Titre: Title:	Using 5 parallel passive data streams to report on a wide range of mobility options
Auteurs: Authors:	Catherine Morency, Martin Trépanier, Nicolas Saunier, Hubert Verreault, & Jean-Simon Bourdeau
Date:	2017
Type:	Communication de conférence / Conference or Workshop Item
Référence: Citation:	Morency, C., Trépanier, M., Saunier, N., Verreault, H., & Bourdeau, J.-S. (septembre 2017). Using 5 parallel passive data streams to report on a wide range of mobility options [Communication écrite]. 11th International Conference on Transport Survey Methods, Estérel, Québec (11 pages). Publié dans Transportation Research Procedia, 32. https://doi.org/10.1016/j.trpro.2018.10.014

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 32 (2018) 82–92

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Steering Committee for Transport Survey Conferences

Using 5 parallel passive data streams to report on a wide range of mobility options

Catherine Morency[a,c]*, Martin Trépanier[b,c], Nicolas Saunier[a,c], Hubert Verreault[a], Jean-Simon Bourdeau[a]

[a]*Department of Civil, Geological and Mining engineering, Polytechnique Montreal*
[b]*Department of Mathematics and Industrial engineering, Polytechnique Montreal*
[c]*Interuniversitary research centre on entreprise network, logistics and transportation (CIRRELT)*

Abstract

Montreal welcomes a variety of transportation options: transit, carsharing, bikesharing, taxi, etc. In the recent years, the "new" alternatives have gained market shares and it has become critical to understand the role each mode plays in the daily mobility of people as well as their interactions. This paper presents an analysis process of passive data streams and a typology of typical days of usage of these modes using clustering techniques. It shows that the most important attributes determining the type of multimodal usage day are type of day (working / non-working), temperature, fuel price and precipitations.

Keywords: transportation options, passive data streams, interaction, clustering, modal share

1. Introduction

There was a time when people were typically travelling at the same time, using the same mode and the same route. Few options were available and the choice was easier and quite stable over space and time. As reported by González et al. in 2005, "human trajectories show a high degree of temporal and spatial regularity". The traveler has more

* Corresponding author. Tel.: 1-514-340-4711 #4502
E-mail address: cmorency@polymtl.ca

10.1016/j.trpro.2018.10.014

transportation options than before and can easily decide at the tip of his fingers which one is the best suited for each trip. He can hop from one mode to another and switch option if conditions change en-route.

In the Montreal Area, multiple alternatives are now available. More than twenty years ago, a carsharing system was implemented and now provides a variety of services (station-based, free-floating hybrid and electric vehicles). In addition, a new free-floating carsharing provider, Car2Go, started its operation in 2013 and provides cars in the central part of the Montreal Island. In 2009, a bikesharing system with more than 5,000 bikes was implemented and continues to provide service 7 months per year. The taxi industry, faced with important challenges due namely to the wide spreading of ridesharing options, is changing its modus operandi with new players providing high-class electric vehicle services. Hence, new heavy transit services were implemented and ambitious projects have been announced for the upcoming years.

In the area, large-scale household travel surveys have repeatedly been conducted since 1970, at a rate of one every 5 years. These surveys are still central in the travel forecasting processes as they allow understanding typical travel behaviors of the Montreal residents during a typical weekday of fall. However, these surveys are unable to provide insights into the variability of behaviours throughout time as well as on the interaction between the various transportation modes since they usually have low coverage of alternative modes such as bikesharing, ridesharing, carsharing or taxi or just do not include the questions that would allow to identify members or use of these services. In this context, passive streams of data generated by operational systems can provide the necessary insights into the use of alternative modes.

This research proposes to use data from at least five different sources (bikesharing, taxi, transit, carsharing, free-floating), over multiple months, to report on the variability of mobility option usage. It relies solely on the mining of passive date streams. It illustrates one way to combine operational data of different formats, reporting on different objects, of different scales and each presenting their own data processing issues.

Hence, the research, on the one hand, aims to understand the usage pulses of each system (results are not presented here due to space limitations) and on the other hand, understand the interactions between these systems while accounting for exogenous factors such as weather, fuel price, activity cycles or events. Passive streams of data covering multiple months are processed using data mining techniques to generate typical activity patterns. Then, the occurrence of these typical days is cross-analysed along with contextual variables to identify interactions.

The paper is organised as follows. First, some background elements regarding the use of various transportation systems are provided. Then, the general methodology is described including the data sources and issues related to the processing of each dataset. Then, the results for all modes combined are exposed. A discussion concludes the paper.

2. Background

Research focusing on specific modes of transportation, namely shared modes, has multiplied over the recent years, pushed by the increasing availability of operational data. It is too numerous to fully enumerate but some examples are provided below.

2.1 Demand modelling of specific transportation modes/systems

Carsharing, either station-based or free-floating, has gained a lot of attention: see for instance Balac et al. (2015) for demand estimation, Ciari et al. (2014) for modelling of both types of carsharing (station-based and free-floating) or Heilig et al. (2017) for an application of agent-based modelling for carsharing (both types as well).

Various attempts to model bikesharing are also found in the literature namely Rixey (2013) who estimates a station-level demand model, Vogel et al (2011) who use data mining approaches to understand the patterns of use of a bikesharing system, Wang et al. (2016) who model bikesharing station activity using log-linear and negative binomial regression models and Morency et al. (2017) who model the number of transactions per day per station using 6 years of bikesharing transaction data.

Research is also conducted on taxi demand. Using taxi data from New York, Hochmair (2016) analyses the variability of taxi trips, over space and time, using 29 million trips, Pele and Morency (2014) propose various indicators using operational data from 1,000 taxis in Montreal while Lacombe and Morency (2015) and Yang and

Gonzales (2014) use regression models to explain taxi trips using variables such as access time to transit, population, age, etc.

Availability of smart card data is also at the core of the numerous research on transit use. Morency et al. (2007) use data mining technique and smart card data to assess transit use variability, Huang et al. (2015) do similar work using data from Chengdu and Agard et al. (2013) use more advance data mining techniques to classify travelers into clusters of typical behaviors.

2.2 Multimodal analysis

Thus, research about parallel examination of transportation data stream is rather rare in the literature, probably due to the limited availability of data on many modes at the time.

Chaolong et al. (2016) proposed an architecture to store and analyze data, but did not use their model in a practical application. Trépanier and Yamamoto (2015) have discussed about the challenges related to the analysis of different transportation data streams such as GPS, smart card, Bluetooth and mobile phone.

Current literature is related more to data fusion. Bayart et al. (2009) confirm the utmost importance of data fusion but the poorness of transportation-related examples. Recent work has shown the potentialities of fusing data from many sources and for example solved some methodological problems related to the spatial and temporal matching of bus network datasets (Giraud et al. 2017). However, the challenge of dealing with continuous data streams remains. Characterizing travel behavior is still a challenge, due to the variations in networks and systems' use (Morency et al., 2016). Briand et al. (2017) have addressed the problem by using mixtures of Gaussian for the modelling of temporal distributions of transit use. Ghaemi et al. (2017) propose dedicated distance calculations in the case of spatial-temporal smart card data.

However, most data fusion works are related to sensors and other ITS data. The reference for sensor fusion has been set by Ben-Akiva and his team (Antoniou, 2004, Huang et al., 2009). Their approach belongs to model-based data-fusion techniques. In their work, the model performs dynamic traffic assignment using a simulation-based approach, but is limited to a single mode. Zheng (2015) and Zheng et al. (2015) also present several machine-learning based data fusion techniques.

Finally, the concept of integrated mobility and multimodality (use of various transportation modes to fulfill travel needs) is greatly discussed as it is a step towards the implementation of strategies involving all available transportation alternatives and this requires modelling tools that also integrate all modes. Spickermann et al. (2014) propose a conceptualization of future urban mobility systems that builds on the strengths of the various modes. Shuppan et al. (2014, p.553) confirm the "growing dominance of multimodal mobility and the declining role of private cars in everyday mobility in dense urban areas", paving the way for integrated and multimodal understanding of daily travels. Gebhardt et al. (2016) analyze intermodality (use of multiple modes within a single journey) using survey data to estimate the number of intermodal trips and people in the objective to model intermodality in cities using demographic composition and spatial configuration.

3. Methodology

This section describes the various data used for the analysis as well as the methods used to produce typical days of travel and look at the variability of usage.

3.1 General methodology

The research process involves a variety of steps which are exposed in Figure 1. First, due to some data limitations that will be further explained in this section, it is necessary to select a study area as well as a period of analysis. Secondly, depending on data availability, few or many mobility systems can be included. In this research, the systems included in the development of typical days of mobility system usage are: roads, transit (bus and subway), bikesharing, taxi, carsharing (station-based and free-floating). The variability of usage of each system is examined to facilitate the understanding of the dynamics of each system while all data are after integrated to develop typical days of usage for all included mobility systems. Classification techniques are used to create typical days of usage and explanatory variables are afterward included to describe the created clusters.

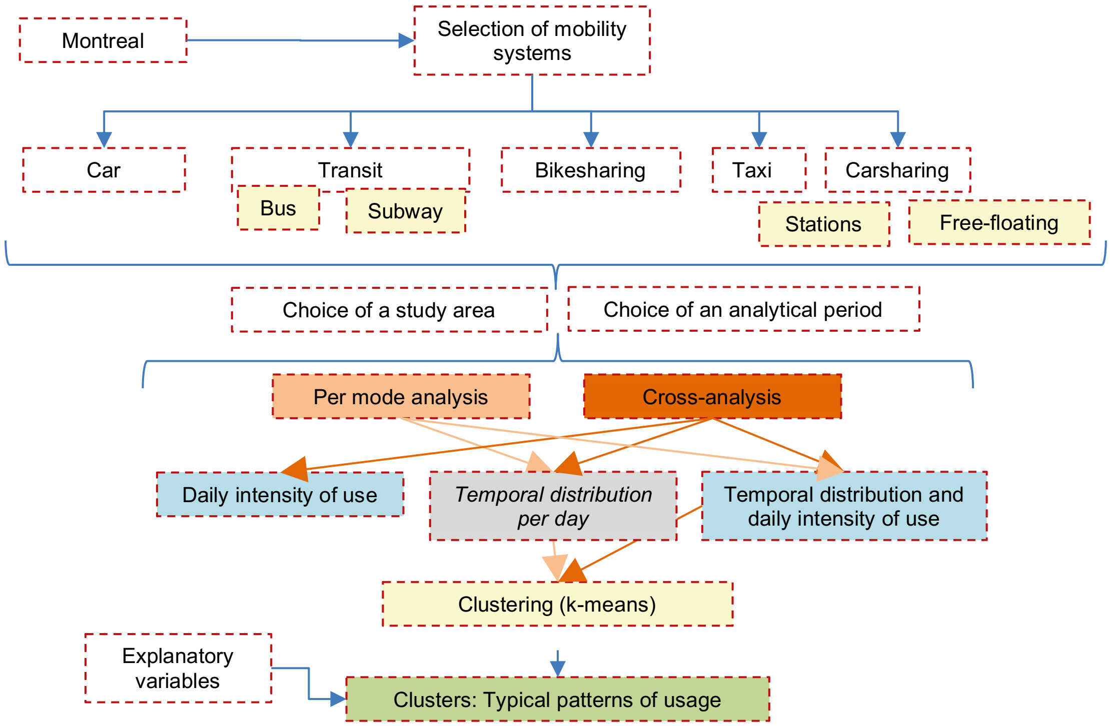

Figure 1. General methodology to produce typical patterns of daily usage

3.2 Datasets

There is no unique dataset providing insights into the use of all available modes of transportation as well as of the variability of this usage over space and time. On the one hand, we have large-scale origin-destination travel surveys that provide a detailed description of typical daily travels of residents every five years and, on the other hand, we have passive data streams monitoring the use of systems. None of these types of data is sufficient to understand the role of each mode of transportation in a city nor inform on how usage change in space and time in the presence of an event or changing context.

In this research, we propose to look at the use of various independent mobility systems using passive streams coming from operational systems. These systems are:

- Car (road system): to account for the level of usage of the road systems, data from a traffic counting system are included. One loop located in an important highway and for which count data is available for enough days is selected. It does not represent the variability of traffic throughout the Montreal region but it gives a good insight on the level of congestion in the central part of the area. The usage value is the number of vehicles using the highway link every hour of each day.
- Transit (subway): the use of the subway system is captured through the smart card validations at the entry station. Only entrance points are available. The number of validations per hour per day is the usage indicator.
- Transit (bus): the use of the bus system is also captured using the smart card transactions though for buses, we don't know where the transaction is occurring (at best, we have the line number). Again, smart cards are only tap-in. The number of bus validations per hour per day is the usage indicator.

- Carsharing (station-based): shared cars are reserved by members and it is possible to measure the number of reservations starting each hour of each day. It is the indicator used to assess the level of usage but it is important to mention that a transaction does not necessarily amount to one trip since a car can be rent to perform a set of trips and that the car can be reserved for multiple days.
- Carsharing (free-floating): for free-floating carsharing, the number of transactions more or less equals the number of trips. Hence, the usage indicator is the number of transactions starting each hour of each day.
- Taxi: the use of taxi is estimated using the number of trips made using the most important taxi company in the area (1,000 taxis accounting for almost 25% of the fleet, assuming their usage pattern is representative of the entire fleet). The usage indicator is the number of taxi trips starting each hour of each day.
- Bikesharing: a 5,000 bikes' service, Bixi, is available in the region. The use of the system is reflected by a database describing each transaction including origin and destination station and timestamp at both points. The number of transactions starting each hour of the day of each day is used as usage indicator. It is worth noting that the system is typically only available from April to November (due to winter).

All these modes are not available across the entire Montreal Area and some datasets do not cover the entire year. Hence, a subset of the area where most services are available is selected (which corresponds to the Bikesharing service area) and the days where all systems (that should operate) are observed are included. This area amounts to 107 square kilometers. It includes most of the bikesharing and subway stations.

The resulting data are summarised in Table 1, namely the total usage over the study area and some details regarding the available data for each system. 275 days of observation are used for the analysis.

Table 1. Summary of datasets used in the analysis

Mobility system	**Days observed**	**Note**	**Total usage on the entire service area**	**Total usage on the study area**
Car	351	A single link on a main highway within the study area, veh/h	50,552,061	50,552,061
Transit - Bus	365	All transit lines crossing the study area, boardings/hour	236,048,727	190,595,595
Transit - subway	365	All subway stations within the study area, boardings/hour	245,575,946	195,713,138
Bixi	215	*In operation from April 15th to November 15th,* some transactions before and after the official start of the season are removed since not all stations available, transactions/hour	3,527,844	3,506,213
Taxi	275	Data unavailable for several days – origin point in the service area, trips/hour	3,648,204	2,729,698
Station-based (SB) carsharing	365	Number of transactions starting in each station located within the study area, transactions/hour	359,570	314,587
Free-floating (FF) carsharing	365	Deployment throughout 2015 so increase in the number of vehicles available, trips starting in the service area/hour	238,533	211,613

3.3 Classification approach

Data mining techniques are used to create typical days of usage for each system independently and using usage from all systems simultaneously; only the latter results are presented. A k-means algorithm is used to cluster days of usage using vectors of attributes describing, for each day:

- The proportion of trips conducted for each hour of the day, the sum of the 24 hours amounting to 1;
- And indicator of usage intensity for this day which is a normalised value, between 0 and 1, indicating whether a day is important for one mode or not. Normalisation is conducted independently for each mode. This indicator is estimated as follows:

$$Daily\ usage\ intensity_{\,j=n} = \frac{\sum_{h=0,j=n}^{h=23,j=n} uses}{Max\left(\sum_{h=0,j=m}^{h=23,j=m} uses\right),\ m \in \{1 \dots 365\}}$$

Hence, for each mode, the vector of attributes contains 25 columns of data. An example is provided below (Table 2): x0 to x23 contain the proportion of usage for each hour of the day (for instance, 0.0489 means 4.89% of the transactions of January 1st occur between midnight and 1 am) while x24 contains the daily indicator of intensity. For the cross-analysis, all the mode-specific vectors are linked together to create daily vectors of 175 values describing both the temporal distribution of usage of each mode and the daily intensity of use (24 hours per day per mode and 7 intensity of usage).

Table 2. Example of the dataset used in the analysis

date	x0	x1	x2	x3	x4	x5	x6		x24
2015-01-01	0.0489	0.0761	0.0707	0.0326	0.0163	0.0109	0.0054		0.127
2015-01-02	0.0081	0.0040	0.0000	0.0000	0.0000	0.0081	0.0040		0.172
2015-01-03	0.0183	0.0046	0.0091	0.0046	0.0000	0.0091	0.0000		0.151
2015-01-04	0.0105	0.0105	0.0053	0.0053	0.0000	0.0000	0.0000		0.131
2015-01-05	0.0160	0.0107	0.0107	0.0000	0.0053	0.0107	0.0214		0.129
2015-01-06	0.0125	0.0083	0.0000	0.0000	0.0042	0.0042	0.0042		0.166
2015-01-07	0.0101	0.0152	0.0051	0.0000	0.0000	0.0101	0.0202		0.137
2015-01-08	0.0179	0.0090	0.0000	0.0000	0.0000	0.0045	0.0179		0.154

For each clustering, many clusters must be determined. So, with the identified vectors, a within sum of square graph is first constructed using from 1 to n clusters with a k-means. Then, a hierarchical classification is conducted to construct a dendrogram using the centers of 30 clusters determined using k-means. Both the dendrogram and the within-sum-of-square graph are used to assist in the selection of the possible number of clusters. Finally, several k-means are conducted using these possible number of clusters to observe results and make the final decision regarding the most appropriate number of clusters for the research. All the clustering processes were estimated in R using the Stats package (function kmeans and hclust).

4. Results of the cross-analysis

For the cross-analysis, seven multimodal clusters were identified, three including only working days, two including solely non-working days and one with mostly non-working days but few working days (regular holiday days without being officially non-working) – see Table 3 – this cluster will be considered a non-working day cluster.

Figure 2 presents the daily intensity of each mode for each of the 7 clusters. This is one of the indicators used as input to the clustering. We see that some modes have quite changing intensity of usage between clusters such as FF carsharing, taxi and bikesharing. Also, both transit modes follow similar patterns with one of the most intense clusters (C6) related to the least intense car cluster (for working days). While no causality is confirmed, we see that the most intense bikesharing cluster is also the least intense taxi cluster for both working and non-working days. The temporal distribution of each mode for each cluster is also examined (just for working days cluster – but the same analysis can be conducted for non-working days).

Table 3. Composition of the 7 clusters for the cross-analysis

Cluster	Working days	Non-working days	Total
1	14.1%	0.0%	9.5%
2	0.6%	25.3%	8.7%
3	0.0%	44.6%	14.6%
4	35.9%	0.0%	24.1%
5	23.5%	0.0%	15.8%
6	25.9%	0.0%	17.4%
7	0.0%	30.1%	9.9%

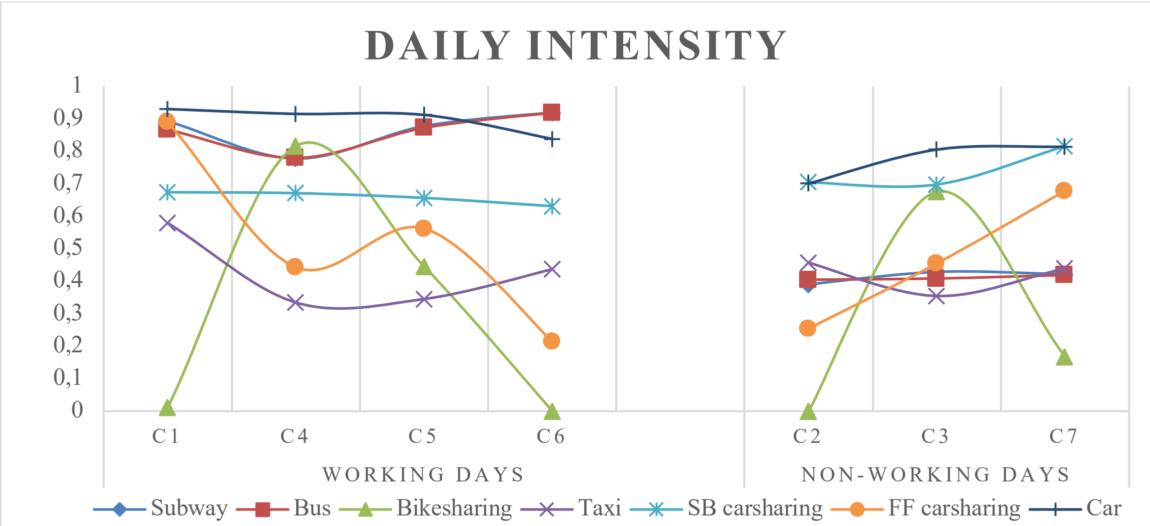

Figure 2. Daily intensity of usage of all modes of each cluster

Figure 3 presents the distribution of clusters throughout the years (for the 275 days of observation). This figure helps us observe the rhythms of usage by the changes in clusters representing usage of the different modes. Clusters related to workdays and non-work days are presented separately. For non-workdays, we more or less see that there is a shift from C2 at the beginning of the year to C3 for the summer months and then to C7 for the end of the year, with few days not following this trend. The difference between C2 and C7 is mostly related to the intensity of usage of the FF carsharing service. During the year, there was an increase in the number of cars available for this service and we see that it is changing, as expected, the intensity of usage of this mode. C3 also has a lower use of taxi and a much more important use of bikesharing and it is related to the clement weather available during the summer months. For the workdays, we see a similar pattern. The start of the year is related to C6 then, as the average temperature increases, we move to C5 and then C4 for the summer month. Then, the end of the year is C1 with a more intense use of FF carsharing.

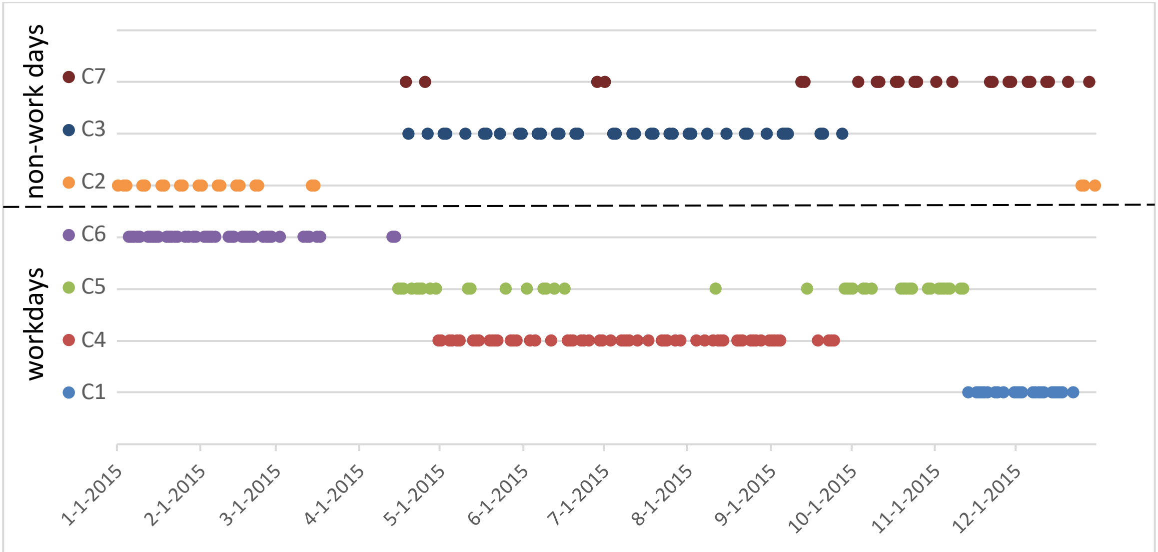

Figure 3. Annual distribution of clusters

Figure 4 presents the temporal distribution of the most important workday cluster, for each mode. The presented graph helps observe the differences between modes for a particular cluster. Hence, we clearly see the differentiated patterns of the car with equal share throughout a large part of the days. The taxi has a similar distribution but with a more important share during the night hours. Still, the other modes mostly have peak-oriented patterns.

C4

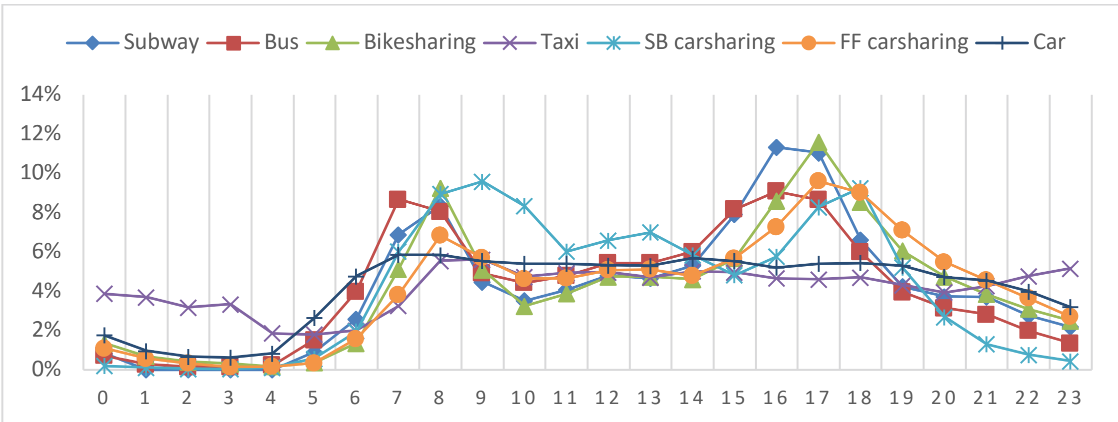

Figure 4. Temporal distribution of usage of all modes for cluster C4

Another analysis tool is used to contribute to help identify the factors correlated with a particular type of day (hence the belonging to a certain cluster). Using various explanatory variables, a decision tree is constructed with the cluster number as the outcome. The tree function from R (tree package) is used for the estimation. The explanatory variables, all estimated daily, are:

- Type of day (workday / non-work day)
- Average temperature
- Millimetres of precipitation
- Fuel price
- Transit level intensity estimated using the number of daily passages at bus stops

The decision tree identifies and classifies the variables which are most important in predicting the belonging of a day to one cluster. The result of the decision tree process is presented in Figure 5. We see that the most important daily feature is the type of day (workday vs non-workday). That was expected since the identified clusters are basically all pure clusters (linked to solely one type of day). For workdays, the second most important variable is transit supply intensity, which is a proxy for high intensity level of usage. As seen on the tree, C6 is entirely forecasted using type of day and intensity of transit supply. For similar transit supply, it is the average temperature and the precipitation levels which are the following variables. Hence, if temperature is above 11 C° degrees and precipitation level is below 6.3 mm, days will belong to C4. For non-workdays, we see that the most important variable is average temperature and that it influences the type of cluster more than once with a first cut-off point at 6.4 C° degrees and a second one at 3.7 C° degrees further in the three. Fuel price is also identified as a determinant factor. The decision tree helps identify which features are determinant with respect to particular patterns of mode usage and can assist planners identify event or features that can affect travel demand for specific modes.

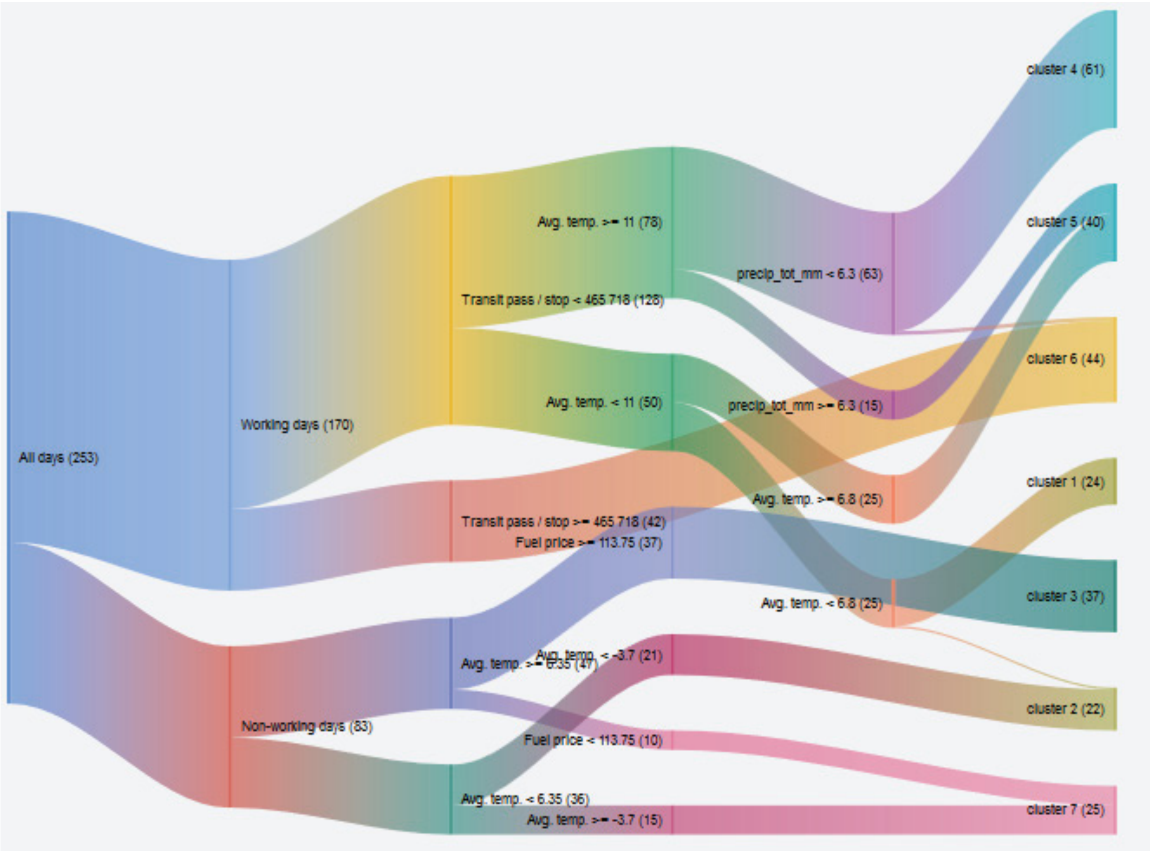

Figure 5. Outcome of a decision tree to predict belonging of days to clusters

5. Discussion

In conclusion, the current section proposes a synthesis of what this research proposes, it discusses the various limitations and proposes some research perspectives.

Using various sets of heterogenous data, this paper has presented the results of a clustering process integrating features of daily usage of various modes. One key element is to be able to identify relevant usage indicators for each mode while making sure the difference in scale does not affect the comparison, some modes supporting more daily trips than others. The cross-analysis confirmed the segmentation power of working vs non-working days with respect to patterns of usage of all the modes –commuting and school trips still have a determinant impact on the way the various transportation systems are used across the year. This analysis also revealed that some modes seem correlated to others: this is the case for bikesharing and taxi, bikesharing and transit – an assumption often formulated that is currently supported by the observations. Hence, other significant features in explaining the type of usage day are the temperature and the precipitations.

Processing various sets of data streams is not without issues and this research has many limitations. There are some unobserved days of travel that had to be excluded for the analysis. They may have different usage levels and patterns and they are not included. Also, the spatial component is currently neglected: the analysis is conducted at the scale of the selected area but there is also variability of usage across space; interactions may be more important in specific locations where some services are more available (near carsharing or bikesharing station for instance). Another challenge is the identification of a usage indicator: for many modes, it is possible to use the number of trips while for others it is the number of vehicles passing by (counts) or beginning of reservations in the case of station-based carsharing (one reservation leading to at least two trips). Also, some modes are still not accounted for: cycling and walking for instance.

There are plenty of expectations and challenges related to the research program overseeing this paper. From a methodological point of view, various elements will be experimented: new usage indicators for the various modes, inclusion of the spatial component, definition of richer vector of attributes for the clustering process, experimentation of other data mining techniques allowing to better compare vectors while accounting for the sequencing, using longitudinal model (growth models for instance) to predict usage of the various systems over time.

From an analytical point of view, other modes of transportation need to be included in the model, most importantly walking and cycling and variables related to service supply and design need to be included to allow for the estimation of scenarios. Longer range of data are also currently being processed (more than two years) to grab evolution trends in addition to yearly variability.

The automatic processing of passive streams of data from administrative systems is an attractive opportunity for transportation system planners. It can provide a unique, continuous view of when and how the various modes are used within an area and provides ground for informed decision making regarding which mode to prioritize, what level of service to provide and how to articulate fare bundles.

Acknowledgements

The authors wish to acknowledge all the partners who provided access to data for research purposes (Montreal Transit Authority, Ministère des transports, de la mobilité durable et de l'électrification des transports du Québec), Communauto (carsharing company), Bixi (bikesharing company) and Taxi Diamond. They also wish to thank the Canada Research Chair Program for funding.

References

Agard, B., V. Partovi Nia, and M. Trépanier. Assessing public transport travel behaviour from smart card data with advanced data mining techniques. Presented at the 13th World Conference on Transport Research, Rio de Janeiro, Brazil, 2013.

Antoniou, C. (2004). Online Calibration for Dynamic Traffic Assignment. PhD thesis, MIT, Department of Civil and Environmental Engineering.

Balac, M., Ciari, F., Axhausen, K. (2015). Carsharing demand estimation: A Zürich area case study, 94th Annual Meeting of the Transportation Research Board, Washington DC.

Bayart, C., Bonnel, P., Morency, C. (2009). Survey Mode Integration and Data Fusion: Methods and challenges. In Bonnel, P., Lee-Gosselin, M., Zmud, J., Madre, J.-L. (Eds). Transport survey methods. Keeping up with a changing world, Emerald.

Briand, A.-S., Côme, É., Trépanier, M., Oukhellou, L. (2017), Analyzing Year-To-Year Changes In Public Transport Passenger Behaviour Using Smart Card Data, Transportation Research Part C: New Technologies, Volume 79, June 2017, pp. 274–289.

Ciari, F., Bock, B., & Balmer, M. (2014). Modeling Station-Based and Free-Floating Carsharing Demand: Test Case Study for Berlin. Transportation Research Record: Journal of the Transportation Research Board, No 2416, pp. 37-47.

Chaolong J., Hanning W., Lili W. (2016), Study of Smart Transportation Data Center Virtualization Based on VMware vSphere and Parallel Continuous Query Algorithm over Massive Data Streams, In Procedia Engineering, Volume 137, Pages 719-728

Gebhardt, L., Krajzewicz, D., Oostendorp, R., Goletz, M., Greger, K., Klötzke, M., Wagner, P., Heinrichs, D., Intermodal Urban Mobility: Users, Uses, and Use Cases, Transportation Research Procedia 14 (2016) 1183 – 1192

Ghaemi, M.S., Agard, B., Trépanier, M., Partovi, Nia V. (2017), A Visual Segmentation Method for Temporal Smart Card Data, Transportmetrica, 13 (5), pp. 381-404

Giraud, A., Légaré, F., Trépanier, M., Morency, C. (2017), Data fusion of APC, smart card and GTFS to visualize public transit use, 96th Annual meeting of the Transportation Research Board, paper 17-03052.

González, Marta C., Hidalgo, César A., Barabási, Albert-László (2008). Understanding individual human mobility patterns, Nature volume 453, pages 779–782 (05 J,une 2008)

Heilig, M., N. Mallig, O. Schröder, M. Kagerbauer, and P. Vortisch. Implementation of free-floating and station-based carsharing in an agent-based travel demand model. Travel Behaviour and Society, Available online 23 February 2017, 2017.

Hochmair, H.H. (2016). Spatiotemporal Pattern Analysis of Taxi Trips in New York City, Transportation research record, Vol. 2542, pp.45-56.

Huang, E., Antoniou, C., Wen, Y., Ben-Akiva, M., Lopes, J., and Bento, J. (2009). Real-time multi-sensor multi-source network data fusion using dynamic traffic assignment models. In Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on, pages 1–6. IEEE.

Huang, J., Xu, L., Ye, P. (2015). Exploring Transit Use Regularity Using Smart Card Data of Students, Conference: Proceedings of the Fifth International Conference on Transportation Engineering, pp 617-625.

Lacombe, A. et Morency, C. (2015). Modeling Taxi Trip Generation Using GPS Data: The Montreal Case, Presented at the 95th Annual Meeting of the Transportation Research Board, Washington DC.

Morency, C., M. Trépanier, and B. Agard. Measuring transit use variability with smart-card data. Transport Policy, Vol. 14, No. 3, 2007, pp. 193 - 203. http://dx.doi.org/10.1016/j.tranpol.2007.01.001.

Morency, C., Trépanier, M., Saunier, N., Verreault, H., Bourdeau, J.-S. (2016) Understanding the interactions between transit, taxi and bikesharing, International Symposium on Emerging and Shared Transportation Modes and Mobility Services, Tel-Aviv, December 2016.

Morency, C., Trépanier, M., Paez, A., Verreault, H., Faucher, J. (2017) Modelling bikesharing usage in Montreal over 6 years. CIRRELT PAPER CIRRELT-2017-33, available online: https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2017-33.pdf

Pele, N. et Morency, C. (2014). When, Where and How Taxis Are Used in Montreal. Presented at the 93rd annual meeting of the Transportation Research Board, Washington DC.

Rixey, R. A. (2013). Station-Level Forecasting of Bikesharing Ridership. Transportation Research Record: Journal of the Transportation Research Board, No 2387, pp. 46-55

Shuppan, J., Kettner, S., Delatte, A., Schwedes, O. (2014). Urban multimodal travel behaviour: towards mobility without a private car, Transportation Research Procedia 4 (2014) 553 – 556.

Spickermann, A., Grienitz, V., von der Gracht, H.A. (2014). Heading towards a multimodal city of the future? Multi-stakeholder scenarios for urban mobility, Technological Forecasting & Social Change, 89 (2014) 201–221

Trépanier M., Yamamoto T. (2015), Workshop Synthesis: System Based Passive Data Streams Systems; Smart Cards, Phone Data, GPS, In Transportation Research Procedia, Volume 11, Pages 340-349

Vogel, P., T. Greiser, and D.C. Mattfeld. (2011). Understanding bike-sharing systems using data mining: Exploring activity patterns. Procedia - Social and Behavioral Sciences, Vol. 20, 2011, pp. 514-523. http://dx.doi.org10.1016/j.sbspro.2011.08.058.

Wang, X., Lindsey, G., Schoner, J.E., Harrison, A. (2016). Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. Journal of Urban Planning and Development, Vol. 142, No. 1, 2016. ttp://dx.doi.org/10.1061/(ASCE)UP.1943-5444.0000273.

Yang, C. et Gonzales, E. (2014). Modeling taxi trip demand by time of day in New York City. Transportation Research Record: Journal of the Transportation Research Board, (2429), pp. 110–120.

Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. IEEE transactions on big data, 1(1):16–34.

Zheng, Y., Zhang, H., and Yu, Y. (2015). Detecting collective anomalies from multiple spatio-temporal datasets across different domains. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, page 2. ACM.