

C. Morency, M. Trépanier, B. Agard, Measuring Transit Performance using Smart Card Data, World Conference on Transport Research, San Francisco, USA, June 24-28, 2007.

Measuring Transit Performance using Smart Card Data

Catherine Morency^{1,2}, P.Eng., Ph.D., assistant professor

Department of Civil, Geological and Mining Engineering

Ecole Polytechnique de Montréal, P.O. box 6079, station Centre-Ville, Montréal, Qc,

CANADA, H3C 3A7

email : cmorency@polymtl.ca

Martin Trépanier^{1,2}, P.Eng., Ph.D., associate professor

Department of Mathematics and Industrial Engineering

Ecole Polytechnique de Montréal, P.O. box 6079, station Centre-Ville, Montréal, Qc,

CANADA, H3C 3A7

email : mtrepanier@polymtl.ca

Bruno Agard¹, Ph.D., assistant professor

Department of Mathematics and Industrial Engineering

Ecole Polytechnique de Montréal, P.O. box 6079, station Centre-Ville, Montréal, Qc,

CANADA, H3C 3A7

email : bruno.agard@polymtl.ca

1 – Centre de recherche interuniversitaire sur les réseaux d'entreprise, la logistique et le transport (CIRRELT)

2 – Groupe MADITUC

Measuring Transit Performance using Smart Card Data

Catherine Morency, Martin Trépanier and Bruno Agard

Session E2 - Transportation system performance measurement

Abstract

Smart card automated fare collection systems (SCAFC) for transit have mainly been considered for their administrative function of controlling the access to the service and for revenue management. However, we believe that data from these systems can also be used to describe both transport demand and supply. This article illustrates the use of smart card data to estimate various transit network performance measures. Combined to well-established evaluation processes, such measures could assist operators monitor in more details their network. The performance of the network supply (vehicle-kilometres, vehicle-hours, commercial speed, etc.) and the statistics on passenger service (passenger-kilometres, passenger-hours, average trip length, etc.) can be calculated from these datasets for any spatial and temporal levels of resolution.

Acknowledgments

The authors wish to acknowledge the Société de transport de l'Outaouais (Gatineau's Transit Authority) which graciously provided data for this study. The research project is also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT) and the Agence métropolitaine de transport of Montreal (AMT).

Measuring Transit Performance using Smart Card Data

1. Introduction

Smart card data systems have mainly been considered for their administrative function of controlling the access to a service, namely a transit network. They help improve the transit user's satisfaction, with simplified ticketing options (unique card, security), while enhancing revenue collection for public authorities (fraud, multilevel validation) (Conklin et al., 2004). In order to do their job, these systems need to record a large amount of information on the daily use of the transit network. Each transaction is recorded along with spatio-temporal details: time, spatial location, operational information (line, stop), card type (fare type and privileges). Even if the resulting dataset was not, a priori, conceived for analytical purposes, we believe that it can be processed in order to reveal information on how the network is rendered and used every day. Actually, with systematic procedures, such data can be exploited to measure the use of the transit service on every single day of observation.

In order to validate this opportunity, a set of continuous data from a mid-size transit authority was recently obtained and first exploited to observe the variability of the travellers over a ten-month period (Morency et al., 2006). These data are from the Société de transport de l'Outaouais (STO) smart card fare collection system. This public transport authority operates 200 buses and services 240 000 inhabitants in Gatineau, Quebec. Its smart card system was implemented in 2001. Today, more than 80% of all STO passengers hold a smart card. With these continuous data, daily, weekly and seasonal activity cycles were identified for various transit card types using data mining techniques (regular adults, students and elderly for instance).

This confirmation of the variability of transit use during these cycles suggests that the supply might not always be adequately matching the real transit demand.

“Measuring the performance of a transit system is the first step toward efficient and proactive management” (Bertini and El-Geneidy, 2003). With this question in mind, the current research steps into the supply side of the equation and aims to estimate some transit performance measure using smart card data. Actually, we believe that smart card data offer a unique opportunity to simultaneously monitor the use and supply of a transit network on every particular day. With a high and continuous sampling of the transit network, we demonstrate that it is possible to estimate indicators of the transit network, as it is offered everyday. Hence, the purpose of the paper is to evaluate the potential of smart card data to derive operational indicators revealing the service that was truly offered to the users on a specific day (compared with the planned service: vehicle-kilometres for instance), estimate the use of the service on this same day (demand: passenger-kilometres for instance) and observe how these two sides of the transit network evolve in time.

The paper is organized as follows. First, we present some work regarding the use of smart card data for analytical purposes. Some transit performance indicators are then presented keeping in mind that the main purpose of the paper is to evaluate the usability of smart card data to estimate some of these classical indicators. Then, the methodology is presented namely the dataset used for the experiment, some imputation operation and the description of the transit performance measures. Results of the estimation of indicators from a demonstrative dataset are then presented. A discussion concludes the paper.

2. Background

2.1 *Smart card in transit*

The use of smart cards in transit network is increasing all over the world. Smart cards are based on a technology patented in 1968 by German researchers Dethloff and Grotrupp (Shelfer and Procaccino, 2002). They serve to identification and transaction by exchanging information with readers installed within vehicles. Smart cards are useless without a strongly integrated information system that link planned, operational and user (card) data: Smart Card Automated Fare Collection System (SCAFC). The complex fare system that is used by many public transport authorities can be better managed with the help of a SCAFC, because smart cards can store more than one transport document at a time and the card is validated automatically. The need to integrate fare policies within large metropolitan areas will continue to promote smart card usage (Bonneau, 2002).

Bagchi and White (2005) reported several advantages brought by the analysis of SCAFC data like the access to larger sets of individual data, the continuous data available for long periods of time and the better knowledge of a large part of the transit users. These authors conducted a study on the passenger transfer behaviours on the Bradford and Merseyside transit networks (UK). The absence of alighting location information was then identified as the main issue for further analysis.

Moreover, privacy is an important issue that could retain smart card implementation. The French Council for Computer and Liberty recommends being careful with such data because one may reconstitute the personal movements of a specific person (CNIL, 2003). But Clarke (2001) recalls that smart card data is not different from

other individual data collection systems like credit card usage, road tolls, police corps database.

In the case of the Société de transport de l'Outaouais (STO), Trépanier et al. (2004) have shown the potentialities of using SCAFC data for public transport network planning with the help of a Transportation Object-Oriented Modelling approach. Furthermore, it has been possible to develop a destination location estimation model for each trip (Trépanier and Chapleau, 2006, Trépanier et al., 2007). The model, which relies on individual spatial patterns of use, has been applied to the 6.2 millions trips of the dataset with a 94% success rate (an alighting stop was imputed). Smart card data can also be useful for travel behaviour analysis. Morency et al. (2006) have shown the potentialities of applying data mining clustering methods to datasets to observe similar behaviours and variability among the card users.

2.2 Transit performance measures

The first edition of the Transit Capacity and Quality of Service Manual (TCQSM) was published in 1999 (Kittelsohn and associates, 1999). This manual recommended the evaluation of transit systems using six performance measures: service frequency, hours of service, service coverage, passenger loading, reliability, and transit vs automobile travel time. Many examples of transit performance evaluation using the TCQSM can be found in the literature (among others Perk and Foreman, 2003, Caulfield and O'Mahony, 2004). These measures are generally estimated using surveys or on-board counts that give a snapshot of the service on a particular day. There is no information regarding the variability of the service or the validity of the day when the data were collected (is it representative of the other days). This is where smart card data become interesting.

Gillen et al. (2001) study the impact of Automated Vehicle Location technology (AVL) on transit firm productivity and observe that the use of such system “*will result in fewer buses used given the number of vehicle miles and number of passenger trips*”. Cost per mile and annual maintenance hours also reduce. The acceptance of Intelligent Transportation Systems (ITS) by the transit system users has also been investigated. Using intercept-mail-back survey (approx. 3000 respondents), Conklin et al. (2004) reported that users are widely supportive of such systems.

Bertini and El-Geneidy (2003) demonstrate that archived stop-level data can be converted into valuable Transit Performance Measures (TPM). Actually, they argue that many TPM have been proposed in the past but rarely been implemented because of data limitations. They confirm that the deployment of Intelligent Transportation Systems (ITS) to monitor and manage transit networks also enable transport authorities to monitor the performance of the network in real time or in retrospect. Smart card data have this potentiality. They use archived data from the Portland local transit provider to determine various performance measures and compare them from day to day and from route to route.

3. Methodology and concepts

3.1 Dataset

Data from the Société de transport de l’Outaouais (STO) smart card fare collection system were used for the study. The STO is a medium-size public transport authority operating 200 buses and servicing 240 000 inhabitants in Gatineau, Quebec. The STO operates its smart card system since 2001. Today, more than 80% of all STO passengers hold a smart card. Moreover, every STO bus is equipped with GPS

reader. At each boarding, stop location and bus route are stored in the database along with a timestamp. Since the STO uses a high-level secured procedure to ensure the privacy of the data, smart card data are completely anonymous. No nominal information on user is known in any kind.

Table 1. Information on smart card transaction in the dataset

The set of data used for the experiment relates to February 2005 and consists in 28 consecutive days of observations (4 complete weeks). Table 1 shows the general structure of the information available for each smart card boarding transaction.

For this study, the dataset has been organized by run number. A run is a sequence of stops serviced by a transit route, in one direction. For the study period, the number of runs varies between 1341 and 1448 for weekdays (avg: 1400), 629 and 667 for Saturdays and 382 and 419 for Sundays. These are the number of runs reported in the SCAFC system. It is thus possible that some runs were operated but not reported.

The estimated alighting location has been added to the dataset to obtain more precise load profile for each run. Destination stop number is estimated with the Trepanier et al. (2007) algorithm. No other method is currently available. For a given trip, the destination (final or alighting point of a multi-segment trip) is estimated from:

1. the location of the next boarding during the day, if any;
2. the location of the first boarding of the day, if it is the last trip;
3. the other trips of the same card made at same time and on same route.

A field indicates the level of accuracy of the estimation for each trip. Imputation of alighting stops for an entire day of observation requires app. 30 minutes.

3.2 *Transit performance measures*

The calculation of transit performance measures are made with the help of database queries involving Structured Query Language (SQL) and customized database functions. Two dimensions are examined:

- Query on transit operation. This involves the linkage of tables on route geometry, run length and travel time, and effective runs (run that were operated regarding the transaction database). Main results are number of runs for each route, vehicle-kilometres and vehicle-hours.
- Query on transit use. This query links the transaction database, the destination location estimation and the route geometry. Trip duration is estimated from the travel length on the route and the commercial speed (average speed which takes into account the stops), because the exact alighting time is not known, especially at the terminal stop (let us remember that the SCAFC store the boarding times only, so there is no transaction at the last stop of the route). Main results are number of passengers, passenger-kilometres and passenger-hours for each run segment.

3.2.1 **Supply-oriented measures**

The supply-oriented measures are calculated as follow. The number of vehicle-kilometres D_{ij} of a single run i at day j , is equal to the sum of the travel distance Δd_{ij} between each stop of the run. The same logic is applied for the number of vehicle-hours T_{ij} . Since each measure is disaggregated, the total number of vehicle-kilometres for a route or the whole network is calculated by adding the run's numbers. The number of vehicle-kilometres D_{rj} of a route r on a given day j is :

$$D_{rj} = \sum_{i \in r} \sum_{k \in i} \Delta d_{kj}$$

The same applies for the number of vehicle-hours. Thus, the average commercial speed of a route (C_r) is :

$$C_r = \frac{D_r}{T_r} = \frac{\sum_{i \in r} \sum_j D_{ij}}{\sum_{i \in r} \sum_j T_{ij}}$$

These indicators are for the operated service only and do not consider the dead time from and towards the depot, or the waiting time during the day.

Average length and duration of runs can then be estimated using veh-km, veh-hre and total number of runs operated.

3.2.2 Demand-oriented measures

Similar indicators are obtained for the demand side, where every single trip is put into the calculation. The travel distance in kilometres (d_{iu}) of a user u on a single run i is :

$$d_{iu} = \hat{\sigma}_i^{s=a} - \hat{\sigma}_i^{s=b}$$

where $\hat{\sigma}_i^{s=a}$ is the cumulated distance on the run at the estimated alighting stop location and $\hat{\sigma}_i^{s=b}$ the cumulated distance on the run at the boarding stop. The number of hours (t_{iu}) is estimated with the help of the commercial speed of the run (C_i) because the alighting time is unknown :

$$t_{iu} = \frac{d_{iu}}{C_i}$$

The total number of passenger-kilometres and passenger-hours for a route (a run or a day) are calculated by adding the kilometres travelled by every boarding passenger.

Let us remind that the estimation of passenger-kilometres and passenger-hours require the imputation of an alighting stop for travellers. The estimation method can be found in Trepanier and Chapleau (2006).

Average length and duration of a trip can be estimated using passenger-kilometres or passenger-hours divided by the number of transit network users.

3.2.3 Performance measures

The combination of supply and demand oriented indicators leads to overall performance measures. In this case, partial measures are calculated because some user do not have a smart card. A correction coefficient would need to be applied to expand the number of users for each run. This coefficient can be calculated from external passenger count observations. However, this demonstration paper presents statistics on smart card use only since adjusted statistics remain confidential to the transit operator.

The average bus occupancy β_{rj} of route r on a given day j is :

$$\beta_{rj} = \frac{\sum_{i \in r} \sum_u d_{iju}}{D_{rj}}$$

where d_{iju} is the travel distance of the passenger u on run i and day j , and D_{rj} the number of vehicle-kilometres on run i and day j . The capacity use ratio α_r is then estimated similarly by adding in the equation the vehicle capacity V_i at each run :

$$\alpha_r = \frac{\sum_{i \in r} \sum_u d_{iju}}{D_{rj} V_i}$$

Vehicle capacity is available for each run because the bus number is recorder at each transaction. The capacity of a regular bus is approximately 75 passengers (41 seats) and 65 passengers for a low-floor bus (38 seats).

The punctuality of the bus service can be evaluated by comparing the planned schedule to the actual bus stop passage time. This time is supposed to be the time of the first of all smart card transactions at each stop, for a given run. A special dataset is built to measure the punctuality of the bus service. It contains the bus passage time for each stop of each run of February 2005. This case study is restricted to route 37, one of the most loaded routes of the STO network.

4. Results

The list of transit performance measures is quite straightforward and has been part of the operator knowledge for some time. However, their estimation is still a burden for ill-informed authorities. The following results demonstrate the use of smart card data to estimate these measures globally but also their ability to produce specific indicators depending on the critical issues faced by operators. Actually, every day of operation, run or even transit stop can be described with respect to supply and demand conditions. All this results can be estimated daily in few minutes.

4.1 Supply-oriented measures

More than 32 000 runs have been offered during the month of observation resulting in almost 500 000 veh-km and 23 200 veh-hre. Average commercial speed is 21.4 km/h and runs have an average duration of 43.2 minutes. These global figures can be segmented according to the type of day to see if the service is equivalent on weekdays and also to measure difference during the weekends. From Table 2, we see that 87% of the runs are offered during weekdays. The distribution of runs, veh-km

and veh-hre is quite uniform across weekdays (around 17.5%) and less during Saturdays (half of what is offered during weekdays) and Sundays (quarter of weekdays). The average speed varies between 21.1 km/h and 22.6 km/h and peaks on Sundays, as expected.

Table 2. Key facts regarding transit supply during the four weeks of February 2005 by day of travel

The daily variability of such indicators can also be estimated, mostly to observe if there were some out of average days of supply.

4.2 Demand-oriented measures

Table 3 presents key facts based on demand data. For every day of the week, indicators of the use of the transit system by all types of users holding a smart card are presented. It is noticeable that:

- Almost 650 000 passengers have used the system in February 2005, more than 92% during weekdays.
- The average in-vehicle speed of journeys is 21.2 km/h and it is quite stable across the various days
- The average length of a journey is less than 9 km. It is shorter during week-end days; there is also a slow decrease of the length and average duration from Monday to Friday. The average duration of a transit trip is 25 minutes.

Table 3. Key facts regarding transit demand during the four weeks of February 2005 by day of travel

These summary indicators can also be estimated for the main types of smart cards. This segmentation allows understanding the travel patterns of student, elderly and express card holders for instance. A classification of the various card type in classes

is used for the purpose. First, the same indicators as presented previously are shown for the main classes of card types (see Table 4). It shows that regular cards for adults account for almost 50% of passengers boarding during February 2005. The second most important card type is for students. These figures also reveal some distinctions between the features of trips done on the transit network by these client groups. Apparently, interzone cards have access to more efficient service since with a much higher speed. This is due to the nature of interzone service which uses high capacity road network instead of local streets to link distant destinations. Express cards are second in terms of average speed. All other client groups have comparable average speeds of travel. The average length and duration of journeys on the transit network reveal similar facts. Interzone cards are used for much longer trips, followed by express cards. Elderly card holders do the shortest trips on the network, followed by students.

Table 4. Key facts regarding transit demand by card type during the four weeks of February 2005

The study of passenger boarding patterns at each stop of the network over a 4 months period is done using data mining clustering techniques. The dataset for this study contains the number of boarding at each stop, for each day and for each 1-hour period during the day. The purpose of the study is to identify different clusters of stops. These clusters have similar temporal load levels. These results could be used for marketing purpose, installing new equipments, targeting investments, etc.

Figure 1. Temporal distribution of boardings for the stops' clusters (Feb. 2005 data)

The result of the data mining clustering is illustrated at Figure 1. It shows the typical boarding distribution for three stop clusters that were found: “daytime” (boarding

are spread through the day), “AM peak” (mostly AM) and “PM peak” (mostly PM). About 70% of the stops always are in the same cluster for all weekdays.

Figure 2 maps the stops to show the spatial distribution of the clusters. “AM peak” stops are mostly found in residential areas, while “PM peak” stops are located in industrial parks and business centers. “Daytime” stops are placed along the main commercial streets. In fact, this kind of study helps to identify activity zones without having any prior knowledge of the territory.

Figure 2. Spatial distribution of stops by stop cluster, Feb. 2005 data

4.3 Performance measures

The cross-examination of demand and supply indicators reveals that the passengers/run ratio is higher during the weekdays (see Table 5). Moreover we observe that the difference in supply between Saturdays and Sundays is larger than the difference observed between corresponding demand. Hence, Saturdays show the lowest passengers/runs ratio as well as the lowest average bus occupancy. In average, the bus occupancy (only considering smart card holders) varies between 11.4 (Fridays) and 13.2 (Wednesdays) during the weekdays. These figures are quite coherent with known patterns of travel during the week.

Table 5. Key facts regarding the network performance during the four weeks of February 2005 by day of travel

Figure 3. Average occupancy of buses per day (Feb. 2005 data)

4.4 Study of a specific route

The variability of some of the performance measures is examined for a specific route of the network. Route 37 is selected for the study since it is the most important route

that runs seven days a week (see Figure 4). A total of 2389 runs are enumerated during the month of February 2005.

Figure 4. Map of route 37

Schedule adherence on route 37 is examined using a special dataset extracted from the smart card data. Figure 5 shows the distribution of delays on this route based on 4 months of operation. This representation is inspired by the work of Bertini and El-Geneidy (2003). Perfect schedule adherence occurs for 17.5% of the observation (each observation represents the difference between scheduled and operated time at a specific stop of the route). 18.9% of them are before time (avg. 1.6 min) and 63.6% are late (avg. 3.0 min). Delays are more important in inbound direction because the peak hours are more concentrated in the morning for this area.

Figure 5. Schedule adherence on route 37 based on Feb. to Apr. 2005 observations

Data mining clustering techniques have been applied on route 37 data to characterize the runs operated. The load profile of each run of a 4-month period has been gathered in the dataset, for inbound direction. Figure 6 shows the typical load profile for the five clusters that have been identified.

Figure 6. Run load profile's clusters for route 37, inbound direction

With a further analysis of the temporal and daily distribution of the clusters, we can find that:

- The C1 run type user population (4,7% of runs, average load of 34,1) is mostly composed of Outaouais College students. The college is located at stop # 1. It is a typical weekday, scholar run.
- The C2 type is a typical mid-afternoon or week-end low profile run (34.4% of runs, average load of 15.8).

- The C3 type is a typical mid-morning or week-end low profile run (37.4% of runs, average load of 27.9).
- The C4 type is mostly a peak-hour, weekday morning run inbound downtown (37.8% of runs, average load of 80.1). The slight decrease at stop #24 is due to the alighting of the Casino workers.
- The C5 type regroups midday and evening runs (19.6% of runs, average load of 47.0).

The belonging to clusters is quite stable. For weekdays, a total of 53.9% of the runs are always in the same cluster (thus same type) all over the 4-month period, for a same departure time. Only about 16% of the runs can be classified in more than 2 clusters for this period.

5. Discussion and conclusion

This paper introduces the use of smart card automated fare collection system data to measure transit supply and demand indicators. Even though measures like vehicle-hours and passenger-kilometres are quite easy to conceptualize, they are very difficult to gather in any transit network on a day-to-day basis. With smart card data, the statistics can be easily compiled if the boarding and alighting locations are available. The planners can obtain these statistics on a daily basis, in order to evaluate the operated service and possibly to correct the forthcoming days. In this case study, data of the Société de transport de l'Outaouais have been used and completed with the Trepanier and Chapleau destination estimation algorithm. Currently, no dataset allows to assess the validity of the imputed alighting stop. The availability, in the short term, of complete travel patterns either from on-board surveys or Origin-Destination surveys will help refine the algorithm. Results show

that while most of the performance measures are stable over time, the approach allows identifying and classifying specific characteristics of the operation. The data mining techniques were especially useful to find different types of bus stops and operated runs.

However, the use of smart card data to obtain transit performance measures has some limits. First, this approach is very data-intensive, so there is a need to have a strong transit information system to support the analysis. In addition, data must be pre-validated because there are always a certain number of errors in smart card systems (missing data, wrong run number, equipment malfunction, etc.). Finally, let us remind that not all transit users have a smart card. In this specific case, the number of card holders is quite high for AM peak inbound routes, but is lower for week-ends and for routes that carry Ottawa users, which do not have smart cards. There is a need to compare and balance the statistics from on-board passenger counts in order to obtain the complete figure. Comparison between synthesised statistics from smart card data and operational data from the transit authority will help refine the presented techniques.

Further researches will focus on disaggregate analysis of both the supply and the demand. For the demand side, the analysis of individual user behaviour will provide additional information to the transit planners about the habits of users: departure time, preferred origins and destinations, preferred routes, etc. The better knowledge of the user behaviour is especially useful for marketing purposes (fares, advertising campaigns, etc.). For the supply side, there is a possibility to find ways of “optimizing” equipment use looking at operated data and run load information. A good knowledge of the variability of the demand could lead to weekly built schedules, instead of months-ahead planned schedules.

6. References

- Bagchi, M., White, P.R. (2005). The potential of public transport smart card data, *Transport Policy*, 12 (5), September 2005, pp. 464-474
- Bertini, Robert L., El-Geneidy, Ahmed (2003). Generating Transit Performance Measures with Archived Data, *Transportation Research Record* 1841, pp.109-119.
- Bonneau, W. and editors (2002). The role of smart cards in mass transit systems, *Card Technology Today*, June 2002, p.10.
- Caulfield, Brian, O'Mahony, Margaret (2004). Transit Capacity and Quality of Service Manual Applied to a Bus Corridor in Dublin, Ireland, *Transportation Research Record* 1887, pp.195-204.
- Clarke, R. (2001). Person location and person tracking: Technologies, risks and policy implications, *Information Technology & People*, 14 (2), 2001, pp. 206-231.
- CNIL –Commission nationale de l'informatique et des libertés (2003). Recommandation relative à la collecte et au traitement d'informations nominatives par les sociétés de transports collectifs dans le cadre d'applications billettiques, CNIL, Délibération N° 03-038.
- Conklin, Joana, English, Larry, Shammout, Khaled (2004). Transit Customer Response to Intelligent Transportation System Technologies Survey of Northern Virginia Transit Riders, *Transportation Research Record* 1887, pp.172-182
- Fielding, Gordon J., Gauthier, Roy E., Lave, Charles A. (1978). Performance indicators for transit management, *Transportation*, no.7, pp. 365-379.
- Gillen, David, Chang, Elva, Johnson, Doug (2001). Productivity Benefits and Cost Efficiencies from Intelligent Transportation System Applications to Public Transit.

Evaluation of Advanced Vehicle Location, Transportation Research Record 1747, pp.89-96.

Kittelsohn & Associates, Inc. (1999) Transit Capacity and Quality of Service Manual (1st ed.). TCRP Project A-15. TRB, National Research Council, Washington, D.C., 1999.

Morency, C., Trépanier, M., Agard, B. (2006). Analysing the Variability of Transit Users Behaviour with Smart Card Data, 9th International IEEE Conference on Intelligent Transportation Systems, Toronto, Canada.

Perk, Victoria A., Foreman, Chandra (2003). Florida Metropolitan Planning Organization Reports on Transit Capacity and Quality of Service. First-Year Evaluation, Transportation Research Record 1841, pp.128-134.

Shelfer, M., Procaccino, J.D. (2002). Smart card evolution, Communications of the ACM, July 2002/Vol. 45, No. 7, pp. 83-88.

Trépanier, M., Chapleau, R. (2006). Destination Estimation From Public Transport Smartcard Data, 12th IFAC Symposium on Information Control Problems in Manufacturing (INCOM), Saint-Étienne, France.

Trépanier, Martin, Chapleau, Robert, Tranchant, Nicolas (2007). Individual Trip Destination Estimation in Transit Smart Card Automated Fare Collection System, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Taylor & Francis, vol. 11, no. 1, pages 1-15.

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, Congrès annuel 2004 de l'Association des transports du Canada (Québec), pp. 10-14.

7. List of tables

Table 1. Information on smart card transaction in the dataset	8
Table 2. Key facts regarding transit supply during the four weeks of February 2005 by day of travel.....	13
Table 3. Key facts regarding transit demand during the four weeks of February 2005 by day of travel.....	13
Table 4. Key facts regarding transit demand by card type during the four weeks of February 2005	14
Table 5. Key facts regarding the network performance during the four weeks of February 2005 by day of travel.....	15

Table 1. Information on smart card transaction in the dataset

CARD ID	The card number is not related to an individual, and issued only for cross-relating records.
FARE (CARD) TYPE	The fare type related to the card (regular, student, express, etc.).
DATETIME	The date and time of the card reading aboard the bus.
ROUTE & DIRECTION ID	Route number and direction on which the bus is operating.
STOP ID	Stop number of the boarding, obtained with the help of a GPS device aboard the vehicle.
VALIDATION RESULT	Indicates whether the boarding was valid, was a transfer or was refused by the reader.
OPERATIONAL INFO	Information on run number, vehicle number and bus driver is also available.

Table 2. Key facts regarding transit supply during the four weeks of February 2005
by day of travel

Supply	% runs	Veh-km	Veh-hre	Avg Speed (km/h)	Avg LenRun (km)	Avg DurRun (min)
Monday	16.93%	17.07%	17.28%	21.13	15.53	44.10
Tuesday	17.48%	17.48%	17.70%	21.13	15.40	43.74
Wednesday	17.52%	17.51%	17.45%	21.47	15.40	43.04
Thursday	17.69%	17.69%	17.80%	21.27	15.41	43.46
Friday	17.46%	17.55%	17.60%	21.33	15.48	43.55
Saturday	8.03%	7.72%	7.47%	22.12	14.82	40.19
Sunday	4.90%	4.97%	4.71%	22.57	15.61	41.50
TOTAL February	32,200	495,400	23,200	21.39	15.40	43.20

Table 3. Key facts regarding transit demand during the four weeks of February 2005
by day of travel

Demand	% Passengers	% Pass-km	% Pass-hre	Avg Speed (km/h)	Avg LenTrip (km)	Avg DurTrip (min)
Monday	16.88%	17.65%	17.83%	20.99	9.23	26.39
Tuesday	19.17%	19.85%	19.90%	21.15	9.14	25.93
Wednesday	19.53%	19.93%	19.73%	21.42	9.01	25.24
Thursday	19.39%	19.58%	19.65%	21.13	8.92	25.31
Friday	17.40%	17.30%	17.18%	21.36	8.78	24.66
Saturday	4.36%	3.20%	3.21%	21.12	6.47	18.39
Sunday	3.28%	2.49%	2.49%	21.23	6.71	18.95
TOTAL February	649,700	5,736,400	270,500	21.21	8.83	24.98

Table 4. Key facts regarding transit demand by card type during the four weeks of February 2005

Demand	% Passengers	% Pass-km	% Pass-hre	Avg Speed (km/h)	Avg LenTrip (km)	Avg DurTrip (min)
Adult - Interzone ¹	3.00%	11.60%	7.16%	34.35	34.13	59.62
Adult - Express ²	14.71%	21.86%	20.51%	22.61	13.12	34.83
Adult - Regular ³	48.37%	41.49%	45.47%	19.35	7.57	23.48
Elderly	3.34%	2.02%	2.24%	19.16	5.34	16.72
Student	30.48%	22.93%	24.52%	19.84	6.64	20.09
STO employee	0.10%	0.09%	0.10%	19.76	7.69	23.34
TOTAL February	649,700	5,736,400	270,500	21.21	8.83	24.98

¹ Unlimited access to regular, express and interzone network

² Unlimited access to regular and express, supplement (\$) on interzone network

³ Unlimited access to regular network, supplement on express and interzone network

Table 5. Key facts regarding the network performance during the four weeks of February 2005 by day of travel

Performance	Passengers / runs	Avg Bus occupancy
Monday	20.14	11.97
Tuesday	22.15	13.15
Wednesday	22.52	13.17
Thursday	22.15	12.82
Friday	20.13	11.42
Saturday	10.98	4.80
Sunday	13.50	5.80
TOTAL February	20.20	11.58

8. List of illustrations

Figure 1. Temporal distribution of boardings for the stops' clusters (Feb. 2005 data)	14
.....	
Figure 2: Spatial distribution of stops by stop cluster, Feb. 2005 data	15
Figure 3 : Average occupancy of buses per day (Feb. 2005 data)	15
Figure 4. Map of route 37	16
Figure 5. Schedule adherence on route 37 based on Feb. to Apr. 2005 observations	16
.....	
Figure 6. Run load profile's clusters for route 37, inbound direction	16

Figure 1. Temporal distribution of boardings for the stops' clusters (Feb. 2005 data)

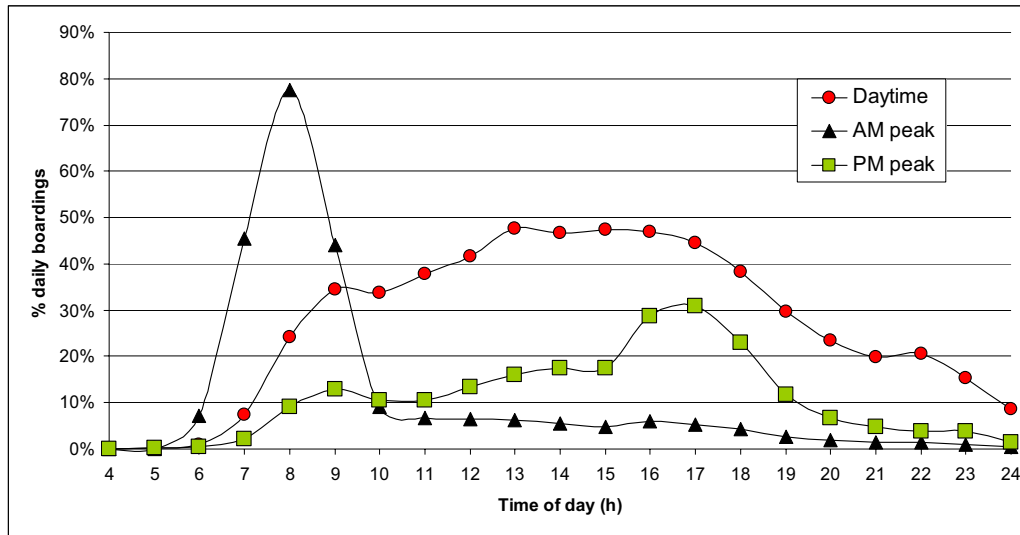


Figure 2. Spatial distribution of stops by stop cluster, Feb. 2005 data

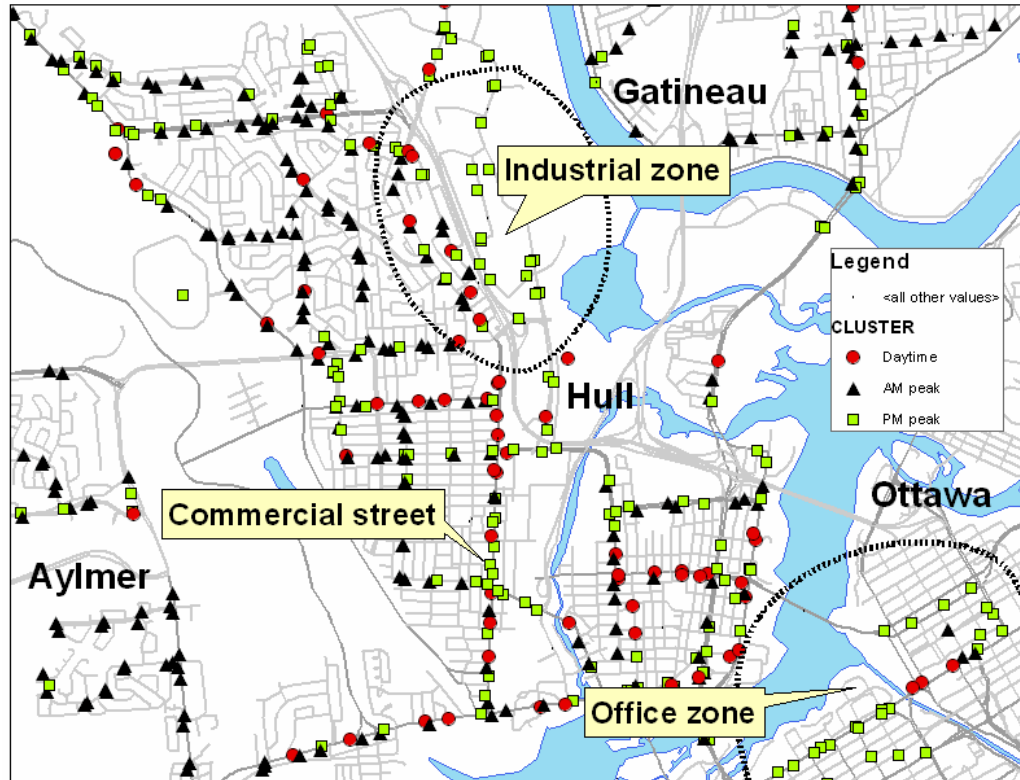


Figure 3. Average occupancy of buses per day (Feb. 2005 data)

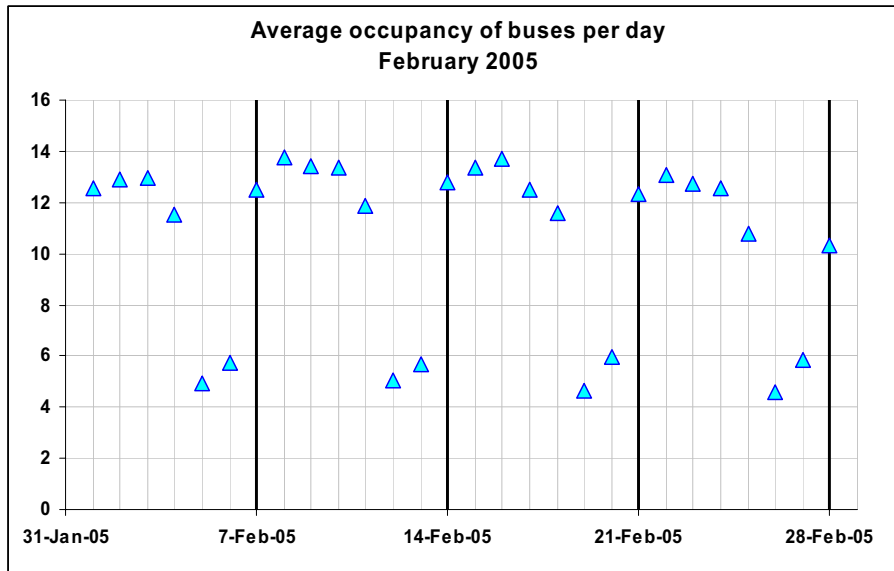


Figure 4. Map of route 37

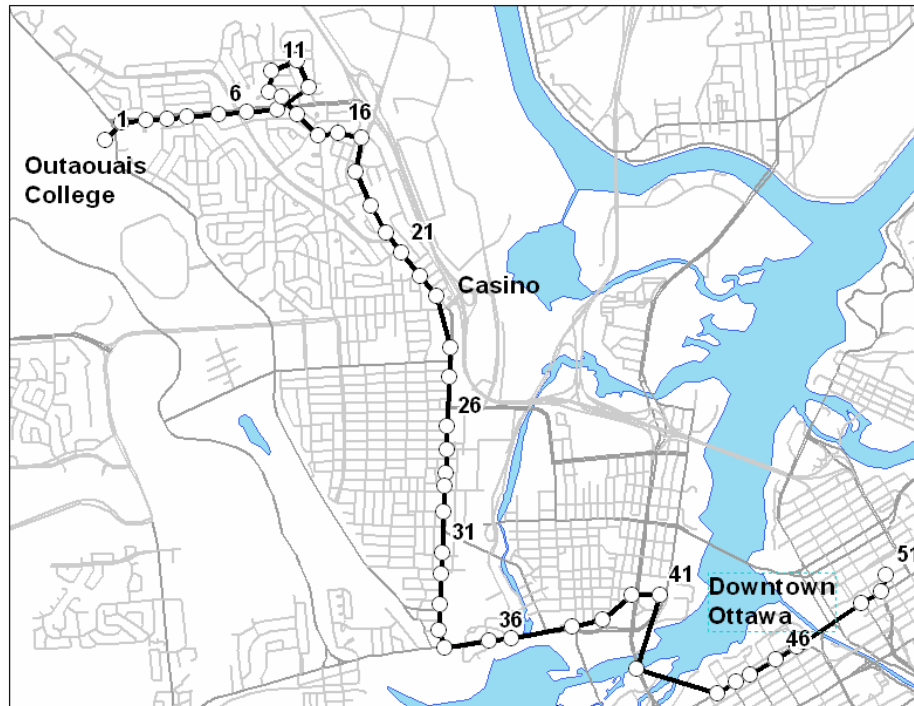


Figure 5. Schedule adherence on route 37 based on Feb. to Apr. 2005 observations

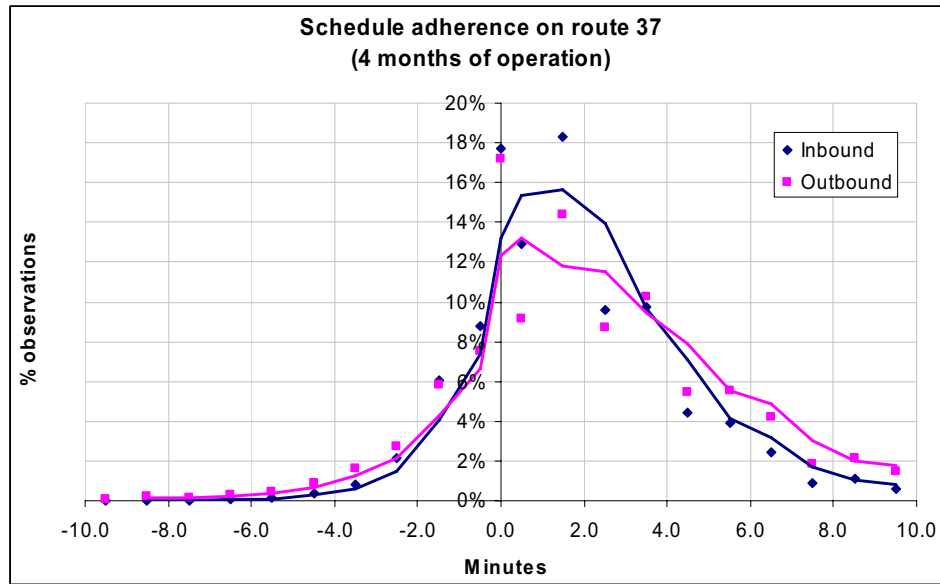


Figure 6. Run load profile's clusters for route 37, inbound direction

