

# L'EXPLOITATION DES DONNEES DE CARTES A PUCE A DES FINS DE PLANIFICATION DES RESEAUX DE TRANSPORT COLLECTIF URBAINS

*Martin TRÉPANIÉ, ing., Ph.D., Professeur agrégé  
École Polytechnique de Montréal  
Département de mathématiques et de génie industriel*

## RESUME

Les systèmes de paiement par cartes à puce sont maintenant légion dans les réseaux de transport collectif urbains. En plus de jouer leur rôle premier, qui est le contrôle des revenus, ces systèmes collectent continuellement d'énormes quantités de données transactionnelles sur l'utilisation des services d'autobus, de train, de tramway et de métro. De par la nature spatio-temporelle des informations récoltées, la carte à puce constitue un instrument de planification des transports très riche, en autant que les données soient correctement traitées, le tout sous la plus grande confidentialité. Cette communication présente plus de cinq années de travaux de recherche effectuées en collaboration avec la Société de transport de l'Outaouais, au Québec.

## INTRODUCTION

La carte à puce, de par la versatilité et la puissance du médium, constitue une des plus intéressantes avancées technologiques dans les services de transport collectif. Intégrée à un système de perception électronique et automatique des titres de transport, la carte permet une collecte en continue de données transactionnelles individualisées sur l'utilisation des réseaux de transport collectif. Bien que l'objectif soit d'offrir une meilleure gestion des revenus et de diminuer la fraude, il est clair qu'un observateur avisé, planificateur ou non, sait que cet enregistrement localisé dans le temps et souvent dans l'espace, pourra servir à des fins de planification du réseau. En effet, la connaissance apportée par ce système dépasse même celui des systèmes de comptage automatiques, qui peuvent déterminer la charge à bord des véhicules mais ne permettent pas de déterminer quels types d'usagers s'y trouvent.

Cet article présente plusieurs travaux de recherche réalisés avec la Société de transport de l'Outaouais (STO), une société de transport de taille moyenne située dans l'Outaouais québécois. Équipée d'un système de carte à puce depuis 2001, la STO s'est montrée

proactive en désirant déterminer quelles sont les potentialités d'exploitation de ses données, qui sont de surcroît localisées dans le temps et dans l'espace grâce à des GPS embarqués. Après avoir résumé quelques rares travaux pertinents dans la littérature, l'article présente la méthodologie utilisée en toile de fond aux différents projets, l'approche orientée-objet en transport. On y décrit également le système de validation des titres de la STO, source des bases de données. Ensuite, les analyses sont articulées sous forme d'exemples et de concepts touchant différents domaines :

- le traitement des erreurs détectées dans les données brutes;
- la détermination d'un lieu de destination (descente) probable pour chacune des transactions (montées);
- la diffusion de statistiques d'achalandage au sein de la société de transport;
- la modélisation des comportements quotidiens et saisonniers de la clientèle;
- l'obtention de statistiques opérationnelles de première main sur le service;
- la caractérisation de l'utilisation du sol autour du réseau;
- la mesure de la loyauté des usagers.

Quelques perspectives de recherche viendront conclure cet article.

## **FONDEMENTS**

Les systèmes de paiement des services de transport collectif par cartes à puce permettent une meilleure gestion des revenus, en plus de faciliter l'intégration tarifaire et les opérations de perception à bord des véhicules. L'utilité de tels systèmes est reconnue dans la littérature, malgré certaines limitations souvent occasionnées par les coûts importants d'implantation : les opérateurs décident de retirer certaines fonctions (nombre de titres, paiement électronique) plutôt que de risquer de voir leur projet stoppé (Bonneau 2002). Cette section rapporte quelques travaux pertinents du domaine.

### **Genèse des données**

Bagchi et White (2004) recensent le type de données qui peuvent être récoltées par les cartes à puce qu'ils catégorisent de la façon suivante :

- Information spatiale : lieu d'embarquement, lieu de débarquement dans certains cas, durée de trajet (dans le cas où on connaît les lieux de montée et de descente).
- Information temporelle et structurale : date et heure d'embarquement, mode de transport et service emprunté.
- Information personnelle sur l'usager : nom, adresse, sexe, âge, possession automobile.
- Information d'achat : type de billet acheté, prix du billet, lieu d'achat.

Le lieu de débarquement n'est pas toujours enregistré. Par contre, en Corée, ces informations sont recueillies puisqu'ils ont introduit une tarification qui est basée sur la distance de parcours (Park & Kim, 2008). Conséquemment, un lecteur est positionné à la sortie des autobus ou du métro et les usagers doivent présenter au passage leur carte à puce afin de calculer le coût du trajet réalisé. De la même façon, au Royaume-Uni, une

estimation de la durée des trajets doit être réalisée pour les processus de remboursement (Bagchi & White, 2005). Ce processus vise à rembourser les usagers lorsque surviennent des retards dans le service. Par contre, les usagers ne sont pas tenus de signaler leur descente avec leur carte à puce. Cette information est donc obtenue par le biais de sondages (Bagchi & White, 2005). Il est aussi possible d'imputer un lieu plausible de débarquement. À cet égard, Trépanier et al. (2007) proposent un modèle afin d'estimer les lieux de débarquement en supposant que l'individu va réembarquer au prochain trajet à l'arrêt le plus près du lieu où il a débarqué. Un résumé plus détaillé de ce modèle sera ultérieurement exposé.

## **Exploitations**

Les avantages des systèmes de cartes à puce en transport collectif sont multiples (Bagchi et White 2005):

- un accès à un volume plus large de données sur les trajets individuels;
- la possibilité de relier les cartes aux usagers;
- l'accès à des données continues couvrant de plus larges périodes de temps;
- la connaissance d'une grande proportion de la clientèle.

Bien que les contributions des données issues des cartes à puce soient reconnues, il n'est pas toujours facile d'utiliser ces données (Chu et Chapleau, 2008). Des processus d'amélioration de la qualité des données sont souvent nécessaires afin d'éliminer les erreurs (Utsunomiya, Attanucci et Wilson, 2006). Dans cet ordre d'idée, Chu et al. (2009) proposent d'utiliser la logique spatio-temporelle ainsi que les concepts de transport en commun afin de détecter les valeurs suspectes dans les données et d'améliorer la qualité de ces dernières. Ce sujet sera également traité dans les sections subséquentes de cet article.

## **Confidentialité**

Ces systèmes soulèvent des questions quant à la confidentialité des données. Clarke (2001) rappelle que la carte à puce n'est qu'une des très nombreuses technologies permettant de retracer l'identification et les habitudes d'une personne, parmi les systèmes bancaires, policiers, d'immatriculation, de péage routier et qu'en ce sens, des précautions similaires à celles de ces domaines doivent être prises pour protéger les informations nominales.

Conformément à la législation en vigueur quant à la protection de la vie privée, les données récoltées par les systèmes de perception par carte à puce peuvent être utilisées entre autres pour déterminer les patrons de déplacements des usagers. Ces données sont avantageuses également pour plusieurs usages, que ce soit à des fins tarifaires, marketing ou d'amélioration de service (Blythe 2004).

## METHODOLOGIE

Dans cette section, il sera d'abord question du système d'information utilisé à la STO pour la gestion et la collecte des données de paiement par carte à puce. Ensuite, le modèle-objet viendra mettre en perspective les objets et les dimensions à analyser.

### Système d'information et de validation des titres de la STO

La Société de transport de l'Outaouais (STO) opère une flotte de plus de 200 autobus dans la ville de Gatineau (240 000 personnes). Le système informatisé de validation des titres (SIVT) est opérationnel sur l'ensemble de son réseau depuis la fin 2001 et touche maintenant tous les titulaires de cartes mensuelles, soit plus de 80% des usagers.

Le système implanté à la STO utilise la technologie des cartes sans contact (*contactless*). Chaque usager dispose d'une carte identifiée à son nom (avec photo) contenant une puce de stockage de données et une antenne. La puce contient un identifiant et des données indiquant le type de titre et la date de validité. Lorsque l'usager monte dans un autobus de la STO, il passe sa carte devant un lecteur. Ce dernier transmet un signal d'excitation à la carte, qui lui retourne alors ses données. Le lecteur indique ensuite, par un signal lumineux et sonore, si la carte de l'usager est valide pour la ligne et la direction courantes.

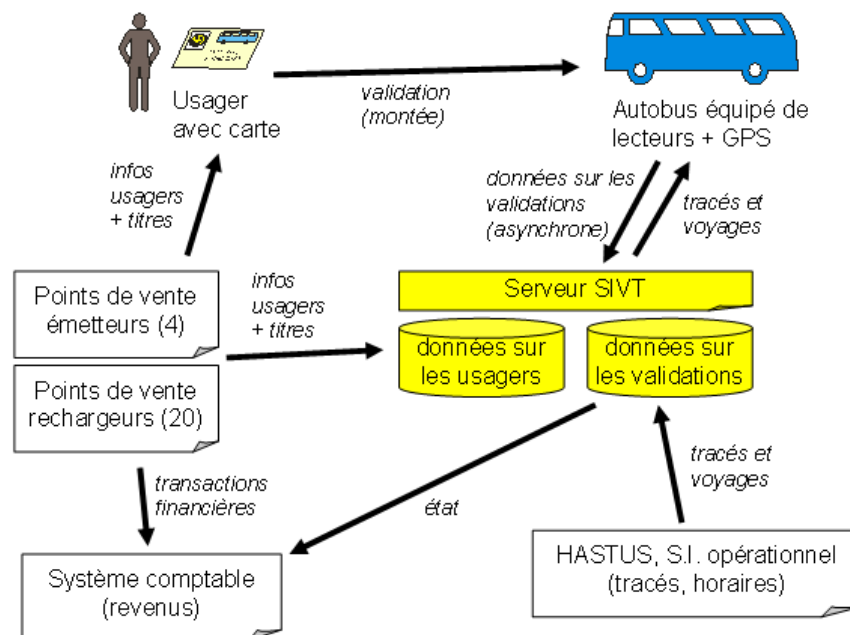


Figure 1. Diagramme fonctionnel du paiement par cartes à puce à la STO (Trépanier, Chapleau et Tranchant, 2005)

À la fin de la journée, les données collectées dans les autobus sont transférées dans une base de données centralisée. Aux fins de recherche, seules les données sur les validations nous sont rendues accessibles. Aucune information nominale sur les clients n'est disponible. Sous la plus stricte confidentialité, voici les variables examinées:

- Date et heure de la transaction ou de la validation;
- Type de tarif (Régulier, express, interzone, étudiant, aînés, etc.);
- Numéro et direction de la ligne empruntée;

- Numéro de l'arrêt où a eu lieu la transaction;
- Données opérationnelles telles que le numéro du conducteur, de l'autobus, du voyage, etc.

## **Modèle-objet**

L'approche orientée-objet en transport a été utilisée pour définir et quantifier les concepts inhérents à l'analyse de données de cartes à puce. Cette approche, développée par Trépanier et Chapleau (2001a,b), préconise l'identification des objets en présence selon quatre métaclasse qui encadrent les propriétés et méthodes pouvant être appliqués à ces objets : 1) dynamiques (qui se déplacent, tels que les véhicules et les personnes) ; 2) cinétiques (descripteurs du mouvement, tels que les routes et lignes) ; 3) statiques (supports immobiles faisant partie du système de transport) et 4) systémiques (fédérateurs des autres objets).

Le modèle-objet de la Figure 2 présente les relations entre les objets. Ceux-ci sont regroupés en quatre groupes selon leur fonction. Lorsque disponible, le nombre d'instances de ces objets est indiqué pour l'intervalle de dates couvrant la période d'analyse la plus large des expérimentations effectuées, soit celle de la modélisation du taux de survie (2004 à 2009).

Les objets du réseau de transport collectif (A) représentent les éléments « visibles » du réseau de transport. Les objets « ligne-direction », définis comme étant des tracés de lignes unidirectionnels, parcourent les objets « lignes-arrêt » selon une séquence d'objets « arrêt » déterminée pour chaque objet « ligne ». Le parcours d'un objet « ligne-direction » par un véhicule est appelé objet « voyage » dans le vocabulaire québécois. Enfin, citons l'objet « ligne de fuite », qui représente la séquence d'objets « ligne-arrêt » parcourue par l'usager qui effectue un déplacement (52,0 millions de montées dans ce cas-ci).

Les objets opérationnels (B) définissent la mécanique de fonctionnement du service. À la STO, sur la période donnée, 1017 chauffeurs distincts ont utilisé 343 autobus différents pour offrir le service. Cela a représenté plus de 400 000 journées de travail, divisée au Québec en pièces de travail (plans d'horaire contigus).

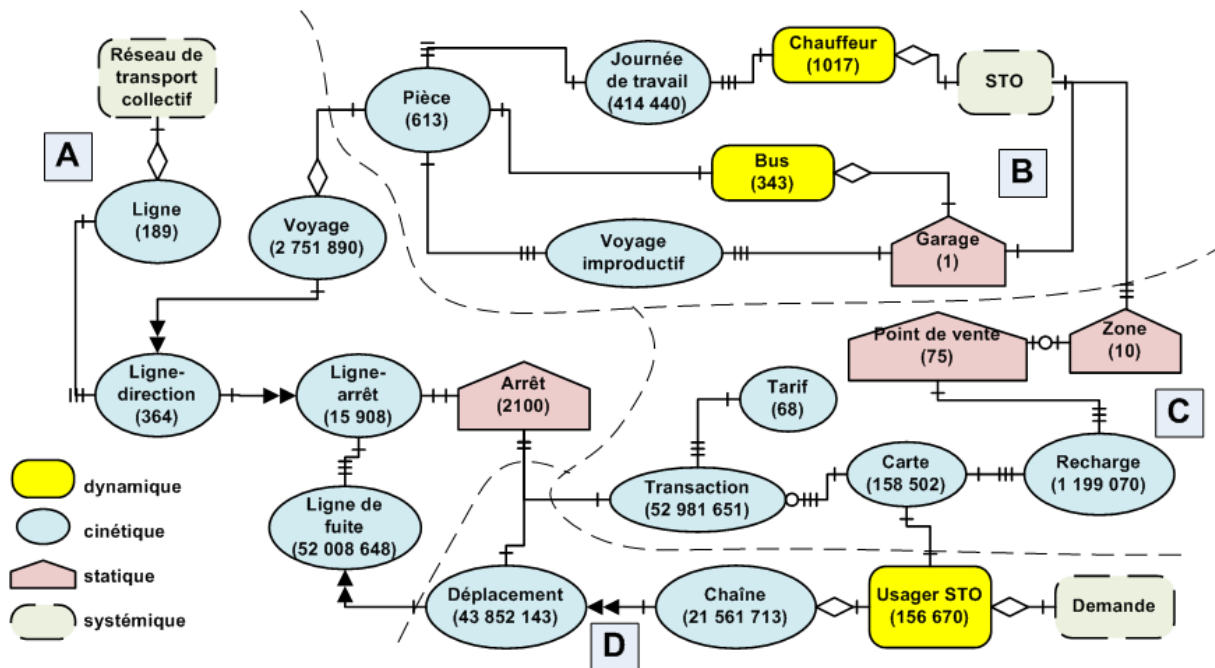


Figure 2: Modèle-objet correspondant à la période du 1er janvier 2004 au 30 septembre 2009

Les objets administratifs (C) regroupent les points de vente et sont caractérisés par les objets « recharge » (1,2 millions de fois durant la période), « carte » (158 000 émises), « tarif » (68 types différents, certains étant spécifiques à une année donnée) et « transaction » (plus de 52 millions).

Pour ce qui est des objets liés à la demande (D), on compte 156 670 usagers distincts, ayant fait un total de 21,6 millions de chaînes de déplacements (chacune définissant une journée de déplacements) pour un total de 43,9 millions de déplacements.

## EXPERIMENTATIONS

Cette section présente différentes expérimentations effectuées au cours des années avec les données du SIVT et de la STO. Le lecteur est invité à consulter les références citées pour obtenir plus de détails sur chacun de ces travaux.

### Traitement des erreurs

Le système de perception des titres de la STO est sujet à la présence d'erreurs, comme tout système d'information d'entreprise. Différents éléments expliquent ce fait. Il peut s'agir d'erreurs systématiques dues à des bris ou des défauts de l'équipement embarqué. Il peut également y avoir une désynchronisation entre les données opérationnelles (horaire planifié) et la réalité, comme par exemple dans le cas de pannes de véhicules ou de détours aux parcours. Bref, bien que la qualité des données se soit considérablement améliorée au cours des années, il a fallu proposer des méthodes pour détecter et, le cas échéant, corriger les données.

*L'exploitation des données de cartes à puce à des fins de planification des réseaux de transport collectif urbains  
(TRÉPANIÉ, Martin)*

Les premiers travaux effectués par Tranchant (2004) sous la direction de Trépanier et Chapleau ont permis de répertorier et de catégoriser les erreurs pour une période limitée. Par exemple, à la Figure 3, on constate que le taux d'erreurs est d'environ 1/15. On retrouve des transactions associées à une ligne « 0 » (sans attache), des arrêts situés sur une ligne inexistante pour la journée donnée, des arrêts en bout de ligne, etc. Notons que les erreurs n'ont pas pour conséquence d'éliminer la donnée associée. Elle est plutôt consignée et utilisée en vertu de la statistique calculée. Par exemple, le genre d'erreur décrit précédemment n'a pas d'incidence sur le nombre total de montées, mais pourrait affecter le nombre de montées d'une ligne précise sur le réseau.

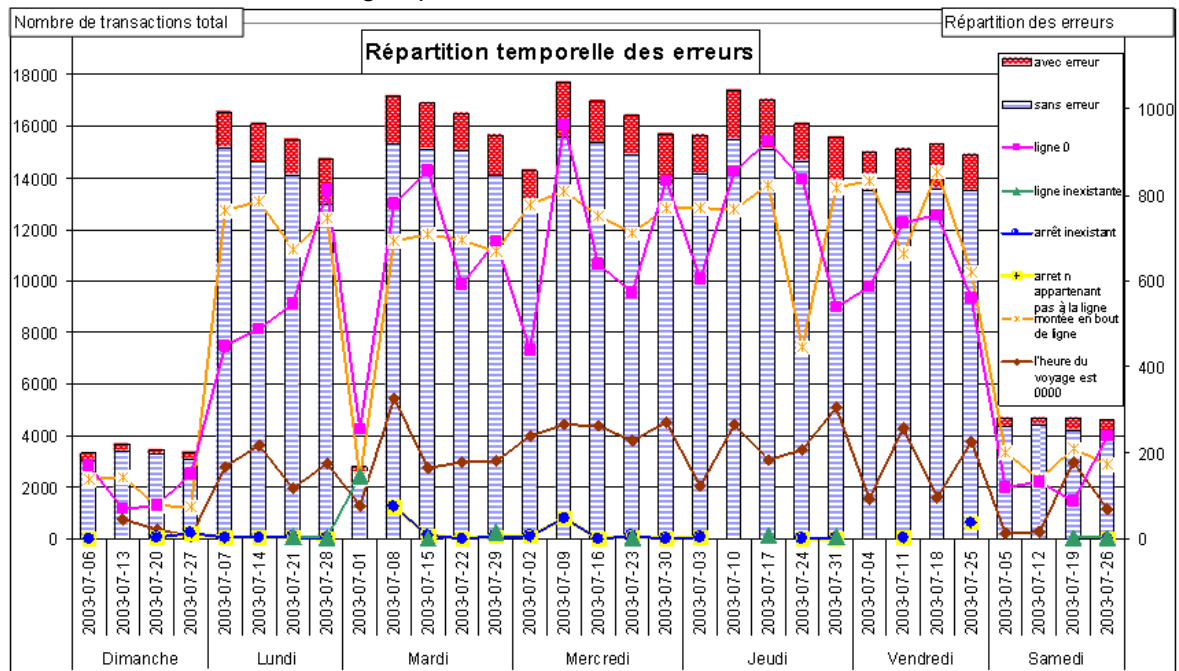
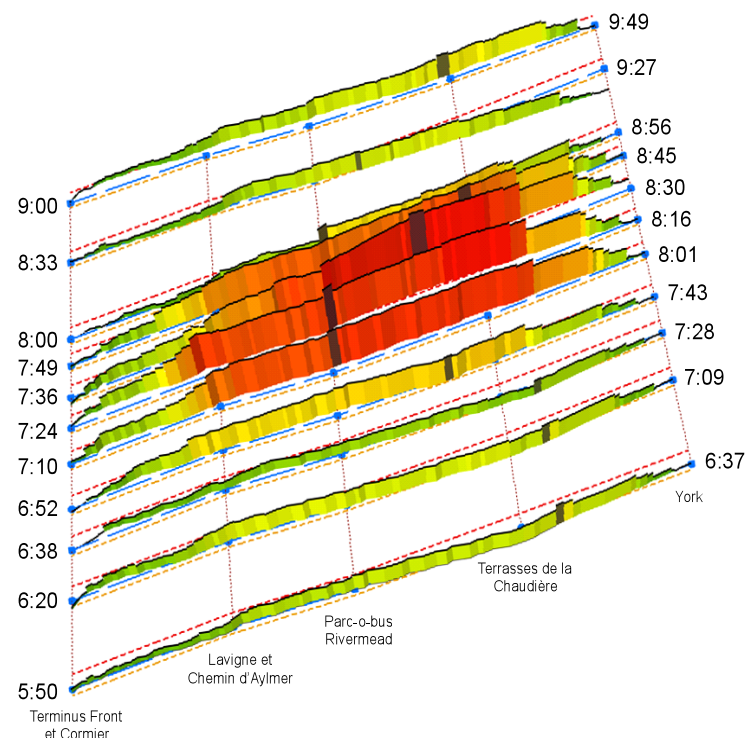


Figure 3: Statistiques d'identification des erreurs pour les données de juillet 2003

Dans des travaux plus récents (Chapleau et al. 2008), les méthodes avancées de validation des données ont permis de produire des profils de charge dénués d'erreurs (Figure 4), permettant d'examiner la charge sur une ligne d'autobus dans le temps et dans l'espace. On y constate le phénomène d'achalandage accru de la ligne pour la partie centrale de son parcours et à l'intérieur d'une période de pointe bien définie. L'obtention de ce niveau de détail permet au planificateur de mieux définir son service à partir de la courbe enveloppe.



**Figure 4: Profil de charge longitudinal dans l'espace-temps de la ligne 44 vers le centre-ville**

## Détermination de la destination

Le profil de charge tridimensionnel présenté à la section précédente ne peut être obtenu sans que l'on ait préalablement estimé le lieu de descente de chaque usager sur les lignes. En effet, le SIVT de la STO est conçu de telle façon que seule la transaction de montée est enregistrée, ce qui est le cas de la plupart des systèmes existants à travers le monde. Cependant, le SIVT a l'atout de pouvoir enregistrer la localisation spatiale du point de montée, ce qui facilite l'identification du point de descente.

La méthode développée repose sur la continuité des déplacements au cours de la journée (Trépanier et al. 2007). Ainsi, à chaque montée correspond une ligne de fuite regroupant les arrêts de descente potentiels (Figure 5). Le lieu de montée suivant de l'utilisateur est utilisé pour retrouver, dans cette ligne de fuite, l'arrêt le plus près (avec une tolérance maximale), constituant ainsi le lieu de descente plausible. Le déplacement est ainsi qualifié de « régulier ».



L'exploitation des données de cartes à puce à des fins de planification des réseaux de transport collectif urbains  
(TRÉPANIÉ, Martin)

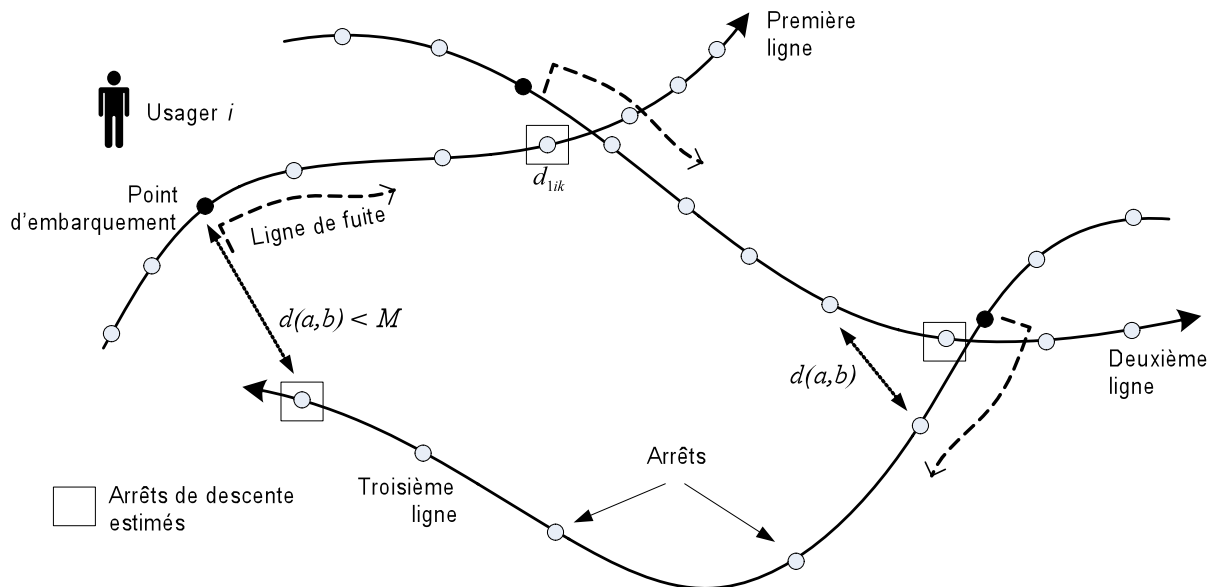


Figure 5: Schéma décrivant le modèle d'estimation des lieux de descente, pour les déplacements de type "normal"

Cette étape ne traite pas tous les cas. La Figure 6 présente les résultats de l'application de l'algorithme sur les données de février 2009. On doit également tenir compte du déplacement de la fin de journée. Ici, l'algorithme essaie d'identifier la destination en examinant successivement le premier lieu de montée de la journée et le premier lieu de montée du lendemain. Les déplacements qualifiés d'« unitaire » sont ceux dont il n'est pas possible d'identifier la destination à l'aide de la méthode de la ligne de fuite. Pour ces déplacements, on fait appel à l'historique complet de l'utilisateur pour essayer de retrouver des circonstances similaires (heure de départ, ligne, arrêt) pour lesquelles une destination aurait été estimée. En cas de succès, la destination trouvée est imputée au déplacement unitaire. Le taux de réussite de l'algorithme a été de 92,7% pour ce mois de février, mais il s'agit d'un taux hypothétique, puisque chacune des destinations trouvées n'a évidemment pas été validée.

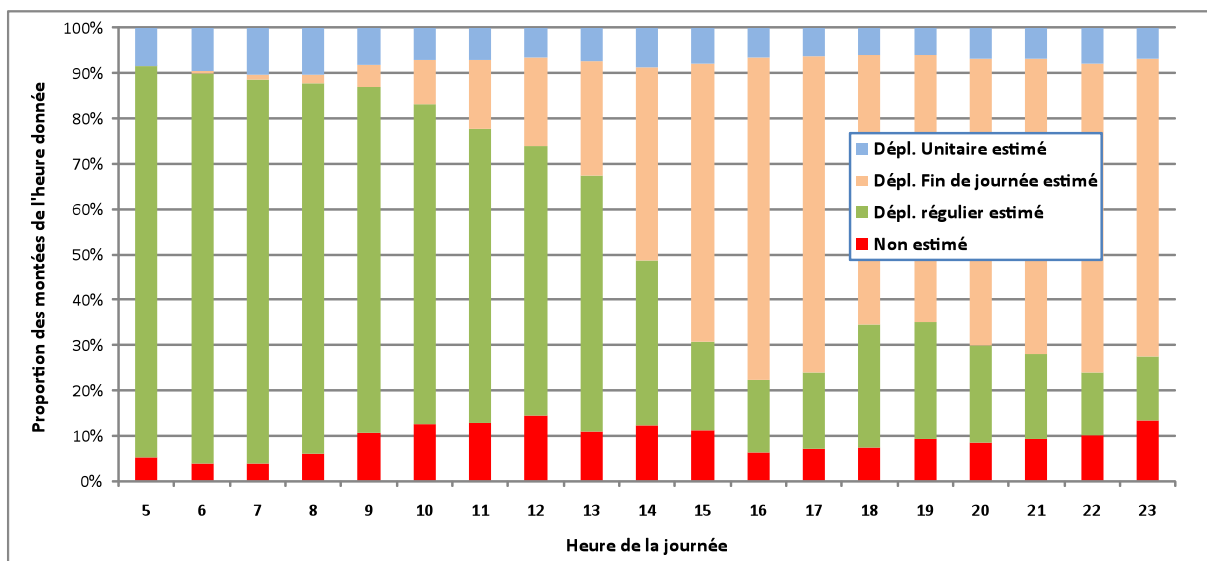


Figure 6: Résultats de l'application de l'algorithme d'estimation des destinations (données de février 2009)

## Diffusion de statistiques

Le système SIVT récolte des données en continu sur toutes les composantes du service d'autobus. Plutôt que de laisser dormir ces millions d'observations dans les entrepôts de données, il convient de diffuser à l'interne de la société de transport une série de statistiques simples découlant de l'opération. En plus de renseigner de façon très précise sur les conditions opérationnelle, cette diffusion a l'avantage de faciliter la détection d'erreurs dans le système, puisque les contremaîtres peuvent confronter la réalité rapportée aux statistiques calculées.

Le système développé calcule une série de statistiques descriptives sur l'opération, telles que le nombre de montants par ligne, le nombre de passagers-kilomètres, la répartition des types de titres utilisés, etc. Généré durant la nuit à partir des données de la veille, les statistiques peuvent également prendre la forme de graphes interactifs, grâce aux technologies XML (*eXtended Markup Language*, stockage et mise en forme de données) et SVG (*Standard Vector Graphics*, affichage). La Figure 7 est un exemple de profil de charge interactif généré en format SVG et pouvant être examiné à même un navigateur web. En déplaçant la souris, l'utilisateur peut interroger chacun des arrêts, déterminer les types de titres associés et connaître le nombre de correspondances, ce pour chacun des voyages opérés.

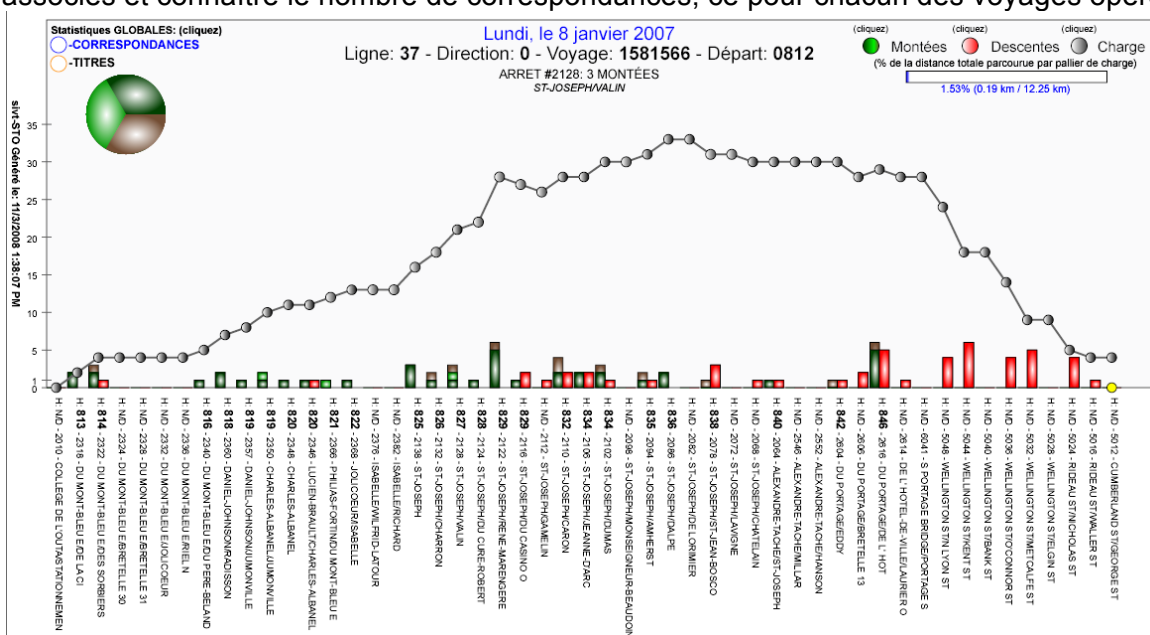


Figure 7: Exemple de profil de charge interactif généré sur demande pour une ligne donnée, à une date et une heure données

## Analyse du comportement des usagers

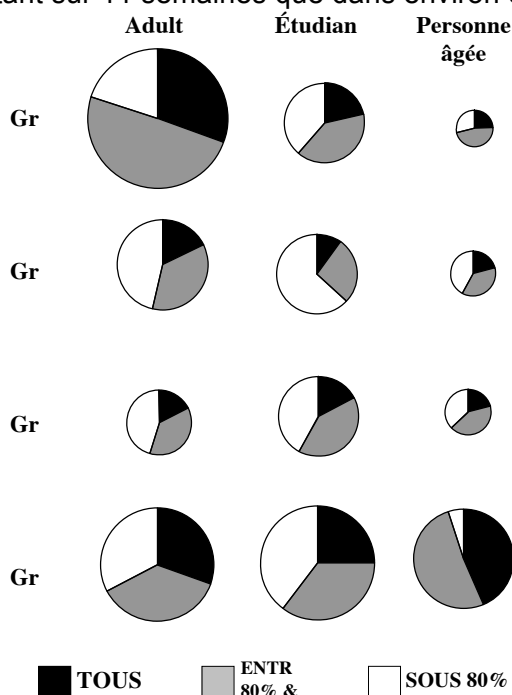
L'examen longitudinal d'un si grand nombre de transactions permet d'obtenir des informations de première main sur les comportements de la clientèle. Des techniques de data mining sont utilisées pour déterminer des groupes d'utilisateurs ayant des similarités en fonction de leurs habitudes de déplacements (Morency et al. 2007). Dans l'exemple rapporté ici, le concept d'utilisateur-semaine représente le comportement typique de l'utilisateur durant la semaine. Chaque observation est caractérisée par le nombre de montées

effectuées dans quatre périodes à chaque jour, pour un total de 28 paramètres. Une technique de « clustering » (k-means) est ensuite appliquée pour déterminer des groupes. Dans cet exercice, 4 groupes principaux ont été identifiés : le groupe 1 représente les déplacements pendulaires matin et soir, le groupe 2 se déplace souvent durant la journée, le groupe 3 représente les lève-tôt, et le groupe 4 se déplace assez peu. Le Tableau 1 présente la distribution des usagers-semaines en fonction du type de titre utilisé.

**Tableau 1: Distribution des usagers-semaines en quatre groupes selon le type de titre**

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

Ce genre d'analyse permet non seulement de savoir dans quel groupe se situe chaque carte prise individuellement, mais également comment varie cet appartenance au fil des semaines. En effet, le comportement des usagers n'est pas constant d'une semaine à l'autre comme en fait foi la Figure 8, où on constate par exemple que les adultes identifiés au groupe 1 n'ont un comportement constant sur 11 semaines que dans environ 30% des cas.



**Figure 8. Proportion des usagers selon leur appartenance aux groupes pendant 11 semaines d'observation**

Ce genre d'analyse peut être également conduit dans le cas des usagers-jours. On examine alors la journée typique en la découpant en 24 périodes d'une heure. Ce raffinement permet de distinguer de façon précise les variations comportementales de chaque individu, comme par exemple à la Figure 9 où on examine l'appartenance quotidienne d'une carte de type « aîné » à un ensemble de 30 groupes définis à partir de techniques de data mining.

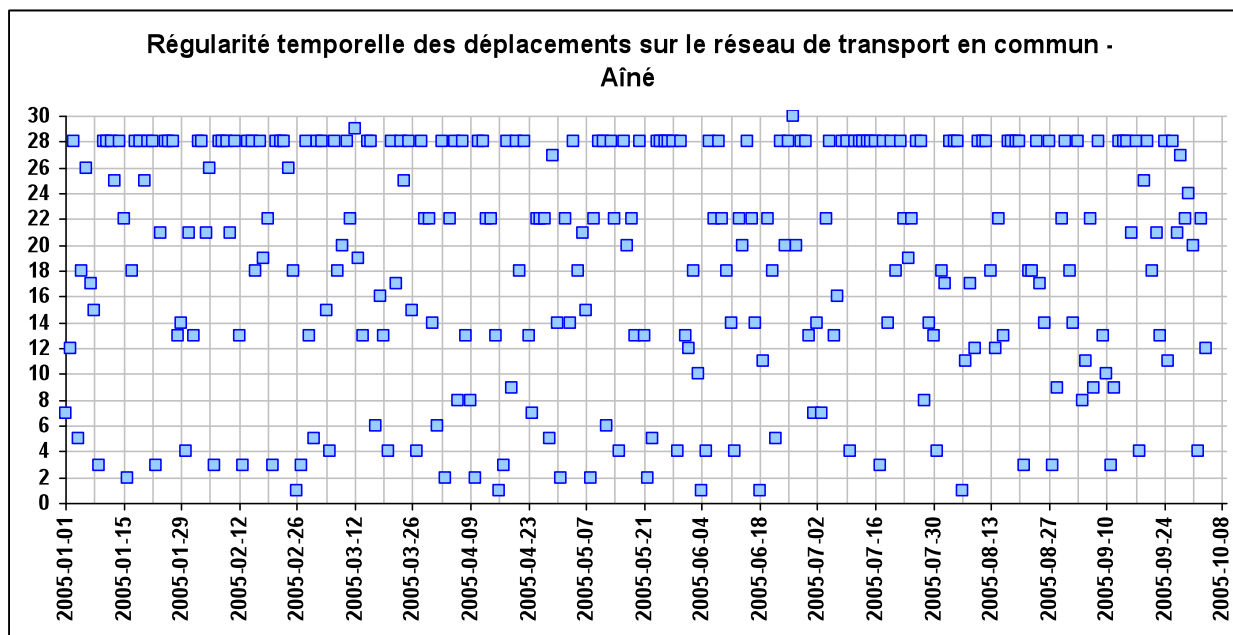


Figure 9. Appartenance des différents jours d'observation aux 30 agrégats formés – titre aîné

La finesse des observations peut également servir à examiner l'évolution de l'apprentissage du réseau par les usagers, c'est-à-dire le nombre d'arrêts différents utilisés au fil des jours. À la Figure 10, on observe une lente augmentation du nombre moyen d'arrêts « assimilés », puis un saut après une période de six mois, dû à l'arrivée de la saison estivale.

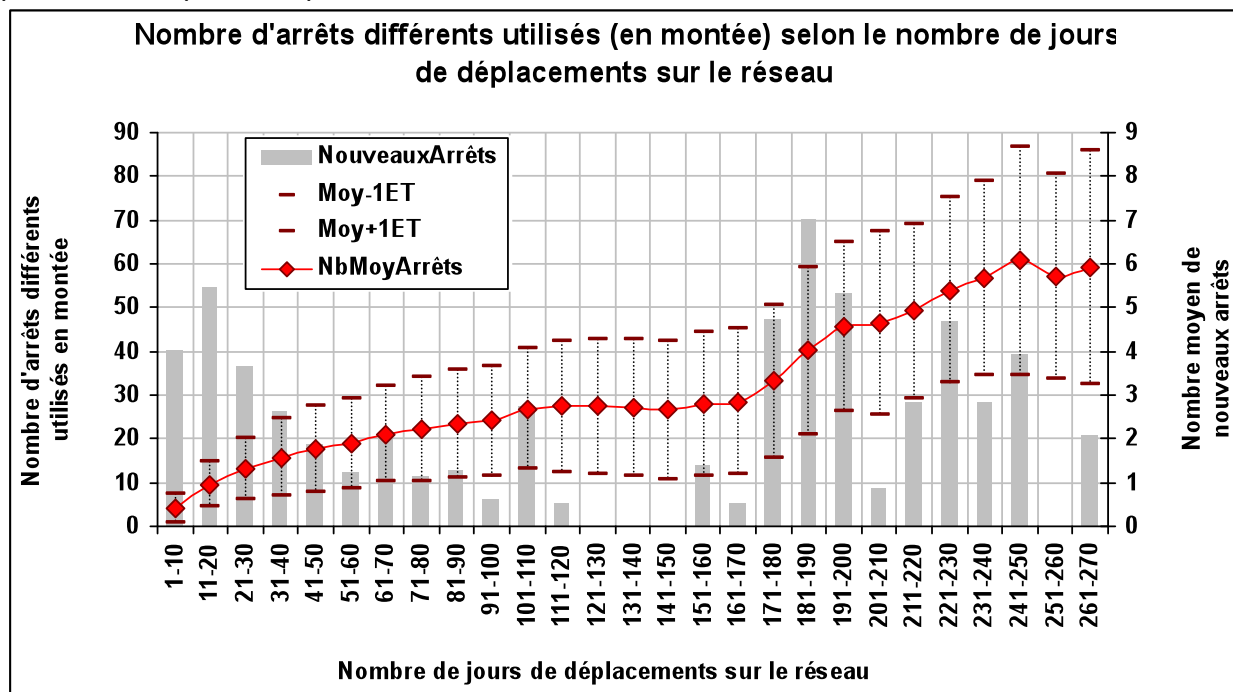


Figure 10. Chronologie d'utilisation d'arrêts différents du réseau de transport en commun sur une plage continue d'observation (janvier à octobre 2005)

## Analyse de l'offre

Bien que les données de cartes à puce soient traditionnellement associées à la caractérisation de la demande, il est tout aussi intéressant de les utiliser pour analyser l'offre de transport collectif. En effet, l'enregistrement de transactions sur les montées à un niveau très précis dans le temps et l'espace permet de calculer des indicateurs de service (Trépanier et al. 2009). Ces indicateurs peuvent être ventilés en fonction du type de titre utilisé, ce qui est impossible à faire à l'aide des systèmes d'aide à l'exploitation, à moins qu'ils soient justement couplés aux données de perception (Trépanier et al. 2009). Le Tableau 2 permet d'apprécier la finesse de ces résultats, qui plus est peuvent être obtenus par ligne, pour chaque journée et à des périodes horaires précises. On y remarque par exemple les différences significatives de vitesse commerciales entre les titre adulte régulier et adulte interzone (lignes suburbaines). De plus, les étudiants, qui effectuent 26,3% des montées, ne comptent que pour 21% des passagers-kilomètres.

**Tableau 2: Statistiques de l'offre de transport calculées à partir de données de cartes à puces et ventilées par type de carte (données de novembre 2006)**

Type de carte	% des montées	% des pass-km	% des pass-hr	Vitesse commerciale (km/h)	Longueur des dépl. (km)	Durée moy. des dépl. (min.)
Adulte régulier	46,2%	38,9%	42,3%	17,5	7,0	24,1
Adulte express	15,1%	21,8%	20,4%	20,4	12,0	35,3
Adulte interzone	3,0%	11,5%	8,4%	26,2	31,8	72,9
Étudiant	26,3%	21,0%	21,7%	18,4	6,7	21,7
Personnes âgées	3,5%	2,1%	2,3%	17,8	5,1	17,2
Autres	6,0%	4,7%	5,0%	17,9	6,6	22,0
Total	100%	100%	100%	19,0	8,3	26,3

Puisque les données disponibles à la STO sont dotées de géolocalisations précises, elles permettent de faire des analyses de ponctualité par ligne. Pour ce faire, il s'agit de comparer les heures de passages aux arrêts, enregistrées lors des transactions de montées, aux heures de passage planifiées selon l'horaire. La Figure 11 présente la distribution des observations de temps de passage en fonction de l'écart à l'horaire pour une ligne importante du réseau. On y remarque un léger retard par rapport à l'horaire, ce qui est conforme aux normes de service, où l'on garantit un passage moins de deux minutes après l'horaire, mais où il faut éviter d'avoir plus d'une minute d'avance.

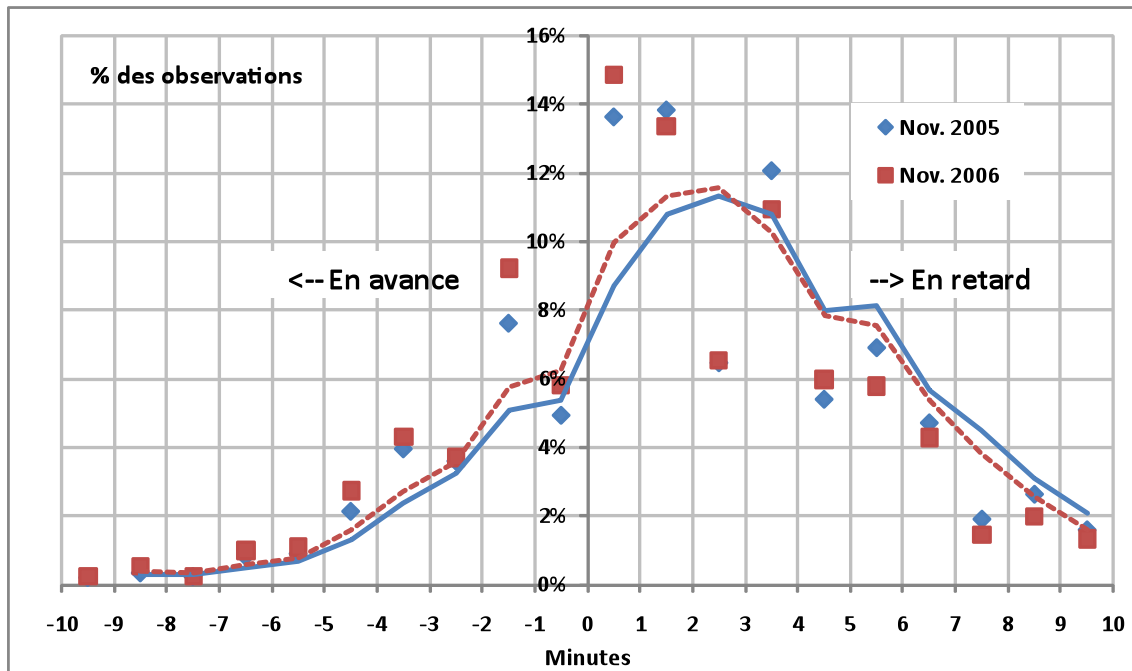


Figure 11: Conformité à l'horaire dans le cas de la ligne 37 pour les mois de novembre 2005 et 2006

## Analyse des générateurs de déplacements

Outre l'offre, le territoire peut également faire l'objet d'analyses à partir de données de systèmes de perception par cartes à puce. En effet, la connaissance fine des lieux de montées des usagers peut servir à caractériser l'utilisation du sol environnant. Chacun des arrêts du réseau est caractérisé à partir des données transactionnelles circonstanciées par date, heure, type de titre de transport. De même, on associe à ces arrêts des « comportements », c'est-à-dire un état représentant l'ensemble des comportements des usagers de cartes qui utilisent l'arrêt. La carte de la Figure 12 présente un certain nombre d'arrêts du réseau gatinois. Pour chaque arrêt, on présente la distribution des groupes de clientèle qui le fréquente. Les groupes ont été établis selon une logique comparable à ce qui a été expliqué précédemment (techniques de clustering), cette fois-ci basée sur les heures de montées de la journée. On y remarque que les arrêts purement résidentiel sont associés à des groupes comportementaux liés aux heures du matin, tandis que les lieux commerciaux et institutionnels (hôpital) sont liés à des groupes de milieu de journée. Le casino, pour sa part, est lié à un groupe de fin de journée (montée servant à retourner au domicile).

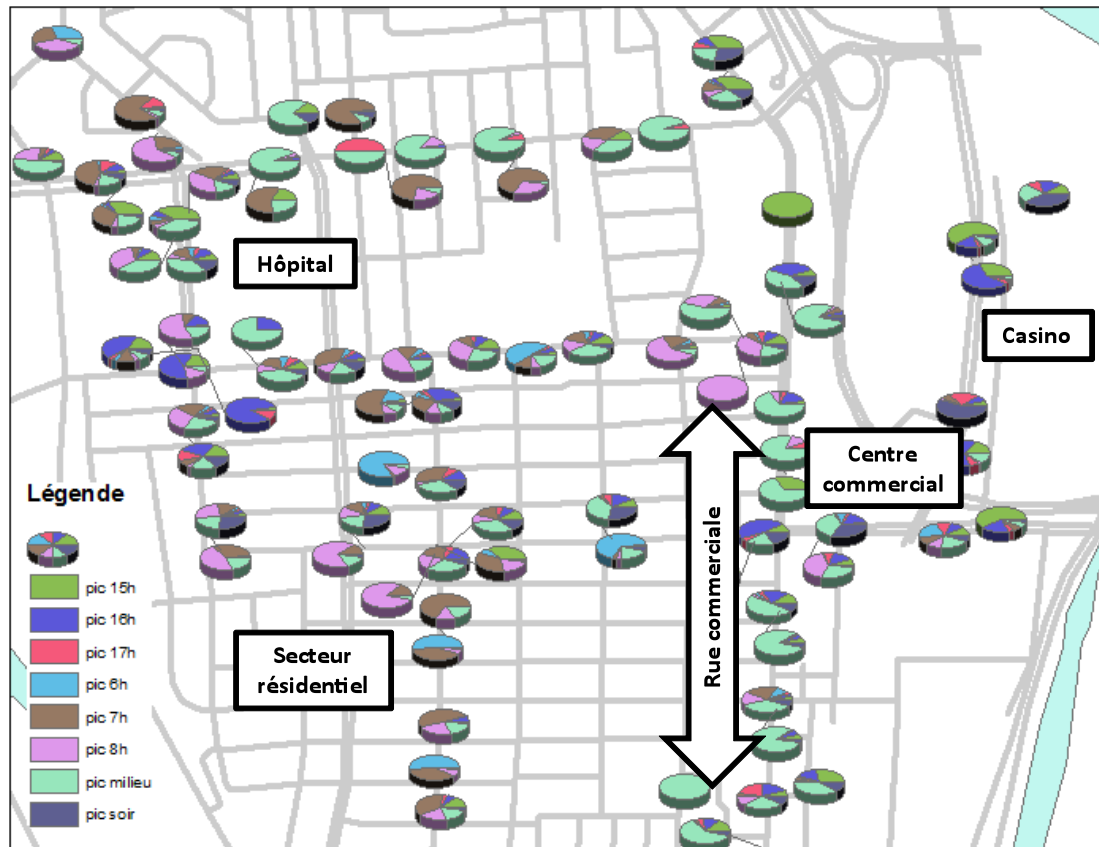


Figure 12: Carte des arrêts selon la répartition des montées par groupe (groupe régulier omis)

### Comparaison avec des données d'enquête ménage origine-destination

Une grande enquête ménage téléphonique s'est tenue dans la grande région d'Ottawa-Gatineau à l'automne 2005. Elle a été réalisée auprès de 23 000 ménages, ce qui représente un échantillon de 5,1%. Il s'agit d'une excellente occasion pour comparer les statistiques compilées dans cette enquête sur les déplacements réalisés par les usagers sur le réseau de transport collectif aux données transactionnelles de cartes à puce enregistrées au cours de la même période.

La Figure 13 présente les résultats de cette comparaison pour trois lignes importantes du réseau. Trois constats peuvent être tirés de l'analyse :

- Les différences sont notables au niveau des volumes observés sur les lignes, montrant d'un côté l'impossibilité d'obtenir une finesse suffisante avec les enquêtes ménage, et d'un autre côté la nécessité d'avoir une méthode d'expansion de l'enquête fondée sur l'achalandage mesuré et non seulement sur la socio-démographie.
- Les enquêtes origine-destination, de par leur concept de jour moyen de semaine établi à partir de deux mois d'entrevues, ne peuvent prendre en compte la variabilité quotidienne de l'achalandage sur le réseau, ce que peut faire l'analyse des données transactionnelles de cartes à puce.

- Cependant, les enquêtes origine-destination renseignent sur la socio-démographie des usagers et les circonstances des déplacements, ce que ne peut faire un système de perception par cartes à puce.

Somme toute, l'exercice a démontré qu'il convient de développer une méthode permettant de fusionner les données d'enquêtes aux données de cartes à puce afin de tirer profit des avantages de chaque ensemble informationnel.

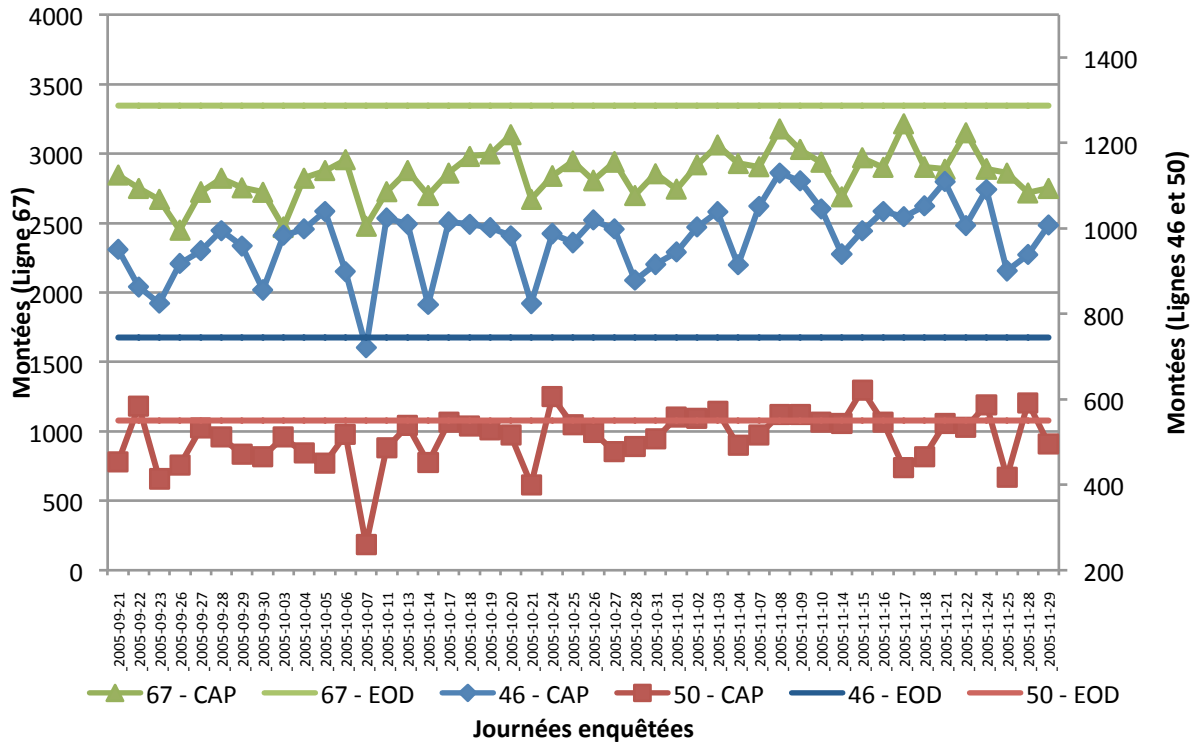


Figure 13 : Analyse de toutes les montées sur les lignes 46, 50 et 67 de la STO (période de l'enquête origine-destination de 2005)

## Survie des usagers

Bien que l'information nominale des usagers ne soit pas disponibles dans les jeux de données, il est possible de suivre les transactions d'une même carte au fil du temps. Ainsi, un examen longitudinal des cartes par type de titre permet de caractériser la loyauté des usagers en leur attribuant un taux de survie sur le réseau. Ce taux est borné par la date de « naissance », c'est-à-dire la date de première transaction sur le réseau, et la date de disparition, où a eu lieu la dernière transaction enregistrée.



*L'exploitation des données de cartes à puce à des fins de planification des réseaux de transport collectif urbains  
(TRÉPANIÉ, Martin)*

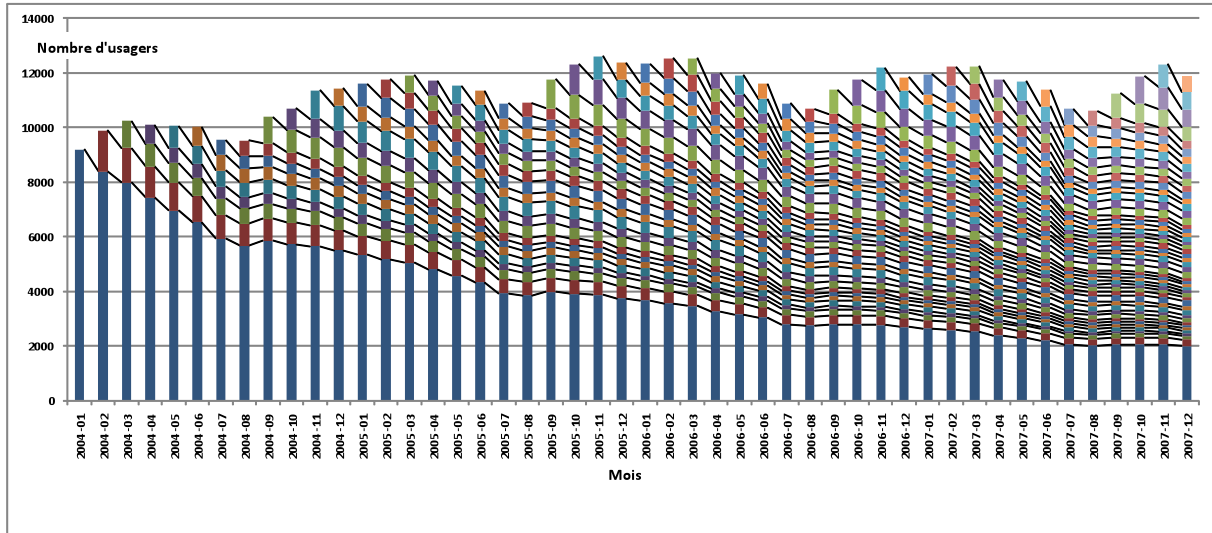


Figure 14: Profil de survie des usager du titre "adultes réguliers"

La Figure 14 permet d'apprécier la finesse de ce genre d'analyse. Pour chaque mois des années 2004 à 2007, on peut mesurer le nombre de cartes en service, et suivre chacune des cohortes de naissance associées à ces mois. On perçoit aisément les variations saisonnières du nombre d'utilisateurs actifs ainsi que la diminution progressive de la taille des cohortes nées à chacun des mois. À partir de cette mesure, on peut obtenir le taux de rétention du réseau, défini comme le nombre d'utilisateurs toujours actifs dans le système après une certaine période de temps. À la Figure 15, on remarque que les titres avec paiement bancaire (PB) ont un taux de rétention plus élevé que les autres. Les usagers des lignes express sont également plus fidèles que ceux des lignes régulières. Dans le cas des étudiants collégiaux et universitaires, on constate les chutes régulières de rétention dues aux saisons scolaires.

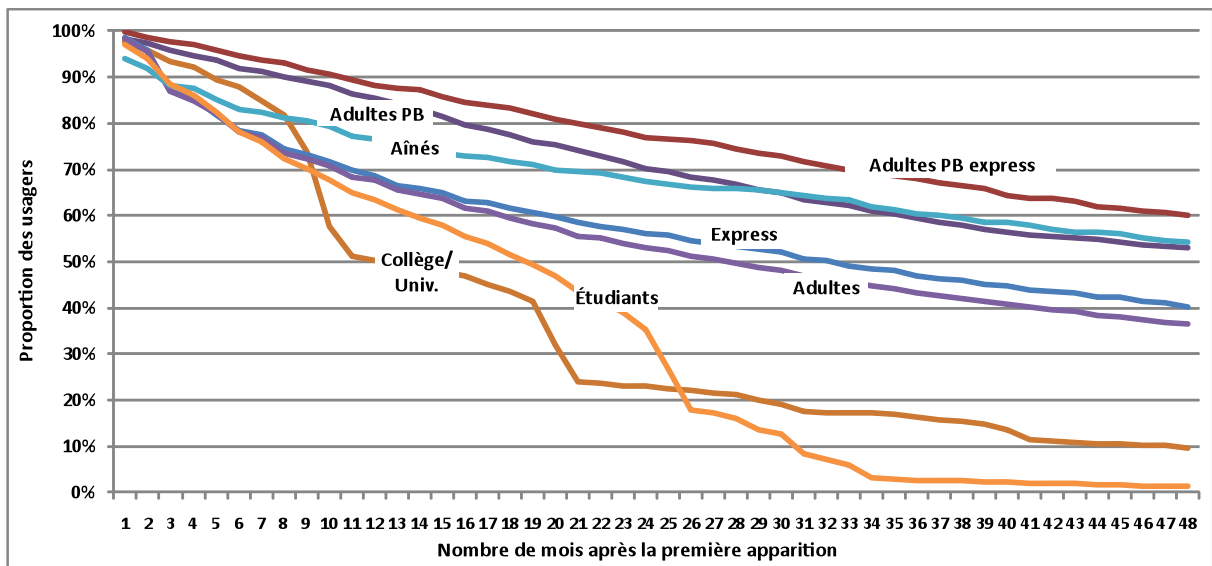


Figure 15: taux de rétention des usagers de cartes à puce apparus sur le réseau en 2004 et 2005, selon le type de titre

## **CONCLUSION**

Cet article a voulu bien humblement démontrer différents résultats de recherches effectuées à l'aide des données du système de perception par carte à puce de la Société de transport de l'Outaouais. L'utilisation d'une approche orientée-objet appliquée aux transports a permis de mettre la table pour une série d'analyse portant sur la demande et l'offre de transport ainsi que sur le territoire. La très grande quantité de données récoltées par de tels système est certes un avantage pour le chercheur, mais les travaux ont démontré que cette lourdeur, sinon cette indigestion, informationnelle, nécessite des nouveaux questionnements quant aux applications possibles, par rapport aux méthodes de planification actuelles, qui préconisent des modèles utilisant beaucoup moins de données factuelles.

Les perspectives de recherche sont nombreuses en ce qui concerne l'utilisation des données de cartes à puce en transport urbain. Premièrement, il est primordial de pouvoir enrichir ces données à l'aide de caractéristiques socio-démographiques afin de profiler les comportements des utilisateurs, comme le font les enquêtes origine-destination. Il est aussi nécessaire de poursuivre les études sur les variations quotidiennes et saisonnières de l'achalandage afin d'offrir aux opérateurs de transport des paramètres leur permettant de mieux ajuster leur service. Les effets externes sur les réseaux, tels que les conditions météorologiques et les événements fortuits ou non peuvent également faire l'objet d'études. Enfin, une intégration plus marquée des données au sein même des outils de planification doit être envisagée, afin de produire des simulateurs de réseau basée sur des données réelles et non microsimulées.

## **RÉFÉRENCES**

- Bagchi, M., White, P.R. (2005) The potential of public transport smart card data, *Transport Policy*, 12(5), pp. 464-474.
- Bagchi, M., White, P.R. (2004), What role for smart-card data from bus system?, *Municipal Engineer*, 157, pp. 39-46.
- Blythe, P.T. (2004). «Improving public transport ticketing through smart cards». *Proceedings of the Institution of Civil Engineers, Municipal Engineer*, vol 157, pp. 47–54.
- Bonneau, W. and editors (2002). The role of smart cards in mass transit systems, *Card Technology Today*, June 2002, p.10.
- Chira-Chavala, T., Coifman, B. (1996), Effects of Smart Cards on Transit Operators, *Transportation Research Record: Journal of the Transportation Research Board*, No. 1521, Transportation Research Board of the National Academies, Washington, D.C., pp. 84-90.
- Chapleau R., Trépanier M., Chu K.K.A., (2008). The ultimate survey for transit planning: complete information with smartcard data and GIS. *8th International Conference on Survey Methods in Transport*, Workshop B1 paper.
- Chu K.K.A., Chapleau R., Trépanier M. (2009), Driver-Assisted Bus Interview (DABI): Passive Transit Travel Survey using Smart Card Automatic Fare Collection System and its Applications, *Transportation Research Record: Journal of the Transportation*

*L'exploitation des données de cartes à puce à des fins de planification des réseaux de transport collectif urbains*  
(TRÉPANIÉ, Martin)

- Research Board, No. 2105, Transportation Research Board of the National Academies, Washington, D.C. pp. 1-10.
- Chu, K.K.A., Chapleau, R. (2008). «Enriching Archived Smart Card Transaction Data for Transit Demand Modeling», Transport Research Board, Washington, D.C., Vol. 2063, pp.63-72.
- Clarke, R. (2001) Person location and person tracking: Technologies, risks and policy implications, *Information Technology & People*, 14(2), pp. 206-231.
- Hofmann, M., O'Mahony, M. (2005), Transfer Journey Identification and Analyses from Electronic Fare Collection Data, Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria, September 13-16, pp. 825-830.
- Morency, C., Nurul Habib, K.M., Grasset, V., Zaman, H. (2009). Modelling Activity Persistency of Carsharing Members using Dynamic Ordered Probability Model, Transportation Research Board of the National Academies, Washington, D.C. paper no. 09-0693, 16 p.
- Morency, C., Trépanier, M., Agard, B. (2007). Measuring transit use variability with smart-card data, *Transport Policy*, 14(3), pp. 193-203.
- Park, J.Y., Kim, D.J. (2008). «The potential of using the smart card data to define the use of public transit in Seoul ». TRB 2008 Annual Meeting, Transportation Research Record, 20 pages.
- Tranchant, N. (2004). Analyse des déplacements d'usagers à partir de données de cartes à puce, Mémoire de maîtrise en génie industriel (M.Sc.A.), 178 p.
- Trépanier, M., Barj, S., Dufour, C., Poilpré, R. (2004), Examen des potentialités d'analyse des données d'un système de paiement par carte à puce en transport urbain, Exposé préparé pour la séance sur "Utilisation des systèmes de transport intelligents (STI) à l'appui de la gestion de la circulation" du congrès annuel de l'Association des transports du Canada, Québec, pp. 10-14.
- Trépanier Martin, Morency Catherine, Agard Bruno, (2009). Calculation of Transit Performance Measures Using Smartcard Data. *Journal of Public Transportation*, 12(1), pp. 79-96.
- Trépanier M., Vassivière F., (2008). Democratized Smartcard Data for Transit Operators. 15th World Congress on Intelligent Transport Systems, New York, États-Unis, 16-20 novembre.
- Trépanier, M., Chapleau, R., Tranchant, N. (2007). Individual Trip Destination Estimation in Transit Smart Card Automated Fare Collection System, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, Taylor & Francis, 11(1), pp. 1-15.
- Trépanier, M. Chapleau, R. (2001a). Linking Transit Operational Data to Road Network with a Transportation Object-Oriented GIS, *Urban and Regional Information Systems Association Journal*, 13(2), pp. 23-27.
- Trépanier M., Chapleau R. (2001b), Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination, *Revue canadienne de génie civil*, vol. 28, no 1, pp. 48-58.

*L'exploitation des données de cartes à puce à des fins de planification des réseaux de transport collectif urbains*  
(TRÉPANIÉ, Martin)

Utsunomiya, M., Attanucci, J., Wilson, N. (2006), Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning, Transportation Research Record: Journal of the Transportation Research Board, No. 1971, Transportation Research Board of the National Academies, Washington, D.C., pp. 119–126.