Conference on Advanced Systems for Public Transit – CASPT12, The Ritz-Carlton, Santiago, Chile, July 23-27, 2012.

Using smart card data to assess the impacts of weather on public transport user behavior

Martin TRÉPANIER*[1,3,4], P.Eng. Ph.D., Professor (mtrepanier@polymtl.ca)

École Polytechnique de Montréal, Department of mathematical and industrial engineering

P.O. box 6079, station Centre-Ville, Montreal, Quebec, Canada H3C 3A7.

Phone: +1 (514) 340-4711 #4911, Fax: +1 (514) 340-4173

Bruno AGARD[1,3,4], Ph.D., Associate Professor (bagard@polymtl.ca)

École Polytechnique de Montréal, Department of mathematical and industrial engineering

P.O. box 6079, station Centre-Ville, Montreal, Quebec, Canada H3C 3A7.

Phone: +1 (514) 340-4711 #4914, Fax: +1 (514) 340-4173

Catherine MORENCY[2,3,4], P.Eng. Ph.D., Associate Professor (cmorency@polymtl.ca)

École Polytechnique de Montréal, Department of civil, geological and mining engineering

P.O. box 6079, station Centre-Ville, Montreal, Quebec, Canada H3C 3A7.

Phone: +1 (514) 340-4711 #4502, Fax: +1 (514) 340-3981

Élodie DESCOIMPS[1], M.A.Sc. (elodie.descoimps@polymtl.ca)

École Polytechnique de Montréal, Department of mathematical and industrial engineering

P.O. box 6079, station Centre-Ville, Montreal, Quebec, Canada H3C 3A7.

Phone: +1 (514) 340-4711, Fax: +1 (514) 340-4173

Jean-Sébastien MARCOTTE[2,3], Masters student (jean-sebastien.marcotte@polymtl.ca)

École Polytechnique de Montréal, Department of civil, geological and mining engineering

P.O. box 6079, station Centre-Ville, Montreal, Quebec, Canada H3C 3A7.

Phone: +1 (514) 340-4711, Fax: +1 (514) 340-3981

1- *École Polytechnique de Montréal, Department of Mathematics and Industrial Engineering, P.O. Box 6079, station Centre-Ville, Montreal, Quebec, Canada, H3C 3A7*

2- *École Polytechnique de Montréal, Department of Civil, Geological and Mining Engineering*

3- *Mobilité Chair in Sustainable Transportation*

4- *Interuniversitary Research Centre on Entreprise Networks, Logistics and Transportation (CIRRELT)*

Final Manuscript submitted to the Conference on Advanced Systems for Public Transit

CASPT12, 23-27 July 2012, The Ritz-Carlton, Santiago, Chile

Organized by Department of Transport Engineering and Logistics

of the Pontificia Universidad Católica de Chile

Keywords: smart card, public transport, weather, user behaviour, transit ridership

Abstract

Public transit systems operation is based on regular schedules established on daily basis: usually about the same schedules during weekdays and specific ones for Saturdays and Sundays. This facilitates the service operation (drivers, buses, assignments) and provides a constant service to the users. However, public transit ridership is not stable within these periods. A series of events will cause variations during weekdays. For example, holidays are known to be low ridership periods; that is why special schedules are assigned to these days. There are also seasonal variations (winter vs. summer), usually taken into account in service planning. Bad weather conditions are known to affect transit ridership, but few studies have been conducted to find at what extent and which users will be most influenced. This kind of study arises two methodological challenges: first, there is a need to have extensive and longitudinal data on ridership; second, it must be possible to isolate the weather effect from other causes of variation. In this project, we try to address both issues with the help of continuous smart card data from Gatineau and Montréal (Québec) public transit networks. Results shows that weather will influence mostly travelers doing non-constrained trips (like for seniors in Gatineau) and will cause a modal shift from bus to subway network in Montreal.

Introduction

Public transport operators know that adverse weather conditions may affect the use of their network. For example, a heavy rain day could make people stay at home, resulting in a decrease in public transport use. In another case, a big snowstorm in a large metropolitan area could encourage users to take subway instead of bus lines, because buses will run slowly through the streets of the city. In all cases, it remains quite hard for the public operators to correctly estimate the impacts on the network. Traditionally, it did not matter too much because the schedules were fixed during weekdays. With the advent of intelligent information technologies that can provide information to the users in real time (mobile phones, interactive messaging, ATIS), the operators could be interested to adjust their service according to the present –or forecasted— weather conditions.

However, before being able to adjust the schedule, there is a need to thoroughly evaluate the impacts of weather conditions on the use of service. We should be able to measure the use of the network for each type of user. Automatic passenger counting systems (APC) could provide useful information, but the information on the type of users not is available. Smart card automated fare collection systems are ideal for this task, because they can provide continuous data on the use of the network related to specific card identification and type. Hence, this permit quite detailed analysis on the use of specific routes, vehicles and other infrastructure of the network.

This paper aims to demonstrate the ability of smart card systems to evaluate the influence of weather events on public transport use. It presents a methodology based on data processing methods, statistical tests and event database integration to better assess the impact of weather events like rain, snow and air temperature on ridership. The approach has been tested on datasets from Gatineau and Montreal (Canada).

The paper first presents some background on weather influence and the use of smart card data in public transport planning. The next section contains further details on the information system used for the study, gathering two datasets on smart card plus some data on weather conditions. The overall methodology is then exposed, followed by the results. First, descriptive statistics are presented on ridership (globally, by fare type and by transit mode). Second, a data mining technique is used to create clusters of transit users, and then the belonging of the users to the groups is examined to foresee if bad weather conditions influence their general behavior. Finally, regression models are developed to measure the impact of weather events (and other elements as well) on transit ridership for different categories of users.

The conclusion of the paper discusses some limitations of the approach and further perspectives that should be addressed by the research community.

Background

This section presents research works related to the use of smart card data for public transportation planning. In addition, it reports some findings on the influence of weather conditions on transport.

Smart card

Smart card automated fare collection systems generate a huge quantity of data on the transactions made aboard transit vehicles. Since data is referenced in time and sometimes precisely in space, it can be used to estimate the ridership across the network. This type of data has multiple applications, as reviewed by Pelletier et al. (2011). First, data can be used to better understand user behaviors. Bagchi & White (2005) used Bradford, UK data to produce indicators like the number of daily trips per card, the turnover rate (card stopped being used compared to new cards) and the number of linked trips (trips using more than one mode in a sequence). Smart card data has also been used to analyze subway ridership and origin-destination matrix, by estimating the egress station of subway users (Munizaga et al. 2010). The examination of transactions over time for single fare card lead to a better understanding of loyalty patterns among users, accordingly to the type of fare (Trépanier & Morency 2010). This kind of data can also be used to rearrange schedules (Utsunomiya et al., 2006).

Individual behaviors of users can be analyzed as well with smart card. Agard et al. (2006) used data mining techniques to classify the card users in four groups showing similar behaviors, and then measured the belonging to the users to each group along the weeks. It shows that there could be a high variation in the behaviors from one week to another. Lee & Hickman (2011) have proposed a method to study travel pattern of transit users with the help of smart card data.

Influence of weather

There are many studies reporting on the influence of weather conditions on transportation. However, very few of them are related to public transit. Handman (2004) classified the weather events that have impacts on transportation: ice storms, snow and hail, heavy rain, thunderstorms, extreme temperature, fog and strong winds. Different approaches were taken in the literature to measure the influence of weather condition. The binary method determines whether an event has occurred or not (Changnon, 1996). Some authors used a threshold-based method, where intervals are used to evaluate the intensity of weather events (Keay & Simmonds 2005). Finally, some authors based their studies on extreme conditions only, like Knapp et Smithson (2000) who examined the impact of 64 snow falls on road networks. These works confirm that studies on weather impacts require a huge quantity of longitudinal data.

Information system

This project uses two types of datasets. First, smart card data has been extracted from Gatineau and Montreal contactless smart card public transit collection systems. Second, weather conditions were collected from Environment Canada, plus calendar events.

Gatineau

Gatineau smart card dataset contains 53.0 million transactions covering a 6-year period of operation of the *Société de transport de l'Outaouais* (STO), a 200-bus network in Gatineau, Quebec (200,000 inhabitants). Figure 1 presents the object-model associated to this database. The model is based on the Transportation Object-Oriented Modeling approach (Trépanier and Chapleau 2001).

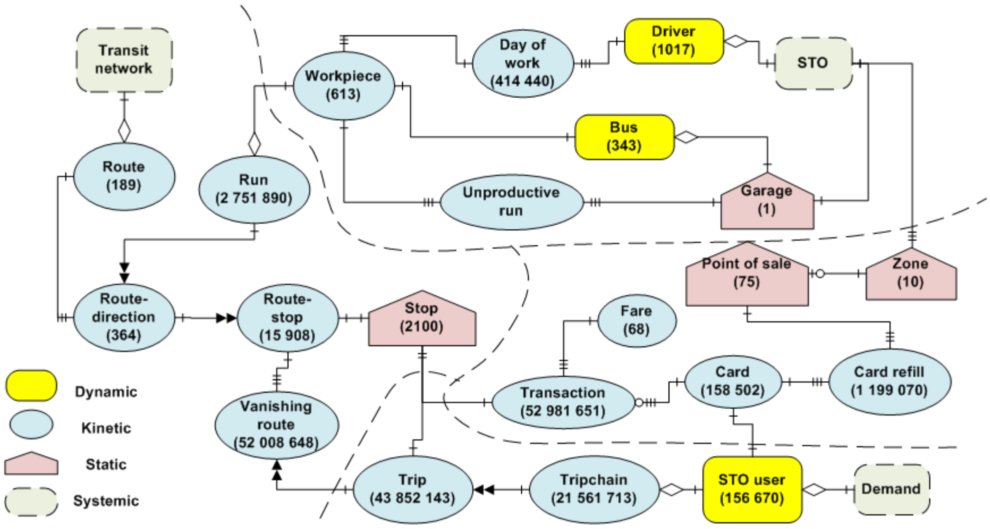

Figure 1: Object model of the STO smart card system (no. of instances shown for period from Jan. 1st 2004 to Nov. 20th 2009)

The dataset used in this study is essentially related to the transactions. Each transaction record contains the date and time, the bus stop, the route and direction, the card number, and the type of fare (adult, student, college and university, senior).

Montreal

The *Société de transport de Montréal* (STM) operates a bus network of 1600 vehicles (regular and articulated) and a completely-underground 68-km subway network (759 cars, 68 stations). The population of the Montreal Island is nearly 1.8 million, and hundreds of thousand users of the 2 million inhabitants suburbs also commute with the STM. The dataset used in this study covers a 7 month period between October 1st 2010 and April 30th 2011. Figure 2 presents the object model of the STM smart card system. Only the number of instances available directly from the dataset is shown.

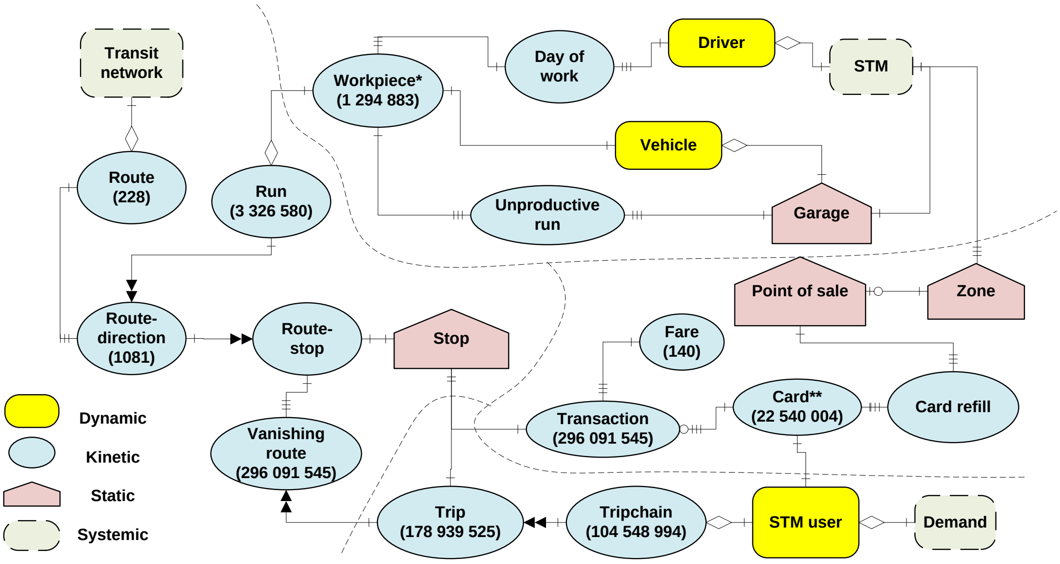

Figure 2: Object model of the STM smart card system (no. of instances shown for period from Oct. 1st 2011 to Apr. 30th 2011) (* bus operations only, ** including magnetic tickets)

The Montreal dataset is similar to the one from the STO, expect that boarding location are not available, except for subway (however, whereas direction is not known). However, the magnitudes of the network are quite different.

Weather and other events

Weather datasets were extracted from the National Climate Data and Information Archive from Environment Canada. For each day, the following data was retained: minimum, maximum and average air temperatures; snow, rain and total precipitations; depth of snow on the ground. In addition, we identified calendar events such as Christmas and Easter Holidays, specific Canadian Holidays (Thanksgiving, Quebec and Canada National days, etc.). The students Holidays were also marked.

Methodology

A three-part methodology is proposed to assess the impact of weather on transit use: 1) descriptive results, 2) data mining and 3) regression models.

A descriptive analysis, including charts, table and simple statistics, is needed to better understand the data and to foresee the major impacts of weather. Daily ridership is examined according to global figures, fare type, and transit mode, to detect disruptions brought by weather events. However, because many other factors may affect transit ridership, seasonal and cycle effects must be taken into account. For the Gatineau network, this approach is interesting because the travel behavior of users is quite consistent: most of the users commute to work or studies. For Montreal, it is more difficult to see the impacts because the behavior is more heterogeneous and counter effects will occur.

The use of data mining techniques has brought interesting results when examining transit user behavior from smart card data (Morency et al. 2007). In this case, a k-means clustering technique is use to create user groups that have about the same daily transaction time pattern. Then, the most dominant group is found for each user and for each weekday. For example, the dominant group for a Monday could be different from the one for Friday. Then, the amount of users that change behavior is monitored for each day of the period. In this study, the method has been applied to Montreal data only.

In the third part of the methodology, we used regression models to measure the influence of weather events (and other events as well) on transit ridership, according to specific fare type, modes, period of day, etc.

Descriptive statistics

This section presents descriptive statistics about daily ridership variation, ridership variations by fare type and transit mode shift possibly related to bad weather conditions.

Daily ridership

The first focus of the analysis is daily ridership. Figure 2 is a chart of the total smart card ridership at the STO for the whole period, expressed in terms of the number of daily transactions. First, we see the huge difference between weekdays and weekends (the latter are seen at the bottom of the chart). Next, there is clearly a seasonal cycle, recurring each year, showing an important decrease of ridership during summer. In addition, the Christmas holiday's period is also marked by an important decrease. Finally, we can see a constant increase of ridership from year to year. This emphasizes on the need to take these variations into account before doing any weather impact analysis.

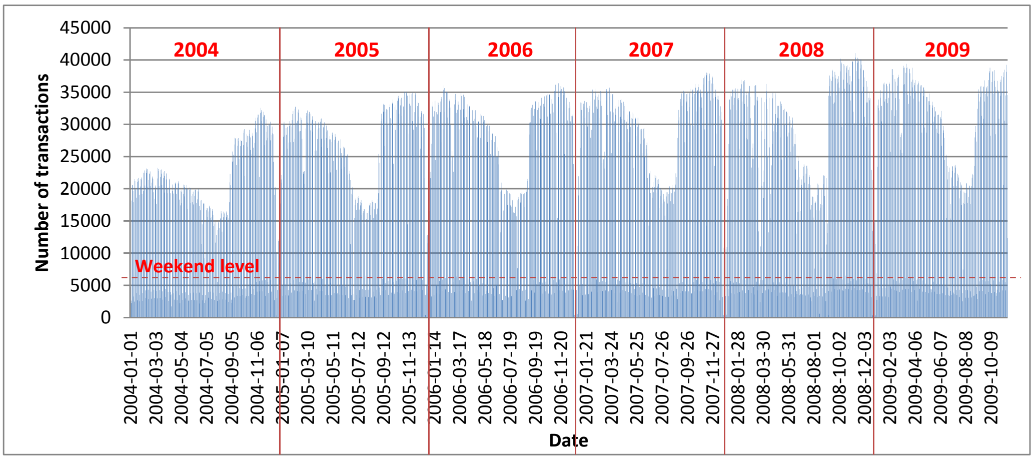

Figure 3: Daily ridership at the STO for the whole period (Jan. 1st 2004 to Nov. 20th 2009)

Hence, it is quite hard to see the effect of weather on total ridership, because other factors can influence the use of public transit. Figure 4 presents the daily ridership at the STM for the first four month of 2011. We can see a slight decrease of transit use for the following extreme weather conditions (Montreal context, of course): minimum temperatures of -25°C and -28°C on January 17th and 24th, a 20 cm snowfall on February 2nd, and a 14 cm snowfall on March 7th. However, the spring break and the Easter weekends cause a much larger decrease in ridership. We also experiment two slight decreases on March 16th and April 20th, while no particular weather event was registered on these days. At this point, only one winter period of data is available, so we cannot compare these figures with last years and validate if these differences are significant.

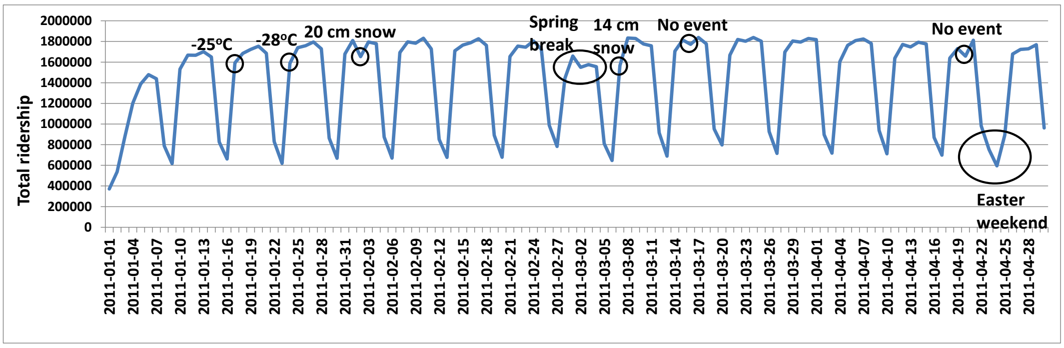

Figure 4: Daily ridership at the STM for the period from Jan 1st to April 30th 2011

Fare type

As expected, weather events will not affect every user at the same level. Figure 3 shows in parallel the senior ridership of STO for January 2007, the outdoor temperature, the rainfalls and the snowfalls. We can see that for low temperature days (e.g. Jan. 17th and 26th), ridership is lower than for similar weekdays (Jan. 10th and Jan 19th). On Jan. 15th, a 10-cm snowfall has also caused a decrease in ridership.

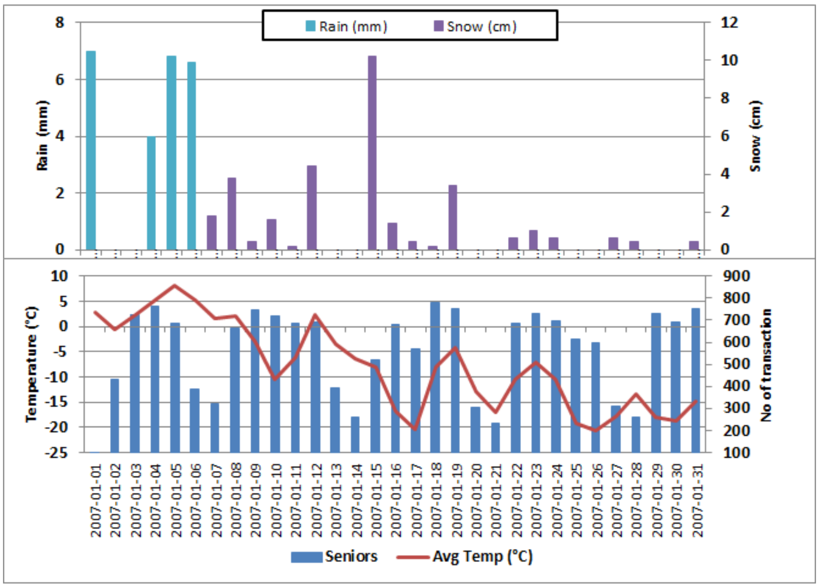

Figure 5: Senior ridership and weather conditions at STO, January 2007

We could expect that the importance of the impacts of weather on transit ridership is related to the importance of the weather event. Figure 4 shows the impact classified by categories of snowfall. This is deseasonalized data; this means that the effects on annual ridership evolution, summer and Holidays have been removed. At left, we can see the weekday and weekend ridership for adults. It shows that for adults, snowfalls have little impact on ridership (the average daily deseasonalized ridership is 24 000 for this categories). The increase for 9-10 cm is not significant (poor sampling). However, for senior, the ridership decreases accordingly to the importance of snowfall. In weekdays, for snowfalls over 10 cm, the ridership is 27% lower than the average.

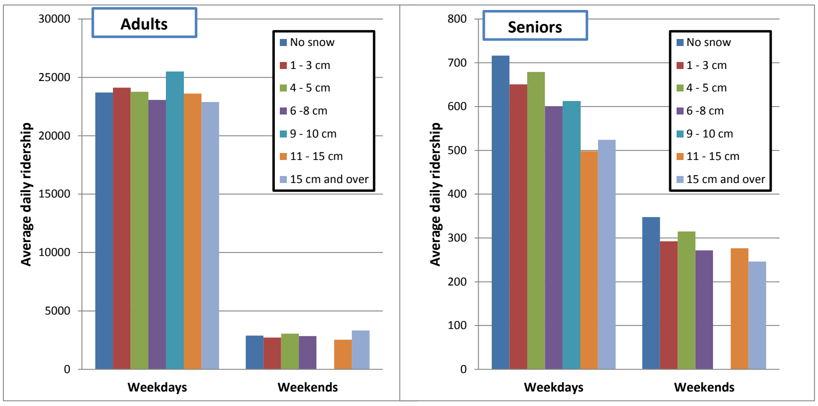

Figure 6: Impact of snowfalls by categories for adults and seniors (STO network, Jan. 1st 2004 to Nov. 20th 2009)

In large cities like Montreal, one could expect that during adverse weather conditions, some commuters would rather use public transit instead of their personal car, because traffic conditions is deteriorating due to poor road surface and the presence of snow banks. In Montreal, while most of the users have a monthly pass on their smart card, about 22% of them will use tickets (registered in smart card or cardboard base). Cash payments are also registered as one-way tickets in the fare collection system.

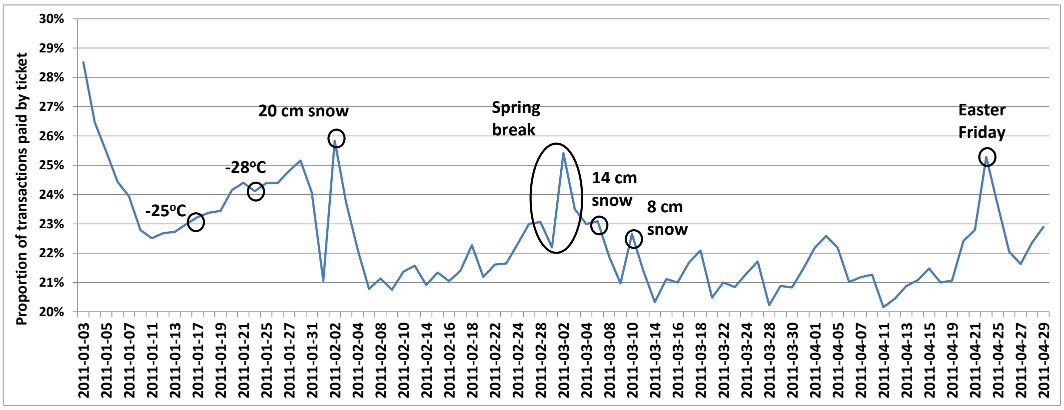

Figure 7: Proportion of transactions paid by ticket at the STM from Jan 1st to April 30th 2011 (weekdays only)

Figure 7 shows the proportion of transactions paid by tickets at the STM during weekdays between Jan. 1st and April 30th 2011. It shows that most of the time, the proportion of ticket use is related to other elements than weather. For instance, the beginning of the year and Easter Holiday cause an increase of this proportion. However, a peak is registered on Feb. 2nd, where 20 cm of snow fell on Montreal. But there is also a peak on March 2nd, where no particular weather event was recorded. No other adverse weather condition seems to have influence the proportion of trips paid by tickets during this period.

Transit mode shift

In multimodal networks, bad weather conditions may cause a transit mode shift. During large snowfalls, the bus service is perturbed because the traffic conditions deteriorate. On colder days, users will be less likely to wait at bus stops. For both reasons, the ridership of the subway network, completely underground and weather-protected in Montreal, may increase.

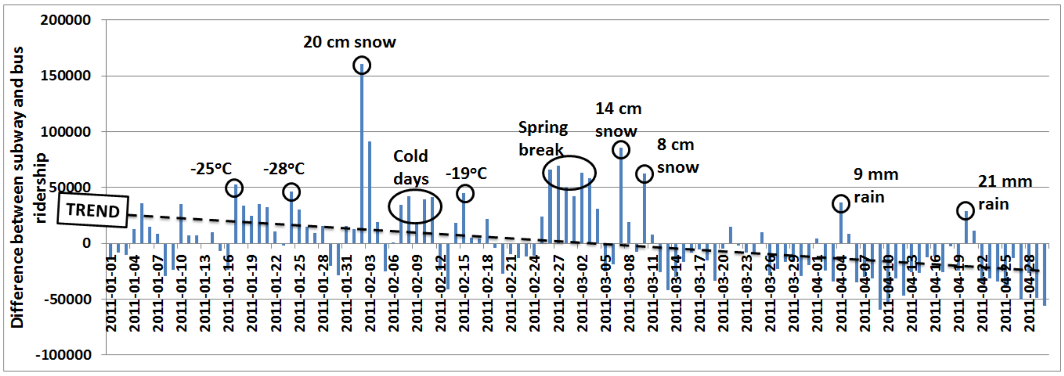

Figure 8: Difference between subway and bus ridership at the STM for the period from Jan 1st to April 30th 2011 (all fares, all days)

Figure 8 shows the difference between ridership for bus and subway networks for the study period. Firstly, what we see with the general trend is that in winter, the use of the subway network is higher. Secondly, we see that there is an increase in subway ridership in large snowfall days. Heavy rains of April also cause a mode shift to subway, counterbalancing the general trend. In cold days, there are also more people in the subway. In Spring break, there is a decline in bus ridership due to the absence of students (major users of the surface network). This descriptive statistics is one of the best ways to show the impacts of weather on Montreal network ridership.

Data mining

This section reports on data mining experiments made with Montreal data.

Clustering

Looking at global ridership may help us to assess the overall effects on the transit system, but it is interesting to look at these effects at an individual level. For this study, a sample of 1 million cards and cardboard tickets was used to create groups of behaviors using a k-means technique applied to the hourly distribution of transactions within the days. The base unit to create the groups was the observation of a single card for each day where it was used (a "card-day"), during a 7-month period (from Oct. 2010 to Apr. 2011). For example, if a card was used for 30 days during the period, 30 observations are available to the analysis. A total of 2.82 million observations were used to create the groups. Seven groups were discriminated from the analysis; their hourly distribution pattern is shown at Figure 9 (showing the average distribution provided by the k-means method). We see that most of the groups are characterized by peak hour dominance. This means for example that a card that has made transaction at 17:00 and sometimes 18:00 PM and 8:00 AM will be associated to the G2 group. Group 6 is related to an evening transactions pattern, while group 7 is more dispersed through the day.

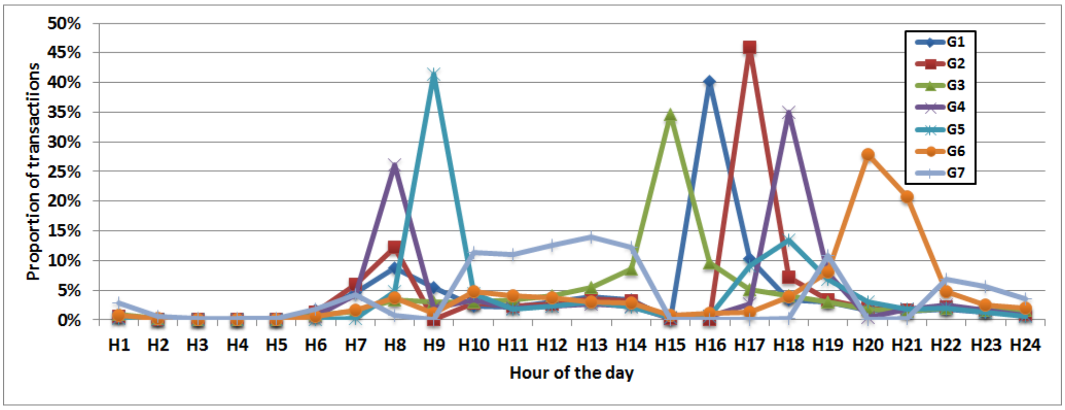

Figure 9: Hourly pattern distribution of transactions for k-means groups, STM data

The number of "card-days" in each group is well balanced. Figure 10 shows the distribution of groups accordingly to the 2.8 million "card-days". Apart from group 6 (evening transactions) and group 7 (mixed patterns), the proportion of other groups is quite the same. The proportion of groups 6 and 7 is higher for weekends, while the distribution of groups is similar between workdays (Mon. to Fri.). It is noticeable that transit users do not always keep the same behavior from a week to another. During the 7-month period, on Monday, monthly pass holders belong to 3.48 different groups in average (including the dominant one). On Friday, this figure rises to 3.75.

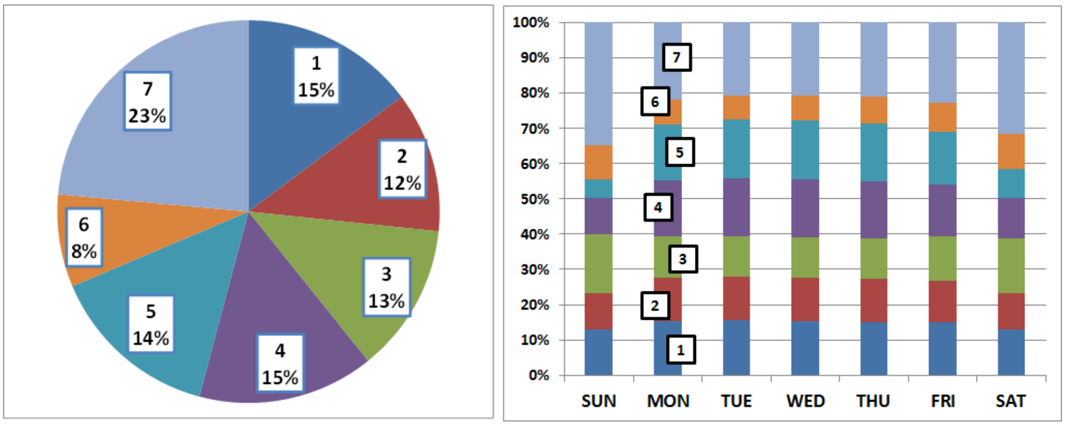

Figure 10: Distribution of groups; globally (left), by weekday (right), for 2.82 million card-days

Dominant behavior

Let us now examine the proportion of use of the dominant behavior group by monthly pass holders during the period. Figure 11 shows that for the beginning of the year, less than half the users will follow their most dominant pattern. After that, the proportion varies between 58% and 64%. For each week, the proportion decreases from Monday to Thursday, and drops significantly each Friday, where users seem to have a more diverse set of behaviors. Hence, this analysis reveals little information on the impact of weather events, except for the big snow fall of Feb. 2^{nd} (and the next day), where there is a decrease of the proportion of users that follow their dominant behavior. We also see a disruption around the spring break period.

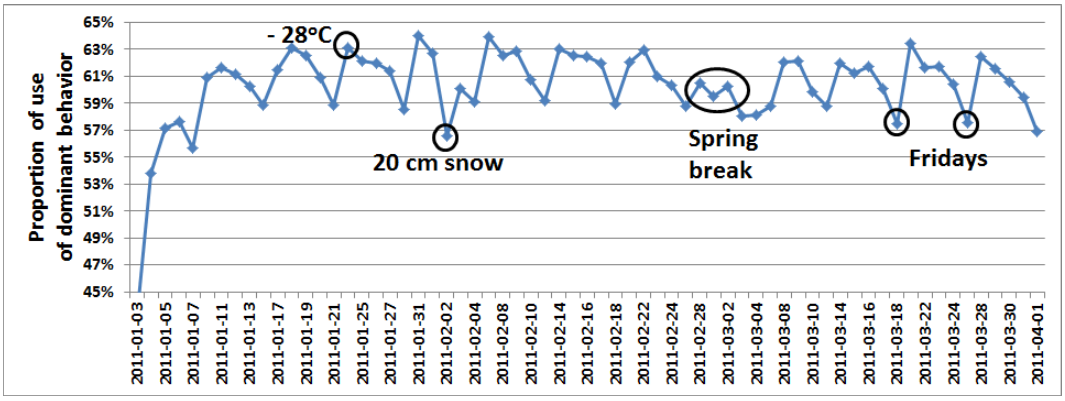

Figure 11: Proportion of users that follow their dominant behavior at the STM for the period from Jan 1st to March 31st 2011 (monthly pass holders, weekdays only)

Models

In this section, we present the development of multivariate regression linear models aimed to measure the influence of different contextual elements on ridership.

Variables

The number of daily transactions was first used as independent variables (Y) for both cases. Thus, there are 2151 observations for Gatineau and 151 for Montreal. Table 1 presents the set of independent variables tested for the Montreal models. A different model was developed to explain each ridership figure. Ratio variable were also tested to take into account the relative importance of one factor over another (e.g. peak periods, use of subway, proportion of pass holders).

Table 1: Independent variables tested for the Montreal case study

Independent variables (Y) – Montreal case				
Name	Description	Min	Max	Average
Y_TR_TOT	Total no. of transactions (TTNT)	336491	1838348	1387024
Y_TR_AM	TTNT AM peak (6 to 9)	18761	446360	283608
Y_TR_PM	TTNT PM peak (16 to 19)	87930	532210	390622
Y_TR_BUS	TTNT bus network	179528	972496	720175
Y_TR_PASS	TTNT pass holders	160417	920910	681360
Y_TR_TICK	TTNT ticket holders	102353	478964	326906
Y_TR_REDU	TTNT reduced fare	73721	535671	378758
Y_TR_METRO	TTNT subway network	154714	918004	664335
Y_TR_AIR	TTNT airport shuttle	1659	3659	2514
Y_RAT_AM	Ratio TTNT AM / total	0,04	0,25	0,18
Y_RAT_PM	Ratio TTNT PM / total	0,23	0,30	0,28
Y_RAT_OFFP	Ratio TTNT off-peak / total	0,44	0,71	0,54
Y_RAT_PASS_TICK	Ratio TTNT pass / TTNT ticket	1,22	2,53	2,04
Y_RAT_BUS_METRO	Ratio TTNT bus / TTNT subway	0,84	1,75	1,09

For Gatineau (Table 2), there is more emphasis on fare type because we can identify young students, college and university students, adults and seniors, which is not the case in Montreal. On some Holidays, the ridership is quite low due to the reduced bus network provided.

Table 2: Independent variables tested for the Gatineau case study

Independent variables (Y) – Gatineau case				
Name	Description	Min	Max	Average
Y_TR_TOT	Total no. of transactions (TTNT)	307	48122	24015
Y_TR_ADULTS	TTNT adults	221	25691	14267
Y_TR_YOUNG	TTNT children and students	0	13526	5534
Y_TR_COLL_UNIV	TTNT college and university	11	6134	1272
Y_TR_SENIOR	TTNT senior fare	13	1509	890
Y_TR_YEARLY	TTNT yearly subscribers	9	8759	3546
Y_TR_AM	TTNT AM peak (6 to 9)	0	17825	8290
Y_TR_PM	TTNT PM peak (16 to 19)	116	17629	8503
Y_TR_EXPR	TTNT express fare	3	7387	3575
Y_RAT_SUBS_ALL	Ratio TTNT subscribers / total	0,02	0,29	0,13

Y_RAT_SEN_ALL	Ratio TTNT senior / total	0,01	0,17	0,05
Y_RAT_YNG_ALL	Ratio TTNT students / total	0,00	0,38	0,22
Y_RAT_COLL_ALL	Ratio TTNT college / total	0,01	0,18	0,05
Y_RAT_EXP_ALL	Ratio TTNT express / total	0,01	0,23	0,12
Y_RAT_AM	Ratio TTNT AM / total	0,00	0,55	0,27
Y_RAT_PM	Ratio TTNT PM / total	0,10	0,54	0,33
Y_RAT_OFFP	Ratio TTNT off-peak / total	0,07	0,71	0,40

The dependent variables are chosen to take into account the seasonal effects (yearly, monthly and daily), the holidays (single, Christmas period and study breaks) and the weather conditions (temperature, snow fall and rain fall). For the weather conditions, four types of variables are tested into separate models: continuous (temperature in Celsius, rain fall in millimeters, snow fall in centimeters), classified (integer 0 to 6 accordingly to the importance of weather event), 3-class dummies (e.g. no snow, light snow, heavy snow) and 6-class dummies.

Table 3: Dependent variables used in both case studies

Dependent variables (X) – continuous				
Variable	Description	Min	Max	Average
TMAX	Maximum temperature of the day	-16,9	23,6	0,38
TMIN	Minimum temperature of the day	-27,9	8,7	-7,63
TAVG	Average temperature of the day	-22,4	16,2	-3,63
HDD	Heating Degree-days	1,8	40,4	21,63
RAIN_MM	Rainfall (mm)	0	26,6	1,85
SNOW_CM	Snowfall (cm)	0	20,4	1,52
TOTPRE_MM	Total precipitations (mm)	0	33,8	3,34
GR_SNOW_CM	Total snow on ground (cm)	0	32	7,72
Dependent variables (X) – classes				
Variable	Description	Min	Max	Average
COLD	Coldness (6 classes)	0	5	2,96
SNOW	Snowfall (6 classes)	0	5	0,62
RAIN	Rainfall (6 classes)	0	5	0,47
Dependent variables (X) – dummy				
Variable	Description	Min	Max	No. = 1
CD_NO	Min. Temp. over 0 ^{o}C	0	1	27
CD_0_MINUS_5	Temp. between -5 ^{o}C and 0 ^{o}C	0	1	32
CD_ MINUS_5_10	Temp. between -5 ^{o}C and -10 ^{o}C	0	1	33
CD_ MINUS_10_15	Temp. between -10 ^{o}C and -15 ^{o}C	0	1	31
CD_ MINUS_15_20	Temp. between -15 ^{o}C and -20 ^{o}C	0	1	23
CD_ MINUS_20_MORE	Min. Temp. below -20 ^{o}C	0	1	5
SN_NO	No snowfall	0	1	80
SN _0_5	Snow between 0 cm and 5 cm	0	1	56
SN _5_10	Snow between 5 cm and 10 cm	0	1	9
SN _10_15	Snow between 10 cm and 15 cm	0	1	5
SN _15_20	Snow between 15 cm and 20 cm	0	0	0
SN _20_MORE	Snow over 20 cm	0	1	1
SN_LIGHT	Light snow (below 10 cm)	0	1	65
SN_HEAVY	Heavy snow (over 10 cm)	0	1	6
RA_NO	No rainfall	0	1	120
RA_0_5	Rain between 0 mm and 5 mm	0	1	12
RA_5_10	Rain between 5 mm and 10 mm	0	1	10
RA_10_15	Rain between 10 mm and 15 mm	0	1	2
RA_15_20	Rain between 15 mm and 20 mm	0	1	2
RA_20_MORE	Rain over 20 mm	0	1	5
RA_LIGHT	Light rain (below 10 mm)	0	1	22
RA_HEAVY	Heavy rain (over 10 mm)	0	1	9
SUND to SATURD	Dummies for days of week	0	1	
JAN to DEC	Dummies for month of year	0	1	
Y2004 to Y2009	Dummies for years (Gatineau only)	0	1	
HOLIDAY	Holiday	0	1	4
XMAS_HOLID	Christmas Holidays	0	1	13
BREAK	Student spring break	0	1	5

Models were tested for the 31 independent variables, combining different sets of dependent variables for weather conditions. A correlation analysis was conducted to check if some variables were too highly correlated to be simultaneously included in a single model, which was not the case.

Results

Multiple models were estimated; the three most interesting are presented here (non-significant variables are removed from the final models). Table 4 confirms the findings of the descriptive analysis on the behavior of senior riders. Weather conditions negatively impact senior ridership, that mostly does non-constrained trips, and the decrease is worsening by the importance of the event. Hence, an air temperature of -20^{o}C has about the same impact as a 20 mm rainfall. However, a 20 cm snowfall is more impactful. The table also shows the seasonal effects on ridership: decrease during holidays, Saturdays and Sundays, summer months, etc.

Other models on adults, young students and college and university students are less conclusive about the effects of weather: only heavy rainfall has a slight impact on their ridership.

Table 4: Regression model for the ridership of seniors, Gatineau

Y_SENIOR				
Number of observations = 2151		R-squared = 0.8229		
F(35, 2115) = 280.86		Adj R-squared = 0.8200		
Prob > F = 0.0000		Root MSE = 131.31		
Dependent variables	Coefficient	Std. Err.	t	P>|t|
MON	582.7157	10.74331	54.24	0.000
TUE	668.2855	10.62779	62.88	0.000
WED	672.257	10.64061	63.18	0.000
THU	681.0297	10.61761	64.14	0.000
FRI	678.2314	10.66572	63.59	0.000
SAT	231.349	10.63012	21.76	0.000
ref: Sunday				
HOLIDAY	-413.0454	16.81063	-24.57	0.000
XMAS_HOLID	-137.5046	16.51097	-8.33	0.000
MAY	64.04239	13.02692	4.92	0.000
JUNE	80.53219	13.22928	6.09	0.000
JULY	29.24613	13.11759	2.23	0.026
AUG	29.96689	13.17659	2.27	0.023
SEPT	35.81441	13.29072	2.69	0.007
OCT	58.21314	12.46176	4.67	0.000
NOV	73.13939	11.78046	6.21	0.000
ref: JAN, FEB, MAR, APR, DEC				
Y2005	24.69133	9.751506	2.53	0.011
Y2006	33.23849	9.792698	3.39	0.001
Y2007	56.215	9.772301	5.75	0.000
Y2008	-24.00871	9.783734	-2.45	0.014
Y2009	-15.95694	10.10986	-1.58	0.115
ref: Y2004				
SN_0_5	-40.70232	10.26901	-3.96	0.000
SN_5_10	-90.5031	18.50452	-4.89	0.000
SN_10_15	-213.4149	30.14862	-7.08	0.000
SN_15_20	-197.3974	66.37838	-2.97	0.003
SN_20_MORE	-315.7003	47.21633	-6.69	0.000
ref: NO_SNOW				
CD_0_MINUS_5	-26.08541	11.33119	-2.30	0.021
CD_MINUS_5_10	-70.98587	13.57991	-5.23	0.000
CD_MINUS_10_15	-84.18821	14.01174	-6.01	0.000
CD_MINUS_15_20	-129.9723	15.37556	-8.45	0.000
CD_MINUS_20_MORE	-215.0836	17.42673	-12.34	0.000
ref: 0_PLUS				
RA_0_5	-54.89504	7.525309	-7.29	0.000
RA_5_10	-130.4172	12.07463	-10.80	0.000
RA_10_15	-178.5551	16.2496	-10.99	0.000
RA_15_20	-185.2767	23.16374	-8.00	0.000
RA_20_MORE	-233.0482	19.0136	-12.26	0.000
ref: NO_RAIN				
CONSTANT	*440.9507*	*13.40021*	*32.91*	*0.000*

Most of the models for Montreal do not show much impact of weather on ridership, except for those that take into account the model shift. Table 5 shows the results for bus ridership. Apart from Holidays, only heavy snows and cold winter months bring a significant decrease in bus ridership. The model also confirms the weekly (busier days: Thursdays and lower ridership on weekends) and seasonal effects (lower ridership in December and January.

Table 5: Regression model for the bus ridership, Montreal

Y_BUS				
Number of observations = 151		R-squared = 0.9343		
F(13,137) = 149.88		Adj R-squared = 0.9281		
Prob > F = 0.0000		Root MSE = 64408		
Dependent variables	Coefficient	Std. Err.	t	P>|t|
MON	461215.5	20360.14	22.65	0.000
TUE	529059.3	19932.75	26.54	0.000
WED	520775.9	19751.00	26.37	0.000
THU	547081.0	19758.72	27.69	0.000
FRI	506916.0	20112.76	25.20	0.000
SAT	79551.9	19655.60	4.05	0.000
ref: SUNDAY				
HOLIDAY	-168010.7	40646.58	-4.13	0.000
XMAS_HOLID	-216693.5	24010.80	-9.02	0.000
BREAK	-122214.3	30379.44	-4.02	0.000
SN_HEAVY	-71526.1	27736.14	-2.58	0.011
DEC	-57025.7	15720.90	-3.63	0.000
JAN	-55827.7	14895.51	-3.75	0.000
APR	-30746.3	14824.62	-2.07	0.040
ref: FEB, MAR				
CONSTANT	*40070.0*	*15837.79*	*25.30*	*0.000*

There is a much clearer figure when we look at the result for the bus-to-subway ratio model (Table 6). While very few variables remain significant, the cold temperature, the heavy snow falls and the heavy rain falls will cause a decrease of the ratio. This means that the ridership of subway will increase proportionally as the bus ridership decrease. For this short period of time (151 days), other models for the Montreal case did not show significant impacts of weather on ridership. However, this model has lower performance that the others with a much lower but still interesting R-squared).

Table 6: Regression model for the ratio bus to subway ridership, Montreal

Y_RAT_BUS_METRO				
Number of observations = 151		R-squared = 0.3830		
F(7, 143) = 12.68		Adj R-squared = 0.3528		
Prob > F = 0.0000		Root MSE = .0688		
Dependent variables	Coefficient	Std. Err.	t	P>|t|
HOLIDAY	.1631679	.0413528	3.95	0.000
XMAS_HOLID	.0771362	.0244162	3.16	0.002
MAR	.0338413	.0155065	2.18	0.031
APR	.0612997	.0169682	3.61	0.000
ref: JAN, FEB, DEC				
CD_MINUS_15_20	-.022219	.0133113	-1.67	0.097
SN_HEAVY	-.0830041	.0289869	-2.86	0.005
RA_HEAVY	-.0570898	.0243488	-2.34	0.020
CONSTANT	*1.073664*	*.011431*	*93.93*	*0.000*

Conclusion

Results

This paper presented several methods aimed to assess the impacts of weather on public transit use using smart card data. Two case studies of different magnitudes were examined: Gatineau and Montreal, Canada, public transit networks. A three-part methodology was applied: descriptive analysis, data mining and regression models.

The descriptive analysis was successfully applied to the STO in Gatineau, which is a suburb-like network in a medium-size city. The ridership charts and the day-to-day comparison showed that seniors will less use public transit during snowy and rainy days, proportionally to the importance of the weather condition. On the other hand, adult passengers, mostly workers with pendular trips, will not be affected. For Montreal, the approach brings less significant results, because of the size of the network and the possible counterbalancing effects of people staying at home vs. commuters that will use public transit to avoid traffic jams. However, there is a significant shift from the bus network to the subway network during bad weather conditions.

A data mining analysis was performed in the case of the Montreal network. Groups of behaviors were created on the basis of the hourly distribution of the trips. The analysis showed that Montreal transit users have a wide set of behaviors and will not remain in the same group all week long. This undermines our efforts to determine if the weather has an impact on their behavior. Hence, there is only one day where the behavioral shift due to bad weather condition was observable with this method.

Regression models were applied to Gatineau and Montreal cases. Except for heavy rain conditions, weather has little impact on ridership of regular adults, student and college fares in Gatineau. Only seniors are impacted by bad weather conditions, probably because some of their trips are not mandatory (not for work or study) and they can choose taxi or another mode to do their trip. In Montreal, the heterogeneous network and the small number of observations led to mitigated results. Bus ridership definitely decreases on heavy snowfall days, while the use of subway increases when it is cold, rainy, or snowy at high level.

Limitations

The work done in this project raises some issues and limitations.

First, it seems evident that studies aimed to assess the impact of weather on ridership should be conducted on the longest time period possible, because weather events are mostly rare events and a large sample of observations help to catch more of these events. For Montreal, the small number of observations, available for only one winter season, may have mitigated the results. In addition, the study looked at daily weather events only. This means that the snow precipitation could occur anytime during the day. A snowfall during the night will not cause the same impact than a snowfall in the evening. Weather events are not the only factor that could influence ridership. In this analysis, we managed to identify Holidays, school breaks and major calendar events. But his may not be enough because other elements such as service disruption, local events, shows, labor strike, smart card system failure, etc. may happen.

Some limitations are bound to the smart card system. The number of transactions may not reflect the exact number of passengers. At Gatineau, occasional users and Ottawa citizens do not have smart cards, so these people are not analyzed. At Montreal, every passenger employs a smart card or a cardboard ticket; however, during weekends, children can board for free when accompanied by an adult, and this ridership is not registered. At the STO, a picture of the user is printed on the smart card; this limits the use of a card by several users. In Montreal, there can be more than one user on each card. This could affect data mining results, based on individual behavior.

Perspectives

This paper shows that smart card data can be useful to analyze the impact of weather on ridership. This passive data is convenient, continuously captured, and more detailed than regular AVL data, because we can follow a card from a day to another and we can discriminate the fare types. Hence, this brings new research perspectives:

- A detailed analysis of the transaction time of the users could reveal the local influence of bad weather on bus service.
- An in-depth analysis of the seasonal, weekly and other cycling variations of the smart card use could help to better assess external effects, like those caused by weather.
- Detailed smart card data can help to microsimulate service adjustment strategies that could be put in place during bad weather events.

The authors wish to thank the Société de transport de l’Outaouais and the Société de transport de Montréal for kindly providing smart card data. They also would like to thank the Natural Sciences and Engineering Research Council of Canada and the funding partners of the Mobilité chair.

References

Agard, B., Morency, C., Trépanier, M. (2006). Mining public transport user behaviour from smart card data. 12th IFAC Symposium on Information Control Problems in Manufacturing – INCOM 2006, Saint-Etienne, France, May 17-19.

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

Changnon, S. A. (1996). Effects of summer precipitation on urban transportation. Climatic Change, 32(4), 481-494.

Cravo, V. S., Cohen, J. E., & Williams, A. (2009). Impact of Weather on Transit Revenue in New York City. Transportation Research Board 88th Annual Meeting (TRB 2009 Annual Meeting CD-ROM), Washington DC.

Deakin, E., Kim, S. (2001). Transportation Technologies: Implications for Planning. University of California Transportation Center, paper # 536, 27 p.

Eisenberg, D. (2004). The mixed effects of precipitation on traffic crashes. Accident Analysis & Prevention, 36(4), 637-647.

Guo, Z., Wilson, N. H. M., & Rahbee, A. (2007). Impact of Weather on Transit Ridership in Chicago, Illinois. Transportation Research Record: Journal of the Transportation Research Board, 2034(1), 3-10.

Handman, A. (2004). Weather implications for urban and rural public transit. Second Annual User's Conference of American Meteorological Society.

Kalkstein, A. J., Kuby, M., Gerrity, D., & Clancy, J. J. (2009). An analysis of air mass effects on rail ridership in three US cities. Journal of Transport Geography, 17(3), 198-207.

Keay, K., & Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. Accident Analysis & Prevention, 37(1), 109-124.

Knapp, K. K., & Smithson, L. D. (2000). Winter storm event volume impact analysis using multiple-source archived monitoring data. Transportation Research Record: Journal of the Transportation Research Board, 1700(1), 10-16.

Lee, S.G. & Hickman, M. (2011). Travel Pattern Analysis Using Smart Card Data of Regular Users. (TRB 2011 Annual Meeting CD-ROM), Washington DC.

Morency, C., Trepanier, M., Agard, B. (2007). Measuring Transit Use Variability With Smart-Card Data. Transport Policy, 14(3), p. 193-203.

Munizaga, M., Palma, C., Mora, P. (2010). Public transport OD matrix estimation from smart card payment system data. Presented at the 12th World Conference on Transport Research, Lisbon, paper no. 2988.

Pelletier M.-P., Trépanier M., Morency C. (2011), Smart card data in public transit : A review, Transportation Research C: Emerging Technologies, vol. 19, pp. 557-568.

Trépanier M., Chapleau R. (2001), 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.

Trépanier M., Morency C. (2010). Assessing transit loyalty with smart card data. presented at the 12th World Conference on Transport Research, Lisbon, paper no. 2341.

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.