cartes à puces. En effet, il est possible d'envisager d'autres mesures de variabilité telles que la dispersion spatiale des lieux utilisées par les usagers. En outre, la fusion de ces données avec d'autres sources telles que des données d'enquêtes à bord, des données de comptage aux arrêts ou des données d'enquêtes Origine-Destination régionale permettra de constituer un système d'information pertinent pour la caractérisation de la mobilité urbaine. La fusion des données de cartes à puces de la STO avec les données de la dernière enquête OD tenue en 2005 dans cette région, portant sur un échantillon d'environ 25 000 ménages, est une expérimentation qui permettrait de valider la faisabilité d'un tel processus.

6. Remerciements

Les auteurs remercient les représentants de la Société de transport de l'Outaouais, pour leur collaboration ainsi que leur ouverture à propos de l'utilisation des données provenant de l'utilisation de la carte à puces en transport collectif pour des fins de recherche.

Les auteurs remercient par ailleurs l'AMT (Agence métropolitaine de transport) qui soutient plusieurs activités de recherche du Groupe MADITUC, notamment l'étude des potentialités analytiques des données provenant des systèmes de perception par cartes à puces.

7. Références

AGARD, Bruno, MORENCY, Catherine, TRÉPANIÉ, Martin (2006). Mining public transport user behaviour from smart card data, conférence proposée à INCOM 2006, à Lyon, en mai 2006.

BAGCHI, M., WHITE, P.R. (2004), What role for smart-card data from bus system?, *Municipal Engineer* **157**, mars 2004, p.39-46

SCHOENFELDER, S., AXHAUSEN, K.W. (2004) Structure and innovation of human activity spaces, *Arbeitsberichte Verkehrsund Raumplanung*, 258, IVT, ETH Zürich, Zürich. (<http://www.ivt.ethz.ch/vpl/publications/reports/ab258.pdf>)

TRANCHANT, Nicolas (2005). Modèle de dérivation des déplacements en transport collectif à partir de données de cartes à puces. Mémoire M.Sc.A., École Polytechnique, Génie industriel, Montréal, 218 pages.

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

TRÉPANIÉ, Martin, CHAPLEAU, Robert, TRANCHANT, Nicolas (200X) Individual Trip Destination Estimation in Transit Smart Card Automated Fare Collection System, soumis au *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* en novembre 2005.