Continuous tracking : exploring the long-term value of Bluetooth

Francois Bélisle (corresponding author)
Data Scientist, Math Mobile inc.
f@mathmobile.io

Nicolas Saunier
Associate Professor, Department of Civil, Geological and Mining Engineering
Polytechnique Montréal
nicolas.saunier@polymtl.ca

Olivier Audet
Engineer, Innovation and Mobility
City of Montreal
olivier.audet@ville.montreal.qc.ca

5024 words + 6 figures + 1 tables = 6769 words
November 15th, 2017

ABSTRACT

This paper explores the value of continuous tracking of Bluetooth devices in a densely populated urban area. Over 80 Bluetooth Media Access Control Scanner (BMS) devices were installed in 2015 in the central portion of Montreal, the Port of Montreal, and part of the city's downtown core, and have been detecting devices since. The resulting database contains over 85 million observations emanating from 1.3 million Bluetooth devices over 640 days. To the author's knowledge, no such wide-ranging data set as been analyzed. With the methodology described in this paper, over 17 million trips were found (approx. 26 thousand per day), with travel-time, travel-distance and origin-destination pairs generally corresponding to expected results. Moreover, the median of the distribution of BT devices per trip shows that trips span over a few detection devices and are not tracked exhaustively over the network. However, extreme values are an important part of the presented distributions. Further investigation is necessary to understand the underlying generating events which are not easily identified without additional information. Finally, this paper explores speed distributions of trips and showed that different families of distributions could point towards different modes of vehicular trips, even multi-modal trips.

INTRODUCTION

The objective of this paper is to better understand the properties of Bluetooth detections and derive methods to extract useful traffic data in a given context. This research focuses on exploring raw detection data and exploiting a “big” dataset provided by the City of Montréal.

The Bluetooth (BT) wireless technology has been developed and used since the early 2000s (*1*) to obtain traffic data, mostly vehicle travel time, and analyze road user behavior. The principle about the use of the Bluetooth Media Access Control Scanner (BMS) for transportation data collection is simply described in (*2*):

> A BMS scanner has a communication range (say around 100 m in radius), here termed as a zone. The zone is scanned to read the Media Access Control addresses (MAC-ID) of the discoverable BT devices transiting within the zone. The MAC-ID is a unique, alpha-numeric string, that is communicated by the discoverable BT device.

Being a passive collection device, BMSs have the major advantage of providing bountiful data of individual travelers on potentially large portions of regional networks and continuously over long periods of time. However, there are several uncertainties related to the BT protocol and BMS characteristics, the BT device characteristics, and the resulting information pertaining to the transportation, mostly trips. Nothing is known about the actual trips made by a vehicle or person between two detections. These limitations have impacts on the quality of the traffic information derived from the raw detections. For example, when calculating travel times in a dense urban core, different kinds of users, such as pedestrians and vehicles (cars, buses, taxis, delivery trucks, etc.) will be recorded without knowing it: the resulting sample will be a mixture of several distributions of unknown proportions.

The dataset used in this research contains more than a year of individual MAC-ID detections covering a continuous period starting in December 2015 and continuing today. This allows us to derive trips per device (user) over that time period. This is different than the data obtained from usual proprietary solutions from vendors such as TPANA, TrafficCast, Acyclica or Traffax, which provide only the cleaned travel times for specific roads or pairs of origin-destination on a network.

The main idea of the proposed methodology is to use traveling habits found over the whole period of data collection to exploit the complementarity and redundancy found in this "big" dataset and filling in unknowns when possible. Information provided by BT detections is in itself limited when considered punctually or individually, but the objective of this paper is to evaluate whether accumulating BT detections of a device over long periods of time and space can provide knowledge about the user’s habits or about the use of the network. For example, can a model be developed to predict the mode and even understand multi-modal trips from BT data?

After a general description of BT characteristics for traffic data collection and related work, this paper presents the dataset and the principal results. To the authors’ knowledge, this is one of the largest datasets of this kind where the MAC-IDs are available to track devices over the whole period of data collection. The paper concludes with the next steps for the possible uses and benefits of such data.

BACKGROUND

One must understand how the detections are made to analyze the resulting data. There are several uncertainties related to:

1. the BT protocol, BMS characteristics:
   - the size and shape of the BMS communication, i.e. detection, zone;
   - the duration of the inquiry cycle, which is the period of time during which each BT device may be discovered: "generally the data acquisition software linked with BMS only provides the MAC-IDs scanned during an inquiry but not the exact time when it is discovered" (*2*);
   - the maximum number of BT devices the BMS can detect during an inquiry cycle;
   - the other data reduction done by the BMS: the discovered MAC-ID are usually linked to the beginning or the start of the inquiry interval. If multiple detections occur over consecutive inquiry intervals, the BMS typically keeps only the first detection instant and may keep the time between the last and first detection or duration (*2*);
2. the BT device characteristics:
   - MAC-ID are supposed to be unique for each device, but can be cloned, in which case devices cannot be differentiated anymore (*3*);
   - there has been recent discussions of both WiFi and BT MAC address randomization to avoid device tracking, in particular for Apple devices (*4*).
3. information about the detection:
   - what? Some information about the type of device (earpieces, phones, tablets, etc.) may be extracted from the MAC-ID, but not about the person who is carrying the device[1], in particular about its mode of transportation;
   - where? The device is in the detection zone, without any other information;
   - when? The detection time is known up to the inquiry cycle.
4. information about trips: the most common use of BMS detection times is to match MAC-ID to derive travel time between the BMS zones but anything can happen between two detections of the same device: the assumption is that the most frequent matchings will correspond to direct trips, but other scenarios may happen (*5*).

It follows in particular that the real volume of user detected by a BMS or moving between two BMS zones is very hard to measure (*6*). The simplest way would be to estimate the capturing rate, i.e. the rate of the detections over the total number of users in the zone. Yet it is not constant over time or space, as it is a function of the penetration rate of BT devices (with BT turned on) in the population, the probability of detection and the maximum number of devices a BMS can detect (*7*). The relationship between the number of detections and the number of users of interest is therefore non-linear, and the last characteristics being particularly important as the number of detections will plateau beyond some number of users in the detection area. A probability of detection of 0.7 is reported in ideal conditions indoor in (*7*), and it fluctuates with with the position of the scanner. Other factors affecting the capturing rate are: distance and angle between the scanner and the car,

[1]This is a feature for privacy, not a bug.

as well as the car speed (*7*), whether the antenna is omni- or uni-directional (*2*) and the antenna gain (*8*).

Once travel times are derived from the BT detections, different cleaning methods have been tried, usually around testing thresholds on different deviation formulations from the central tendency of the data, based on the box and whisker statistics of boxplots or median absolute deviation, the second being recommended by (*5*). Despite these complications, most travel time estimated by Bluetooth data were within 10 % of the ground truth in the study described by (*9*). Vehicle highway travel times obtained from BMSs are even used as ground truth for other sensors for example for the I95 corridor coalition (*10*). BMSs have also been used for other modes, for pedestrians (*8, 11*), cyclists (*8*) and buses (*3, 5*).

Although such data has been available for some time, there are few studies about the properties of detections and travel times obtained from BMSs (*2, 3, 12*), and none to our knowledge on the long-term tracking of BT devices. Furthermore, most studies focus on one mode or type of users and do not try to identify different modes.

GENERAL DATA DESCRIPTION

The data is generated by 80 TPANA BMSs spread over the Island of Montréal in Québec, Canada, as illustrated in figure 1. The earliest set of BMSs have been installed in December 2015; all BMSs record continuously, being connected to the Montréal traffic management center ("Centre de Gestion de la Mobilité Urbain", CGMU). Other groups of BMSs have been installed successively depending on the city's projects. All BMSs are installed on the local network, in the jurisdiction of the City of Montreal, and excludes all highways under the jurisdiction of the Ministry of Transportation, although some sections associated to ramps are equipped with BMSs. In particular, a high density BMS installation has been made on most corner streets in the Quartier des Spectacles, the entertainment district in the heart of Montréal's downtown with a high density of travelers using all possible modes, including Montréal's busiest bike path (on Boulevard de Maisonneuve). In this sector, the distance between BMSs is small by industry standards (*13*), ranging from 100 m to 250 m. Also, BMSs have been installed around the portion of the Port of Montreal where 2500 trucks come to load and unload daily (14).

Technically speaking, the data collected by the system is composed of :

- a timestamp;
- the MAC-ID of the detected device;
- the id of the BMS that detects the device;
- a time interval during which the detection occurs;
- the number of times for which the device was detected during this time interval.

The data provided in the database also contained aggregated speed on pre-determined pairs of detectors. Because only these pairs are considered, it is a subset of the raw data, filtering successive detections that have no corresponding matches. For the present data, we estimate that more than 30 % of raw points were rejected by the proprietary TPANA software. The authors have chosen to analyze raw data representing individual device detections, or pings, in the hope of finding users further downstream in their trip even if they were lost on successive BMSs.

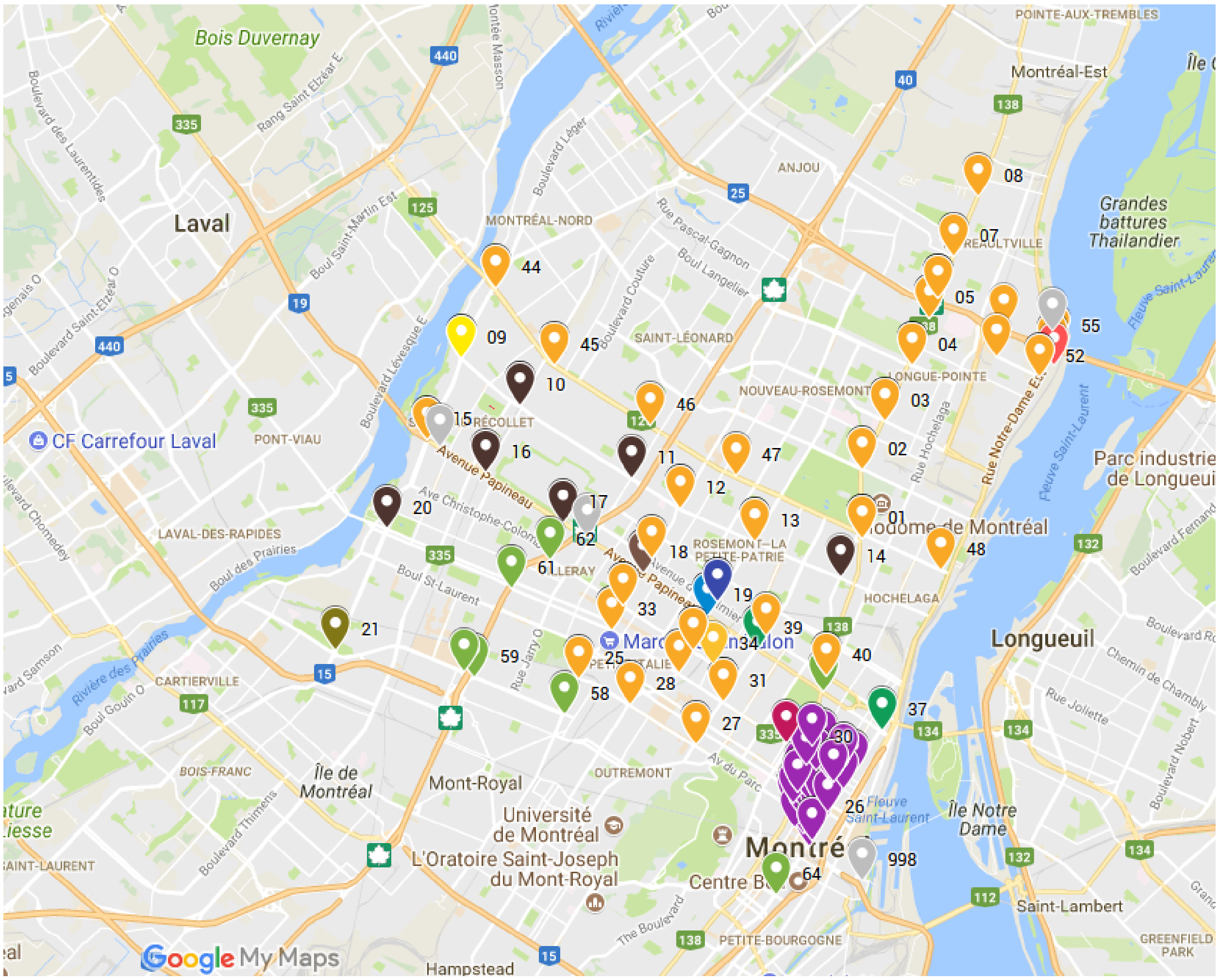


FIGURE 1 : This image shows the location of individual BMSs on the Island of Montréal in Québec, Canada. Individual colors show different project installation, in particular purple in the area of the entertainment district (Quartier des Spectacles) (source: Google Maps).

The following descriptive tables and figures present an overview of the size and importance of the given data set.

TABLE 1 : This table presents basic statistics describing the raw data set of BMS detections

First / Last Detection Date	2015-06-01 / 2017-03-02
Detections	85,139,587
Detections for devices seen more than once	84,770,638 (99.5 %)
Detections for devices seen more than once and at 2 locations or more	81,862,929 (96.1 %)
BT Devices	1,272,350
BT Devices seen more than once	903,402 (71 %)
BT Devices seen at 2 locations or more	811,903 (63.8 %)

A few comments about the numbers presented in table 1.

- The Island of Montréal is populated by roughly 2 million people, while its greater metropolitan area accounts for 4.1 million people. It was estimated that more than 50 % of residents of the province of Quebec, where the city is located, have a smart phone in 2014 (*15*). This rate varies per age and socio-economic group. Although the BMSs cover an important part of higher road-network and the densely populated areas of the city, it does not cover the whole island nor all the principal arterials to the downtown core;

- The devices seen more than once account for almost all of the collected data;

- Roughly less than a third of devices where seen once ; this might emphasize the ephemeral nature of the underlying devices or the transient nature of the data being recorded ;

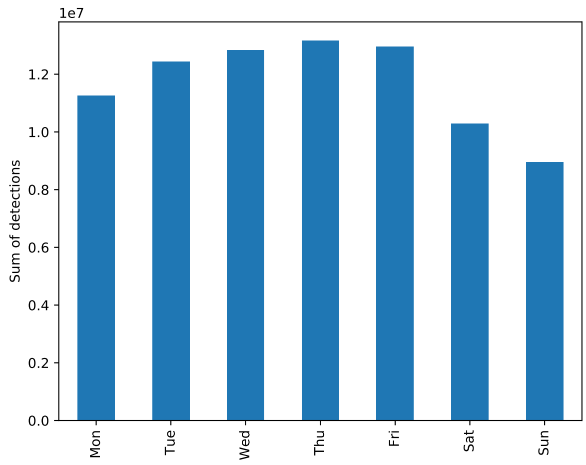
- Roughly more than a third of devices were seen more than twice, but at only one location. This could represent a number of phenomena, from still devices being turned off and on to actual movement of devices within the detector's tracking area ;

- The area covered by BMSs represents a densely populated area (Ahuntsic, Rosemont, Plateau Mont-Royal, etc.) and part of the downtown core (the Quartier des Spectacles, as mentioned earlier), concentrated in the north-south axe around the metro's central and main orange line. Although this significant deployment covers main arterials, it represents a fraction of the island as a whole. The area of the island covered by the BMSs is 11 kilometers long and 5 kilometers wide, representing just over 11 % of its total area, excluding the suburbs.

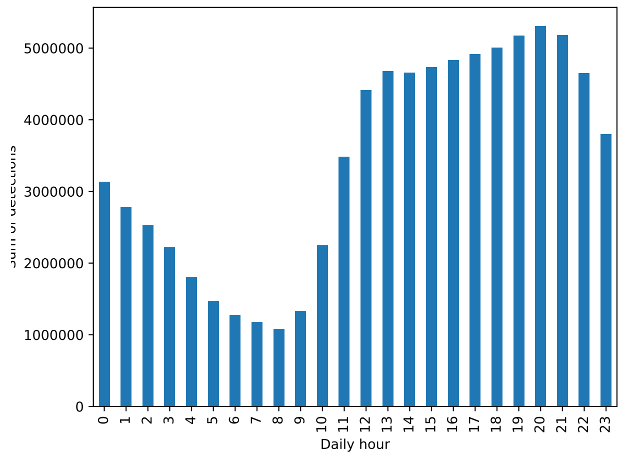
The figure 2 presents general counts and cumulative distribution functions of detections. A few comments :

- Figure 2a shows daily counts of detections, summed over all the days during the analysis period. Weekdays are more active, with the maximum being attained Thursday ;

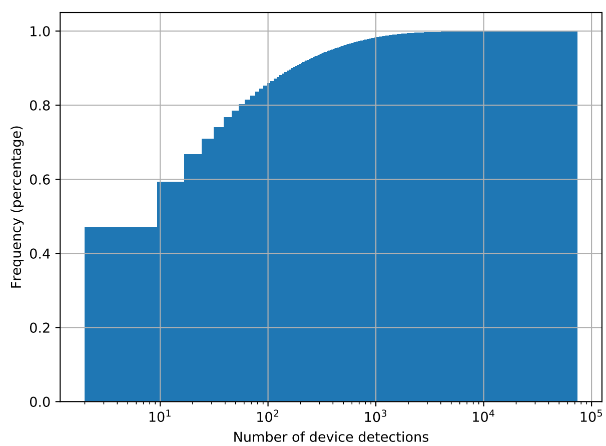
- Figure 2b shows hourly counts of detections, again summed over all the days during the analysis period. It shows a 5 fold difference in the volume of detections in the AM peak-hour (approx. 1 million detections) to the PM peak-hour (approx. 5 million detections)



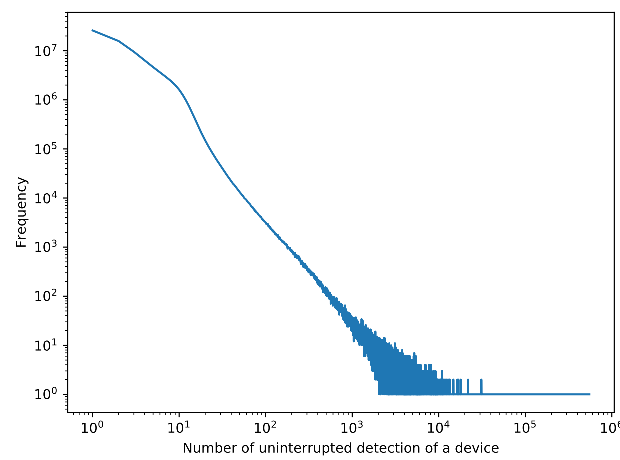
(a) Total detections per day



(b) Total detections per hour



(c) Distribution of device detections



(d) Distribution of continuous device detections

FIGURE 2 : Basic statistics of BT detections.

This suggests an imbalance in statistics during the day, which is contrary to the observed phenomena for vehicles; while this varies from region to region and from seasons during the year, the PM peak is usually associated with higher volumes, ranging from 110% to 150%, but far from 500% ;

- Figure 2c shows the distribution of the number of device detections over the BMS network. The median value is 11 detections per device over the span of the database. More than 80% have about 50 detections and 90% have roughly 200. Please note that these numbers excludes the number of uninterrupted detections in a BMSs detection zone;

- While this paper focuses on trips and movement, the BT detection database provides an information about the number of uninterrupted detections of a device in the BMSs detection zone. As is shown in the loglog graph in 2d, this number is usually low (1 detection of a device for a given tracker 27,361,640 times), but it can also get quite high, the maximum being 820,493 uninterrupted detections of a device over a 2 month period, representing a detection roughly every 11 seconds. As can be seen, the tracking of still objects is far from marginal, and can represent a host of devices from printers to tablets.

Finally, the graph shows that extreme values are an important part of the database and should not be considered “outliers”. Before determining filtering methods, one must understand the underlying phenomena generating the detections and decide if they are of interest to the analysis.

The vicinity of tracking devices are presented in figure 3. It counts (shows the intensity of the counts of) the number of devices having an origin at BMS i on the x-axis (e.g., the BMS where a device is detected) and a destination at BMS j on the y-axis (e.g., the next BMS where a device is detected). Please note that these detections are considered successively for a device and can be considered as a subset of trips. Origins and destinations for trips will be presented in section 6.1. To consider :

- the names represent integers, where similar integers are BMSs "close" to each other (but not always) ;

- the intensity of neighborhood can be observed and confirms the position of surrounding BMSs ;

- the symmetry in the figure represents the geometric nature of the road-network and the bi-directionality of streets ;

- the intensity of the heatmap depends on possible connections between pairs of BMSs (sometimes inaccessible because of one-way streets, etc.), but also of the underlying phenomena : device "flows". BMSs on busy arterials are more visible in the graph ;

- the diagonal (the next BMS is the same as the current BMS) represents 22% of all neighborhood points, limiting the intensity of neighborhood BMSs.

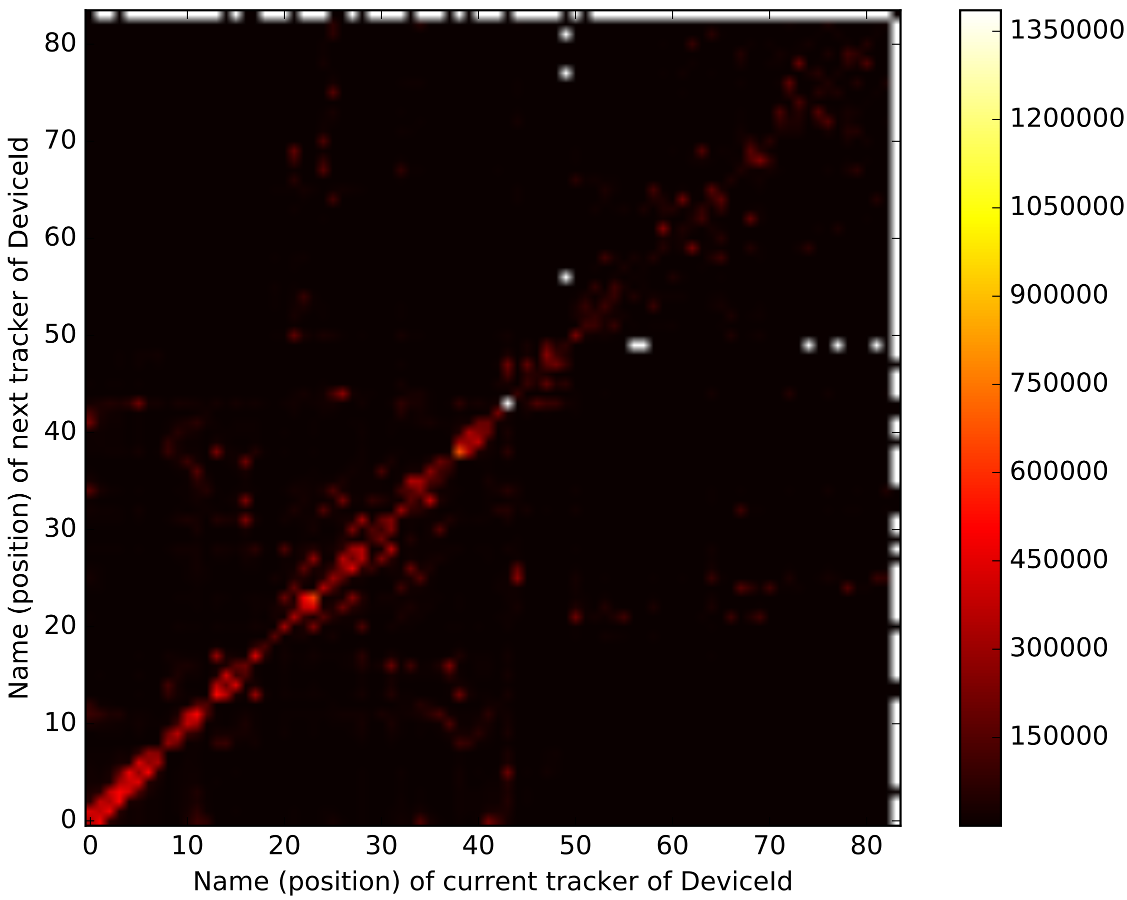


FIGURE 3 : This image shows the frequency heatmap of the next BMS from a given BMS. The integers X and Y axis represent a projection of the position of the BMS

METHODOLOGY

To explore the data, characterize road-network conditions and understand trip patterns, the following methodology is proposed. This paper does not dwell into the technical issues of managing a database of several gigabytes of disk space and the different methods used to manage memory and speed up processing. A multi-threading python-based framework was developed for this project.

Traffic basics

To derive basic traffic statistics and infer trips, the following variables must be defined and calculated :

- **Time** : the time between detections is defined as the difference between the last detection of the previous entry and the first detection of the current entry : $FirstDetection_i - LastDetection_{i-1}$. Because BMSs have influence zones that can overlap, this value can be negative. Indeed, the last detection at BMS $i-1$ can occur while the device is still being detected at BMS i if their influence zone overlap. Because of this issue, varying definitions are possible, such as $FirstDetection_i - FirstDetection_{i-1}$ or $LastDetection_i - LastDetection_{i-1}$. These values should not be negative and offer a different perspective on travel times.

- **Distance** : the distance between two trip detections is considered the same as the distance between the network distance of its two BMSs. This definition ignores the influence zones of the BMS's which reduce the true distance between detections. Also, the influence zone varies under different traffic, spatial and temporal conditions (*2*), (8). Moreover, because BMSs specifications were not supplied to the authors, no hypothesis were made which would introduce further bias in the statistic ;

- **Speed** : The speed has its usual definition, with instantaneous successive detections filtered out. Negative speeds are taken into account to understand trip chains, but are filtered out for cumulative statistics of trips.

Trip characteristics

A trip is defined as having specific characteristics in both time and space. In time, a trip is defined as a series of pings of a device on network devices - not necessarily consecutive. TPANA ends trips after a 1800 seconds idle time between detections (13) on a predefined pair of BMSs. For the purpose of this paper, we continue with the same threshold approach as proposed by the service provider, but use an alternative cut-off value: based on the author's experience, travel-time in Montreal in this section is usually shorter than 45 minutes, North to South, irrelevant of mode, in general traffic conditions. Having a higher threshold improves the probability of connecting two detections anywhere in this network.

Furthermore, successive detections of BT devices that are longer than 45 minutes (2700 seconds) are considered different trips. Of course, a combination of successive detections of a device that are shorter than 45 minutes are combined to form a trip chain that can total longer travel times.

In space, a trip is defined as a device detection on at least two different BMSs. Intra-zonal trips, or trips with single BMS detections, cannot be classified as such because it cannot be asserted weather the tracked device is still or not. This definition includes possible travel times to account for the users "absence" from the network in successive BMS.

EXPERIMENTAL RESULTS

This section presents the exploration experimental results obtained. Results are discussed as they are presented.

Basic Trips

There are 17,060,134 trips that correspond to the definition given in 5.2, which roughly corresponds to 26 thousand daily trips. Note that trips included exclusively in the BMS's zone of influence are not counted in these trips. With a smaller threshold than 2700 seconds, this number is higher. Compared to typical counts in the sector (*16*), the penetration was found to respect typical penetration rates (2), for example with an approximately 15% sampling estimation on Papineau Avenue, a principal North-South arterial connecting the downtown core all the way to the residential areas and suburbs, principally by car.

Further more, figure 4 presents generalized statistics. A few comments about the presented values :

- 4a shows the trips' duration distribution whose median duration is about 500 seconds (approximately 9 minutes) while 75% of trips last under 2000 seconds (about 32 minutes). However, a significant proportion is longer. Assuming there is an underlying portion of the trip that is not on the BMS grid-network, these travel-times are consistent with other local studies (17). Furthermore, the maximum length trip is of 780,616 seconds, or 9 days;

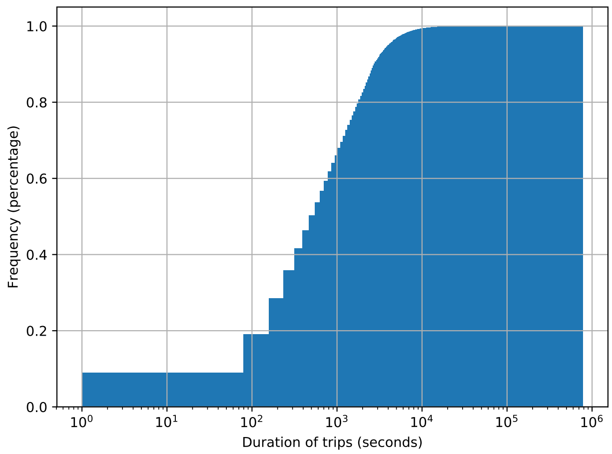
- 4b shows the distance distribution per trip whose median is just under 2400 meters. However, as with the previous statistics, the distribution has considerable values in its upper spectrum, with 10% of vehicles traveling more than 10 kilometers over this network, the highest value being 366 kilometers ;

- 4c shows the number of trips per BT device. It shows that 40% of BT devices make only one trip during the study period, with the median being 2 trips while 75% make less than 9 trips. Anecdotally, the maximum number of trips is 16,783 which represents an average of 26 trips per day ;

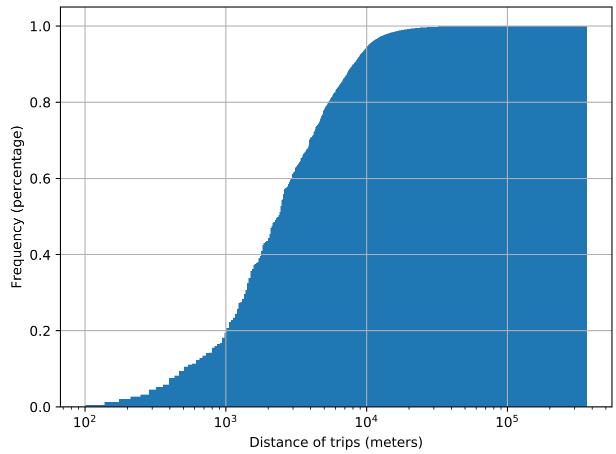
- 4d shows the number of BMS devices per BT trip. The median of this distribution is 2 devices and its 75% quantile is 3, its 90% quantile is 6, meaning that trips span over a few BMS and that devices are not tracked exhaustively over the network. The maximum number of BTs per trip is 2034.

As with the successive detections analysis presented in 4, extreme elements (not to be confused with “outliers”) are an important part of the presented distributions. Again, it is clear that the different statistics are composed of different distributions that are generated by different phenomena. However, these underlying generating events are not easily identified without additional information.

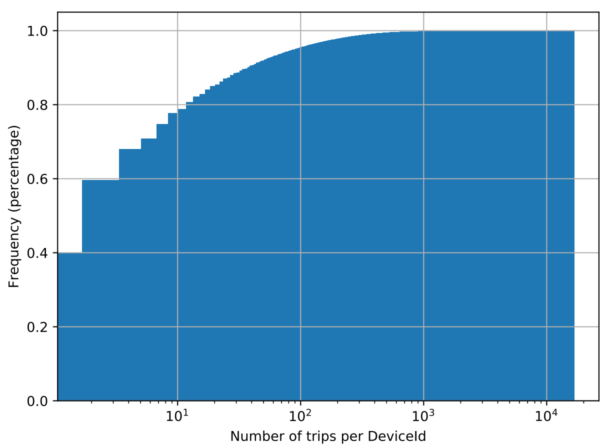
Finally, figure 5 shows the intensity of origin-destination pairs of trips, where as figure 3 showed the intensity of origin-destination pairs on successive BT device detections. Like figure 3,



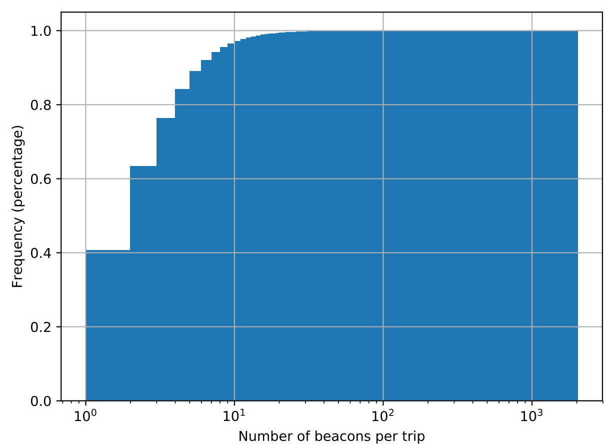
(a) Trip duration distribution



(b) Trip distance distribution



(c) Number of trips per device



(d) Trip chain-length distribution

FIGURE 4 : Basic trip cumulate distributions functions : durations, distances and compositions.

the diagonal is highlighted, indicating that a BT device's trip's destination is not far from its origin, but less so. More interestingly,

- The graph is still highly symmetrical, indicating the presence of pendulum trips if the distance is long and of bi-directional street segments if the distance traveled is short ;
- The graph displays pairs that have no connections, in particular for BMS 49, which, interestingly, is an exit ramp of the Port of Montreal, which is a container port with over 2500 daily truck trips (*14*). This would indicate that most trips which originate from this node probably exit directly the study area and leave the island on the highway network, which is not tracked ;
- The graph shows more scattered points, indicating complex trip patterns ;
- The graph displays clear clusters, for example, the OD square of 0-8 BMSs, which track

Sherbrooke Street, a busy city arterial, or the OD square 43-50, which covers the Port Of Montreal. More generally, the underlying 10-40 square, representing the Rosemont and parts of Ahuntsic and Montreal-Nord neighborhoods shows intra-neighborhood trips ;

- However, for the majority of trips, these clusters seem to have a destination relatively close to their destination, a few BMSs away, which may or may not correspond to an underlying trip. A filtering method could be applied to define origin and destination zones and merging several BMSs. While this would not change the fact that BT trips have an origin and destination that are closely linked (on a street segment such as Sherbrooke Street), the proposed heatmap would more thoroughly highlight the properties of the underlying vehicular trip by filtering out trips that are only partially seen on the BMS network.

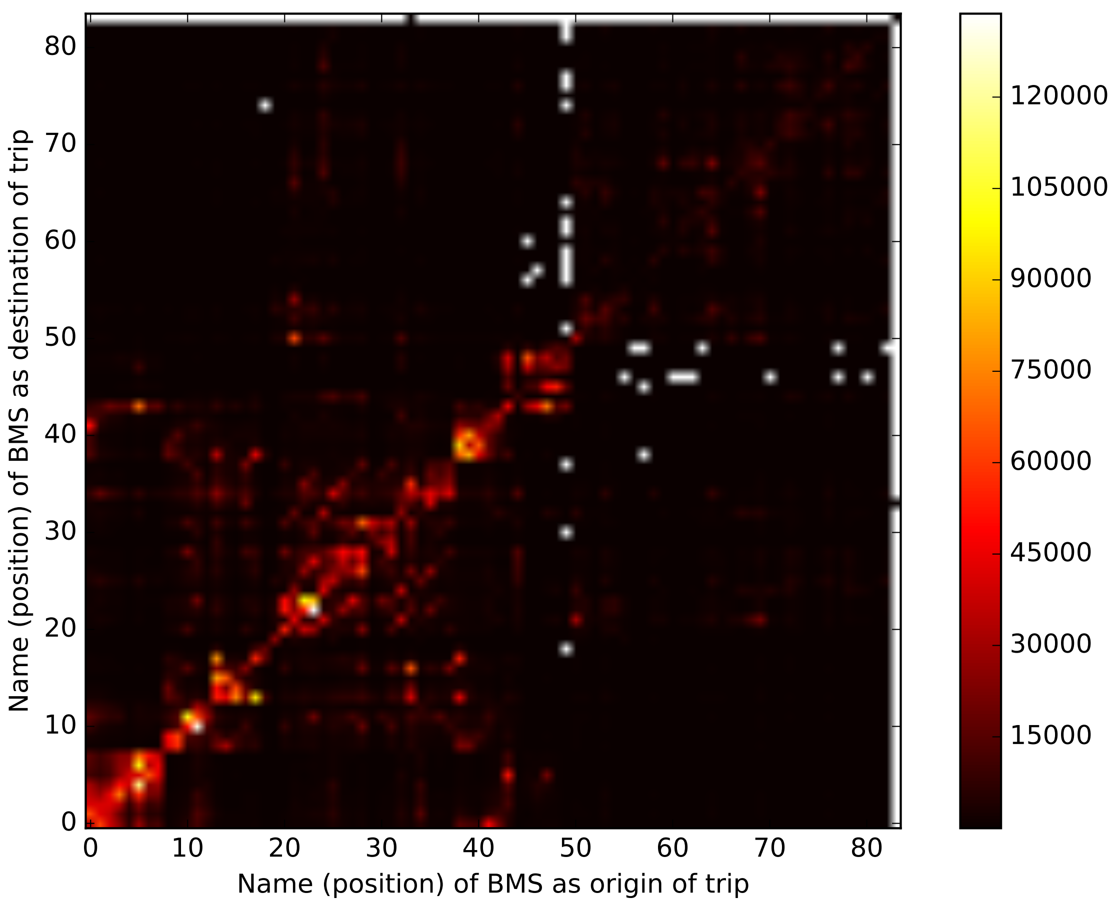
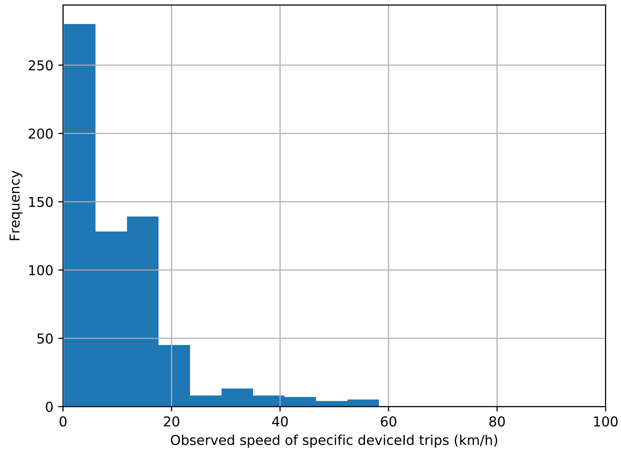


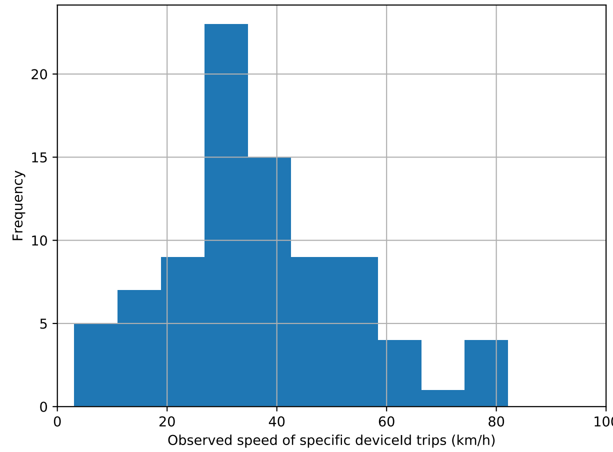
FIGURE 5 : This image shows the frequency heatmap of the destination BMS from a given BMS origin for all BT trips. The integers X and Y axis represent a projection of the position of BMSs.

Individual speeds

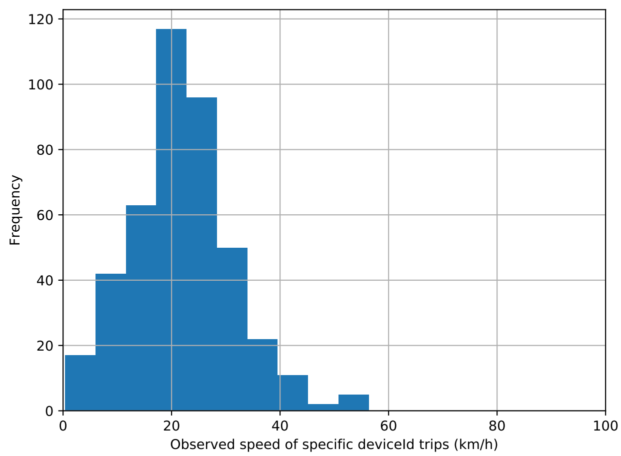
While the mode of the device cannot be asserted directly from the data, the analysis of the speeds of the trips, presented in figure 6, can shed some light on how a particular device traveled through the network over a year and a half. Several observations can be made about the figures :



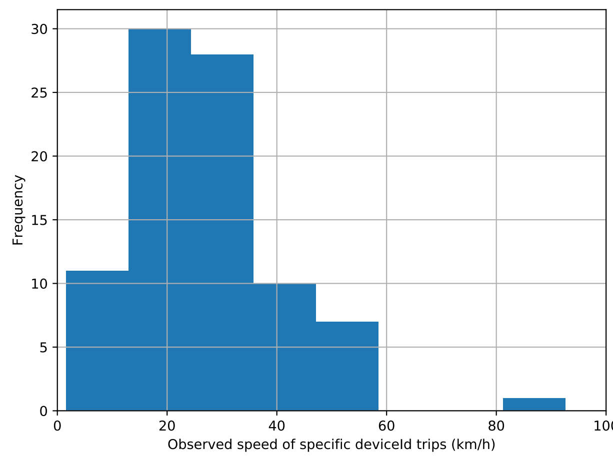
(a) Device of a pedestrian?



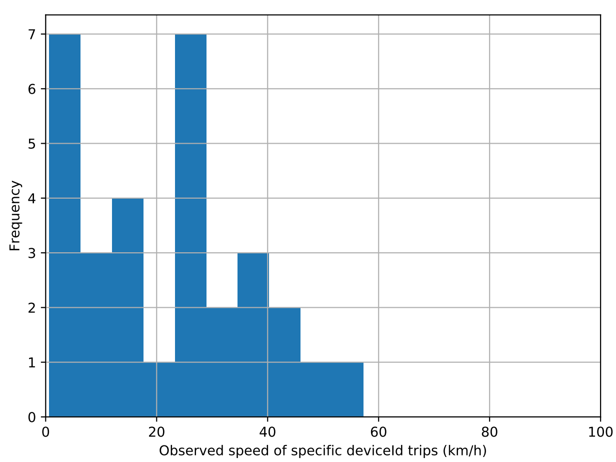
(b) Device of a driver?



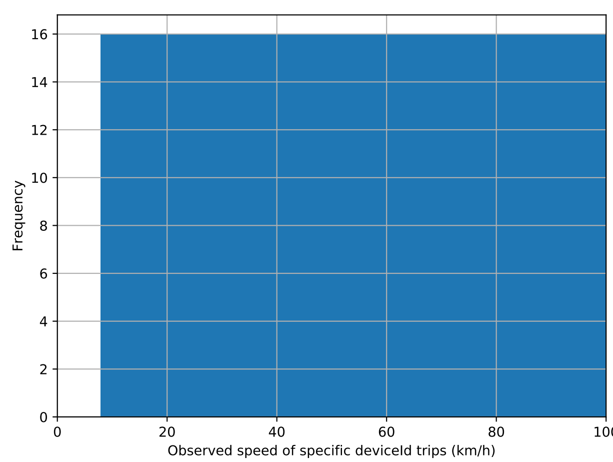
(c) Device of a cyclist?



(d) Device within a bus?



(e) Multi-modal device?



(f) Extreme regularity is suspicious...

FIGURE 6 : Distributions of speeds of 6 different BT devices over the span of the study

- The distributions presented here vary in shapes and corresponding statistical measures (mean, spread, skewness and kurtosis) and represent a multitude of different trips. However, other BT devices have similar speed distributions that can be intuitively matched to a subset of these;

- The number of samples distribution is different for every BT device, but these examples have been chosen because they are sufficiently large ;

- The mean of the distributions vary from a approximately 5 km/h (6a) to 40 km/h (6b) covering mode speeds from walking to driving;

- Most distributions have a clear mode (in the statistical sense), while 6e is clearly bi-modal (in the traffic engineering sense). Considering the speeds, can these trips be associated with a main or principal mode of transportation (walking, public transit, etc.)? As stated earlier, most BT studies focus on a particular mode (see for example, (*18*)) ;

- As can be seen in 6f, suspect distributions needs careful attention before deciding if the underlying phenomena causes a side-effect in the statistics and can be considered, not an outlier, but a different source of data than traffic.

DISCUSSION

As the data description (section 4) and the trips description (section 6.1 show, there are several issues to analyzing BT data.

Elapsed time definition : interval overlap

As explained in section 5.1, there are several ways to define the elapsed time between successive BT detections. These definitions introduce a bias that varies under different flow conditions (*9*). In high-speed highway conditions, it is assumed that the definition has little impact on speed or travel-time statistics, which is not the case for generally dense urban over-saturated conditions as presented here. While this paper has not precisely evaluated this bias, it is assumed that speeds are generally higher than what would be measured because the time interval chosen is the smallest amongst possible values. Further investigation is necessary to understand this link.

BMS influence overlap

A related issue is the distance between BMSs. For simplicity, the position of the detected device is associated to the BMSs position. However, this is an approximation and introduces a bias in the calculated statistics. Indeed, BT devices can be detected a few hundred meters before the BMSs location. As stated in (*8*), different operational characteristics have an impact on the BMS zone coverage. In the analyzed database, BMSs in the Quartier des spectacles are separated by 100 to 250 meters, shorter than the vendor's specifications. In fact, TPANA (the city's service provider) recommends spacing of BMS devices by at least 1 kilometer in an urban setting (*19*).

Because of this and the proposed elapsed time definition, negative times can occur when a BT device is detected at the downstream BMS before it has stopped being detected at the upstream BMS. As a rule, the most direct solution to this issue is to merge BMSs in the range of 1 kilometer and to consider them has a group.

CONCLUSION

This paper explored the value of Bluetooth device detections over the span of a year and a half on a significant portion of the Island of Montreal, Canada. Over 80 BMS devices were installed in the central and densely populated portion of Montreal (Plateau-Mont-Royal, Ahuntsic, Rosemont, etc.) around the subway's central orange line, around truck-accessible facilities in the Port of Montreal, on principal parts of Sherbrooke Avenue (a major East-West axis), and finally, the Quartier des Spectacles in the city's downtown core. The database contains over 85 million observations emanating from 1.3 million Bluetooth devices over 640 days. With the methodology described in this paper, over 17 million trips were found (approx. 26 thousand per day), with travel-time, travel-distance and origin-destination pairs intuitively corresponding to expected results (17). Figure 4d showed that the median of the distribution of BT devices per trip is 2 and its 75% quantile is 3, meaning that trips span over a few BMSs and hence, that devices are not tracked exhaustively over the network. Moreover, extreme values are an important part of the presented distributions, as shown in the paper's log graphs. Further investigation is necessary to understand the underlying generating events which are not easily identified without additional information. For example, if other devices than "traveling" devices are found or if devices included in buses are over-represented (*3*), they must be taken into account.

Moreover, this paper explored speed distributions of trips and showed that different families of distributions could point towards different modes of vehicular trips. Also, multi-modal distributions have been found. Further validation is necessary.

Finally, there are technical issues associated with deriving traffic-related statistics from Bluetooth. For example, understanding the length of detection zones of BMS devices, and the parameters influencing this zone, helps limit the bias when calculating travel-times and speeds from BMS to BMS, especially in an urban area.

Next steps

The abstract nature of the Bluetooth data lets us only hypothesize on trips and their characteristics. Other data sources will be used to validate the paper's hypotheses and its conclusions. In particular, in the context of the Smart City and with the objective of optimizing travel on its network, the City of Montreal invited its citizens to participate in a vast study on travel by downloading a new mobile application: MTL Trajet (*20*). The data was collected for a period of one month, from October 17th to November 17th 2016, overlapping Bluetooth tracking. Through the application, the user anonymously records his trips using her smart phone, specifying the trip purpose, mode and the origin and the destination of the journey. A total of 11,433 downloads were completed during the experiment. In the context of our research, the data will be used to validate the Bluetooth data hypotheses and compare results of trips and their definition. A pairing method is proposed to match Bluetooth trips to GPS trips. Applying machine learning algorithms, predictive models can be built for route, mode choice and other traffic related models, building on the works of (*6*).

REFERENCES

[1] Kasten, O. and M. Langheinrich, First Experiences with Bluetooth in the Smart-Its Distributed Sensor Network. In *Workshop on Ubiquitous Computing and Communications PACT*, 2001.

[2] Bhaskar, A. and E. Chung, Fundamental understanding on the use of Bluetooth scanner as

a complementary transport data. *Transportation Research Part C: Emerging Technologies*, Vol. 37, 2013, pp. 42–72.

[3] Bhaskar, A., L. Kieu, M. Qu, A. Nantes, M. Miska, and E. Chung, Is Bus Overrepresented in Bluetooth MAC Scanner data? Is MAC-ID Really Unique? *International Journal of Intelligent Transportation Systems Research*, Vol. 13, No. 2, 2015, pp. 119–130.

[4] Martin, J., T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. Rye, and D. Brown, A Study of MAC Address Randomization in Mobile Devices and When it Fails. In *Privacy Enhancing Technologies Symposium*, 2017.

[5] Kieu, L., A. Bhaskar, and E. Chung, Bus and car travel time on urban networks : integrating Bluetooth and bus vehicle identification data. In *25th Australian Road Research Board (ARRB) Conference*, 2012.

[6] Barceló, J., L. Montero, L. Marqués, and C. Carmona, Travel Time Forecasting and Dynamic Origin-Destination Estimation for Freeways Based on Bluetooth Traffic Monitoring. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2175, 2010, pp. 19–27.

[7] Tsubota, T. and T. Yoshii, An Analysis of the Detection Probability of MAC Address from a Moving Bluetooth Device. In *International Symposia of Transport Simulation (ISTS) and the International Workshop on Traffic Data Collection and its Standardization (IWTDCS)*, 2017, Vol. 21, pp. 251–256.

[8] Abedi, A. N., A. Bhaskar, E. Chung, and M. Miska, Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, Vol. 60, 2015, pp. 124–141.

[9] Malinovskiy, Y., U. Lee, Y.-J. Wu, and Y. Wang, Investigation of bluetooth-based travel time estimation error on a short corridor. In *Transportation Research Record: Journal of the Transportation Research Board*, 2011, 11-3056.

[10] Haghani, A. and Y. Aliari, Using Bluetooth Sensor Data for Ground-truth Testing of Reported Travel Times. In *Transportation Research Board Annual Meeting Compendium of Papers*, 2012.

[11] Malinovskiy, Y., N. Saunier, and Y. Wang, Pedestrian Travel Analysis Using Static Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2299, 2012, pp. 137–149.

[12] Malinovskiy, Y., *Mobile Device Identifier Data Collection and Analysis for Transportation Intelligence Purposes: Applications, Uncertainty and Privacy*. Ph.D. thesis, 2013.

[13] *Real-Time Travel Time Executive TPANACOLLECTOR and TRAKKER User Manual*. TPA North America inc., 2015.

[14] Lord, S., CargoM: de précieuses données pour le transport routier. *La Presse*, 2017.

[15] *NETTendances 2014*, 2015.

[16] *Montreal Open Data Portal*. `http://donnees.ville.montreal.qc.ca/dataset/comptage-vehicules-pietons`, 2017, accessed: 2017-07-30.

[17] *Évaluation des coûts de la congestion routière dans la région de Montréal pour les conditions de référence de 2008*. Ministère des Transports du Québec, 2014.

[18] Malinovskiy, Y., Y.-J. Wu, Y. Wang, and U. K. Lee, Field Experiments on Bluetooth-Based Travel Time Data Collection. In *Transportation Research Board Annual Meeting Compendium of Papers*, 2010, 10-3134.

[19] Bruce, R., Bluetooth Traffic Monitoring. In *ITS Canada, ATMS Traffic Detection*, 2013.

[20] *Montreal Trajet*. `https://ville.montreal.qc.ca/mtltrajet/`, 2016, accessed: 2017-07-30.