Titre: Title:	Comparing multiple data streams to assess free-floating carsharing use
Auteurs: Authors:	Grzegorz Wielinski, Martin Trépanier, Catherine Morency, & Khandker Nurul Habib
Date:	2018
Type:	Article de revue / Article
Référence: Citation:	Wielinski, G., Trépanier, M., Morency, C., & Habib, K. N. (2018). Comparing multiple data streams to assess free-floating carsharing use. Transportation Research Procedia, 32, 617-626. https://doi.org/10.1016/j.trpro.2018.10.011

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 32 (2018) 617–626

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Steering Committee for Transport Survey Conferences

Comparing multiple data streams to assess free-floating carsharing use

Grzegorz Wielinski[a]*, Martin Trépanier[a], Catherine Morency[b], Khandker Nurul Habib[c]

[a] *Polytechnique Montréal, CIRRELT (Interuniversitary Research Centre on Entreprise Networks, Logistics and Transportation), Department of Mathematics and Industrial Engineering, P.O. box 6079, station Centre-Ville, Montréal, Québec H3C 3A7, Canada*

[b] *Polytechnique Montréal, Mobility Chair on sustainability & CIRRELT, Department of Civil, Geological and Mining Engineering, P.O. box 6079, station Centre-Ville, Montréal, Québec, H3C 3A7, Canada*

[c] *University of Toronto, Department of Civil Engineering, 35 St George Street, Toronto, ON M5S 1A4, Canada*

Abstract

Passive data streams are a great alternative to traditional travel surveys to assess the use and the change in behavior. Because of accessibility issues, researchers have harvested free-floating carsharing (FFcs) web-services in order to estimate the spatiotemporal demand. This paper presents the comparison of multiple data streams in the assessment of trip type distribution for FFcs service. While a full dataset of GPS traces may be considered a good approximation of the ground truth, harvesting of origin-destination data seems to estimate correctly the general trend of certain trip type distributions, while for other trip type estimations, a more extensive set of data is needed (member & stopover information) to fully assess it.

Keywords: Passive data streams; behavioral change; free-floating carsharing

1. Introduction

Carsharing constant increase in popularity over the last two decades makes it a vital part of the mobility cocktail in cities around the world. On the one hand, station-based carsharing (SBcs), most widely used carsharing form (Shaheen et al. 2016), proved itself as a sustainable mobility means with societal, environmental and personal benefits (Millard-Ball et al. 2005). On the other hand, free-floating carsharing (FFcs), mostly popularized in the late 2000s

* Corresponding author. Tel.: +1-514-340-4711 ext. 4911; fax: +1-514-640-4173.
E-mail address: grzegorz.wielinski@polymtl.ca

10.1016/j.trpro.2018.10.011

with the arrival of large car manufacturers (car2go, DriveNow) (Shaheen et al., 2013), disrupted the market by adopting a different business model: cars are not anymore constrained to a station, but within a service area, allowing members to perform one-way trips without reservation *a priori*.

With free-floating services, carsharing operators (CSOs) usually offer the full fleet visibility to its members over a web-based service or an application. This makes sure a user can locate the nearest car and even lock the vehicle for a short duration while he or she processes to reach for it. Carsharing operators being protective of their data, researchers have seen the opportunity to estimate the level of activity of one carsharing service by “harvesting” or “scraping” a web page, or simply querying an API (application program interface, see https://github.com/car2go/openAPI for one example of a CSO web service). One may derive this periodical storage of vehicle’s position into actual estimated trips. Fields of study like vehicle relocation, spatiotemporal dynamic, user behavior and service benchmarking are one of the many applications a systematic carsharing data harvesting process can assess. While multiple studies employ this kind of data, inevitably some limitations arise.

Communauto, the oldest carsharing operator in North America, operates in Montreal since 1995 a station-based service, but since mid-2013 the operator introduced a free-floating service. Since then, the FFcs service has been the object of numerous expansions, in terms of fleet size and type, and also in terms of service area coverage. Strong with a research partnership, multiple passive data streams were made available: 1) transactional (booking) data, 2) detailed trips GPS data, and 3) web-collected (OD) data. This enables the assessment of the user behavior (and its change over time) of free-floating carsharing members, but also the comparison of these data streams. Free-floating settings allow one-way trips, but they also make possible standard round-trips, making it interesting to assess the different dynamic inside a free-floating environment. There fold, this paper contribution is threefold:

- RQ1: What is the trip type distribution inside a FFcs service?
- RQ2: Are we able to assess any behavioral change over time system-wide?
- RQ3: Are there any limitations respective to the used data source? Is the web-based data collection method a good proxy to emulate the results gotten by a GPS dataset?

This paper is structured in five parts. First, a background (section 2) of recent literature focuses on the exploitation of web-based OD data harvested in different contexts, but also a quick review on trips type assessment for the three mentioned passive data streams is undertaken. Then, the case study and the methodology (section 3) for processing these various data sources are presented. The trip type distribution and the underlying hypothesis for each data streams compose the main aspect of the fourth section, while the discussion section looks to compare the strengths and limitations of the respective methods (section 5). Concluding remarks end the paper (section 6).

2. Background

A literature review has been carried out on studies exploiting web harvested data. This process consists to steadily capture multiple snapshots of the available vehicles over a period of time. When one vehicle at period $t-1$ is made available but no more at period t, one may deduce that it is used by a member to perform a trip (a vehicle disappears from the pool of available vehicles when it is used or reserved by a member). Inversely, when the vehicle appears back, the trip theoretically ended and one may deduce the distance between the origin and the destination and also the time taken to perform the trip. Other data as the fuel or energy level delta may also be harvested in the same process. Because users may perform trips with stopovers, the driven distance may not be deducted accurately.

A literature review of pertinent studies revealed 15 papers. They can be split in four main categories: 1) GPS-related studies, 2) transactional-related studies, 3) web-based data used as a supporting role to the study, and 4) web-based data used as a central component of the study.

2.1. GPS-related studies

First, GPS-based studies look to leverage rich datasets of carsharing trips. Leclerc et al. (2013) and Hui et al. (2017) look at the Canadian and Chinese setting. Usually, GPS passive data streams require researchers to develop an algorithm to split the full transaction trace into trips and stops. Both mention the importance to assess good stop-time criteria to limit false positives. This could occur in the case of traffic congestion where the car could stop, but in reality, the member didn't perform a stop to reach its activity, for example. Clustering techniques are employed by both studies, resulting in a transaction classification by trip/stop length and frequency. Lopez et al. (2016) on their end evaluate the data quality of GPS and transactional data through carsharing operator data. Results show that in an urban setting, 13% of travelled distances are not recorded or simply not accounted for when aggregating the total trip length. Finally, Wielinski et al. (2015) applied data fusion techniques with GPS and survey data to split trips according to the carsharing members' trip purpose and the travelled distance.

2.2. Transactional-related studies

Studies leveraging booking data to assess the user behavior in carsharing are plentiful. For our application, this category regroups two studies that look to assess the proportion of round-trips in a free-floating carsharing setting based on transactional data and vehicle position at origin and at destination. Schmöller et al. (2014) split in a table the transactions made in two cities in Germany across two main axes: trip length and OD air-line distance. A short air-line distance between the OD implies that a user either 1) came back near its original departure location, or 2) made a short trip. By looking at short air-line distance and long total trip length, one may assume a higher likelihood of round-trips. Looking at air-line distance shorter than 800 meters and travelled distances of 3 km and more, the authors estimate a proportion of round-trip between 8-11%. In the Montreal case (Wielinski, 2014), the FFcs service is integrated in conjuncture with a station-based one. It is estimated, with a similar methodology used by Schmöller et al. (2014), a 20.5% proportion of round-trips.

2.3. Web-based harvesting study as a complementary role

The third category mainly looks at studies that harvest free-floating vehicles position, but the article's data exploitation is not primarily related to this paper first interest. Most of them relate to fields like vehicle imbalance and relocation techniques, spatiotemporal dynamic and user behavior, or decision support related tools (Heilig et al., 2017; Schulte et al., 2015; Willing et al., 2017). In other applications, they are used as a way to complement an existing operator dataset (Wielinski et al., 2017). OD data are useful for spatiotemporal dynamic studies because the data requirement is minimal: an origin and a destination paired with their corresponding time stamps, independent of how the member uses the vehicle. Vehicle relocations made by the operator are still captured when harvesting web-based services, creating a bias, but it is usually assumed low (Boldrini et al., 2017). Points of interest (POI) datasets may also be used in conjunction with OD data in the modeling effort to assess the effect on the demand for certain part of the service area. In the case of Wielinski et al. (2017), they used some web-scraped data to complement their existing booking dataset (provided by the operator) to get cars' energy level, otherwise not available in the original dataset.

2.4. Web-based harvesting as the main part

The last category comprises studies focusing on the exploitation and the assessment of OD data (Boldrini et al., 2017; Ciociola et al., 2017; Wang et al., 2017) and/or to use the ability to collect larger scale datasets to compare multiple cities/operators in the same time frame (Boldrini et al., 2017; Kortum et al., 2016, Tyndall, 2017). While the collection method is similar amongst them, some particularities occur when real-time insights are required (Ciociola et al., 2017). Authors also mention a limitation for web harvested data. While a vehicle is remotely rented, it immediately disappears from the pool of available vehicles, but it takes between 1 and 4 minutes to reappear back, which would inevitably skew the trip duration distribution.

2.4.1. Trip length based on routing engine calculations

An observed trend to account for the lack of actual trip length (both in terms of distance and duration) is to query a routing engine (like Google Directions API) with the trip's start and end coordinates. This will generate the shortest/quickest path between both coordinates (with or without congestion consideration). Studies like (Boldrini et al., 2017; Ciociola et al., 2017; Wang et al., 2017) use this technique, but assess the data in different ways. For example, Ciociola et al. (2017) compare the shortest driving length in terms of duration to the rental total time. As the rental time takes into account the time to access the vehicle and park it, they found that 12.1% of the rentals are actually quicker than the routing estimate, hinting to a pervert effect of the time-based pricing policy; members may be willing to drive faster to minimize the rental cost. They also compare rental times to transit times with the Google API. Members seem to prefer carsharing to compensate longer transit trips, but not too much for shorter ones. Wang et al. (2017) paper focus primarily on assessing the complementarity of carsharing with public transit. They also leverage routing engines to collect driving and transit times for a similar trip. Compared to Ciociola et al. (2017), they subset their dataset to keep as much as possible only one-way trips. As Becker et al. (2017) mention, a quarter of all free-floating trips began or ended outside the service area, meaning they would be part of round-trips or including a stopover. Wielinski (2014) and Schmöller et al. (2014) also assess a proportion of assumed round-trips. Boldrini et al. (2017) creates a classifier (with estimated driving times and the distance between the OD) to split the rentals in one-way, two-way and one-way trips with a stopover. They estimate the majority (79-90%) of trips being one-way with stopovers. With this in mind, Wang et al. (2017) perform three data subsets on their sample to consider only one-way trips. While these data cleaning steps would effectively reduce the number of undesirable trips, the employed thresholds are still subjective.

2.4.2. Distance-based by fuel consumption

Another technique used to approximate the driven distance is to take into account the fuel consumption of the vehicle and based on the vehicle energy consumption derive an estimated driving distance. Boldrini et al. (2017) and Heilig et al. (2017) use this technique. In the case of Heilig et al. (2017), the studied dataset comprised a fleet of homogenous electric vehicles, while for Boldrini et al. (2017) the average battery and fuel consumption was fetched from a specialized website. While this technique allows effectively to estimate a probable driven distance, the fuel or power consumption varies greatly on factors like driving characteristics (Wang et al., 2008) and thus may incorporate biases in the estimates.

2.4.3. Large-scale data harvesting

Large-scale benchmarking is made possible by easily horizontally scaling the data collection for multiple cities or CSOs for the same time frame. Kortum et al. (2016) p roceed to publish the first study of this magnitude. They proceed to compare the evolution of multiple CSOs (34 different cities) across a 5-year timespan. They discuss growth rates, vehicles use rates and also perform regression techniques to assess factors influencing services' daily bookings. Boldrini et al. (2017) continue on the same path by comparing 10 different cities for a two-month period. Their unit of analysis is more centred around the user than the city, as in Kortum et al. (2016). Tyndall (2017) looks at the social equity perspective. The author collected data for 6 months in 10 different American cities. With data on U.S. census tracts, they can evaluate neighborhood types where vehicles are more agglomerated, resulting in a higher vehicle presence in neighborhoods of naturally socially more advantaged persons.

3. Methodology

3.1. Case Study

Communauto has been operating a station-based carsharing service since the mid-1990s in the Montreal area. Since June 2013, it also operates a free-floating service, called Auto-mobile, in the central part of the city. The free-floating service started its operations with a service area (SA) of ~8 km^2 and 24 electric vehicles (EV), but quickly expanded the SA and the car fleet with mostly hybrid vehicles (HV) over the years. As of June 2017, the 85 km^2 service area had 5 distinct expansion phases and operates more than 605 vehicles. Communauto integrated both the SBcs and the FFcs services into a unified ecosystem. Station-based package owners can access the free-floating vehicles with no additional fee or administrative action. Some pricing mechanisms allow SBcs members to use the FFcs service and pay the same rate as they would have with the station-based service. Free-floating members with no station-based package cannot access the station-based service and have to subscribe to a monthly plan if they wish so. This will give them access to a bigger car fleet, but also to better overall rates.

3.2. Information System

3.2.1. Transactional Data

Two main datasets are made available by the carsharing operator and another one is generated inside the research team (web harvesting). First, the main dataset contains information on all trips made by Communauto members from January 2008 to May 2017. This dataset has various information as the start/end timestamp, travelled distance (odometer delta), member id (anonymous), gender, age, home location (postal code precision), package information, vehicle types (HV/EV), station location, and station capacity.

3.2.2. GPS Data

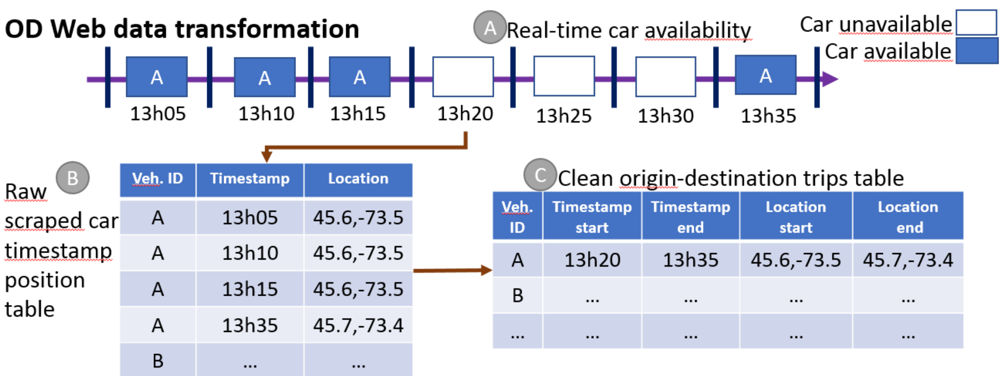
More accurate details on trips are made available via the full GPS dataset of FFcs trips (~12 seconds precision). More than 33,750,000 observations are part of this dataset that represents the whole 2015 period. An algorithm is developed to split the full transaction in one or multiple trips (if one or more activities are detected). The algorithm is based on the vehicle status and speed to achieve it. In all, 238,553 valid transactions are split. This dataset is useful to assess how members are using the free-floating service (the SBcs service is solely used as a round-trip service). While GPS data has been looked over to replace traditional data collection (Wolf et al., 2001), this dataset has its own limitations (Lopez et al., 2016) and may therefore not be the ground truth. Nonetheless, this is the most precise assessment of a member user behavior with a vehicle that is available to us.

3.2.3. Web-Based Data

While GPS coordinates are significantly richer, they are harder to maintain and their availability is limited by a third party. To overcome those limitations, a process to extract the position of all free-floating carsharing vehicles is set up via a call to a web API (application program interface). While this method is less precise and subject to the web service uptime (Kortum et al., 2016), it allows to deduce trips in an origin-destination format with a rounded timestamp to the nearest 5 minutes. This method is also subject to operator-based trips for relocations and thus biased. Nonetheless, this dataset allows to collect precious insights on how free-floating members use the service (908,056 observations).

Vehicle timestamp position data from Communauto web service is recorded every five minutes. When a vehicle is not in use, it will appear as available to other potential users. Thus, the recorded data contains observations of available vehicles (with a timestamp and a location). A data transformation process has to be employed to deduct actual completed trips. When a vehicle is removed from the pool of available vehicles, it is assumed that a member uses the vehicle. Only the difference in time and location is returned between the last observed position of the vehicle and the first location following the car return in the available cars pool. Figure 1 presents the main methodology to convert web car position timestamps to OD data.

For this paper, the transactional dataset was primarily used as a control group to compare the GPS method travel distance calculation (section 4.1), but also mainly for the member information. This dataset doesn't contain spatial data, thus analysis performed in section 4 focused to compare GPS and OD data foremost.



Veh. ID	Timestamp	Location
A	13h05	45.6,-73.5
A	13h10	45.6,-73.5
A	13h15	45.6,-73.5
A	13h35	45.7,-73.4
B	...	...

Veh. ID	Timestamp start	Timestamp end	Location start	Location end
A	13h20	13h35	45.6,-73.5	45.7,-73.4
B	...	...	...	...
...	...	...	...	...

Fig. 1. Process used to transform repeated vehicle position capture to origin-destination trips

4. Passive data streams in free-floating carsharing use assessment

This section is separated in three sub-sections. First, an overview of the recorded and calculated travel distances is presented. Then, the trip classification method is detailed, and finally, the actual trip distribution over time is presented.

4.1. Travel distance

From the three main data sources, a fourth one is calculated. Based on the origin and the destination, the shortest path on the road network is calculated via a routing engine (OSRM, see http://project-osrm.org/) for every transaction found in the GPS dataset. The distributions of the calculated distances for each data source are presented in Table 1.

Table 1. Distributions of the calculated distances (meters) for each data source

Metric	Transactional	GPS	Web-based (OD)	Routing
Mean	13,050 m	12,860 m	3,144 m	3,839 m
25th percentile	3,000 m	2,973 m	1,007 m	1,018 m
Median	6,000 m	5,573 m	2,541 m	2,783 m
75th percentile	11,000 m	10,600 m	4,563 m	5,398 m

Table 1 shows the different distributions for the year 2015. Web-based (OD air-line distance) and routing distances are similar in distribution, while GPS and transactional share comparable numbers. Naturally, routing distances are higher than OD ones, because the shortest path as to take into consideration the road network in its distance calculation. GPS distance is lower compared to the transactional data (which can be considered the closest to the ground truth). This is in line with Lopez et al. (2016) study, indicating that GPS distances may be lower by 13% compared to transactional ones. GPS/transactional distributions appear substantially higher than the routing/web-based combination. This may indicate that members perform other activities along their trips and the trips' type distribution may be more complex than simply the shortest/quickest way to their destination.

4.2. Trip Classification Method

To explore more deeply the user behavior in a free-floating setting, four different trip types are categorized, four for the GPS method and three for the OD technique. In this context, trip types are being refer to the structure of the trip (one-way, round-trip, symmetric) or an estimated purpose (commuting). Inherently, both methods' trip distributions are directly affected by the various arguments used to segment the trips in different classes. For comparison and reproducibility purposes, Table 2 presents the classification features and their values used in this analysis of both methods.

From Table 2, one-way and two-way (round-trip) trips are mostly discriminated by the distance between the origin and the destination of the trip. Commuting represents all trips made in a weekday morning departing near the user home location where in the afternoon a second trip starts nearby the destination of the morning trip to return nearby the user home location. Commuting is exclusive to the GPS method because a member id is not available in the OD method dataset. In addition, no data on trip purpose is available and thus commute-related trips are only an estimation. Symmetric trips are similar to commuting trips but are not restricted on the trip start time nor the user home location. Return trips have to be within 0.5 to 6 hours from the end of the foregoing trip. It has to be noted that trip types requiring to have the same member id for both trips are made within a smaller sample size (only all trips having a correct match are retained).

4.3. Trip distribution

Figure 2 presents the trip distribution for both the GPS and the OD data collection methods from January 2015 to July 2017. By definition (as in Table 2) one-way and round-trips are quasi complementary. The majority of trips are one-way made (68.0%-78.1% with the GPS method). This is coherent by the free-floating nature of the service. For the

Montreal case (Wielinski, 2014), round-trips are estimated to be at ~20.5%, which is consistent with our results. While the Munich and Berlin case (Schmöller et al. 2014) showcase a value between 8-11%, the ability in Montreal to use a FFcs vehicle for longer rides and pay only the cheapest fare between both Communauto services fee structure should help to increase the proportion of round-trips for Communauto use case. Measuring symmetric trips in a FFcs service allows to evaluate the position of a transportation mode with no anchor points as free-floating carsharing inside the mobility cocktail of a city. A lower value increases the probability to use other means of transportation to complete a member trip-chain. For Auto-mobile case, they are in a smaller proportion (5.7%-7.4% for the GPS method). Lessening restrictions 7&8 in Table 2 to 750 metres increases the proportion to 7.5%-9.3%, while a 1000 meters threshold increases it at a range of 8.6%-10.5%. Commuting-related trips are compared to all trips made on weekdays to ensure a proper comparison. The proportion range from 4.1% to 8.8% and seem to be increasing in fall of 2015. This could be explained by the end of the summer vacation period and by a new expansion phase, which lead to a substantial increase of the free-floating car fleet (from 140 at the end of March 2015 to 270 at the end of September 2015). An increase in the car fleet density can result in a higher confidence in the availability of a free-floating vehicle, thus it could increase the willingness to use the free-floating service for commuting purpose. Ciari et al. (2014), through multi-agent simulations, estimated a ~12% work-related trips stating that FFcs increase the desirability (compared to SBcs) for work-related trips because the member is not billed while not using the car during work time.

Table 2. Parameters used to compose the different trip uses for both data collection methods

	GPS method				**OD method**		
Feature	Commuting	Round-trip	One-way	Symmetric	Round-trip	One-way	Symmetric
1-Trip distance	> 500 m	> 1000 m	> 500 m	> 500 m (1st & 2nd trip)			
2-Trip duration					>= 30 min	>= 10 min	>= 10 min
3-Distance between origin and destination (same trip)		< 750m	>= 750m		< 750m	>= 750m	>= 500m
4-Weekday	Mon to Fri						
5-Distance between origin and home	First trip <= 500m						
6-Distance between destination and member home location	Second trip <= 500m						
7-Distance between destination (1st) and origin (2nd trip)	<= 1000m			<= 500m			<= 500m
8-Distance between destination (2nd) and origin (1st trip)				<= 500m			<= 500m
9-Duration between destination (1st) and origin (2nd trip)				Between 0.5 hours & 6 hours			Between 0.5 hours & 6 hours
10-Same member id for both trips	TRUE			TRUE			
11-Start time	Between 5:00AM & 10:00AM (1st) / Between 3:00PM & 8:00PM (2nd trip)						

Until now, the GPS method has been mainly used to characterize the member usage of the service because of its higher precision over the OD method. When comparing both methods, we observe some similar behavior in the period where both data periods are available. For example, one-way trips and round-trips follow a similar trend for both methods from July to December 2015 indicating that the OD method could have the potential to estimate one-way and round-trip trips accurately. Because we can't measure the actual driven distance for a given trip with the OD method, instead of dividing the trips according to the distance, as the GPS method allows, we do it with the trip duration. This inevitably creates some biases between both methods. Also, at the time of the study, the process to generate trips from the OD method would not consider trips longer than 3 hours. Looking at the transactional dataset, the proportion of trips with a length higher than 3 hours is at 8.5%, meaning the discrepancy between the OD and the GPS method could be in part explain by this restriction.

Regarding the OD method, symmetric trips seem to be higher than its counterpart. No restriction about the member, like in the GPS method (feature 10 in Table 2), increases the proportion of these trips. When examining the distribution over time, we note an increase from June 2016 onwards. While no GPS data is available to investigate the validity of this increase, we can specify that over the period between June 2016 and February 2017, the car fleet increased from 300 to 530. This important increase in vehicle density (no major service area zones were added in the meantime) could increase the density of trips in a zone and thus recorded symmetric trips could be the result of symmetric trips made by separate members and potentially limit the use of this method to measure symmetric trips in a higher density car fleet usage.

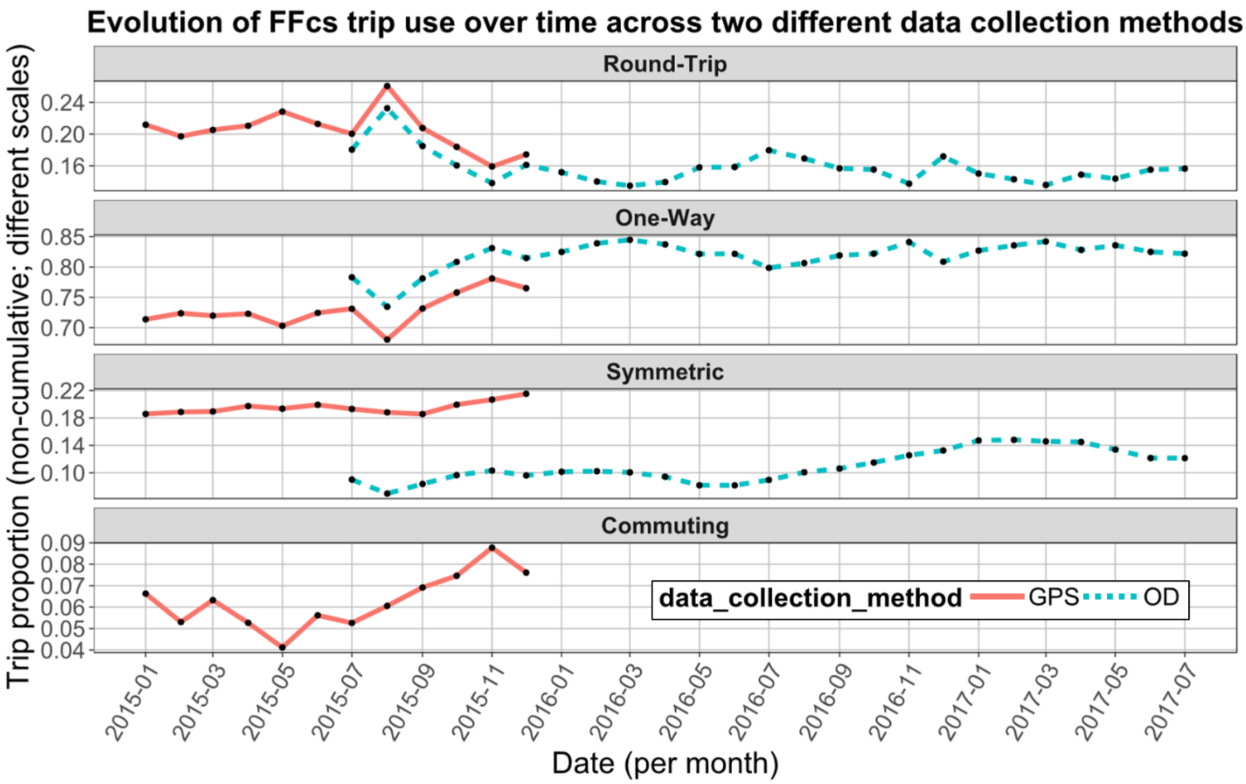


Fig. 2. Longitudinal evolution of Communauto FFcs service uses distributions (non-cumulative) amongst two data collection methods

5. Discussion

Behavioral change can be looked at an aggregated or singular level. In our case, the analyses were made system-wide. About RQ2, over time the CSO upgraded the car fleet and the service area size. This can directly influence the user behavior. Trip type evolution shows multiple trends. First, one-way trips seem to increase at the end of 2015 summer to stabilize itself afterwards. Second, work-related trips follow a positive trend through the end of 2015. Third, symmetric trips proportion seems to be in a more important volume following mid-2016. While these events could be related to the service changes, further investigation must be undertaken (1) to conclude if a significant change really occurred and (2) what caused this change. Member-level change (with data sources having a member identifier) can also be looked at in future work.

Regarding the different data sources, OD data may be a good substitute for the more extensive datasets as GPS data. Results showed a good interrelation between GPS and OD trip type trends. However, both methods have their own limitations. Table 3 presents a brief comparison between them on six different aspects (RQ3).

Table 3. Comparison chart for GPS and OD data

Features	GPS method	OD method
Processing required	Very high	High
Data availability	Third-party dependent	In-house but subject to the service provider
Information level	Transaction, trip, activity, distance, time & location	Transaction approximated time & location
Data fusion capabilities	Easier data fusion with other datasets having unique identifiers	Data fusion limited with assumptions
Change assessment level	Service-wide & per user	Service-wide
Context-specific limitations		May be limited in high density zones (symmetric trips) + no work trip detection possible

Because of its size, GPS data is more process intensive than OD data. Also, GPS data requires to create and refine trip and stop detection algorithms. OD data allows a better control of the data availability (near real-time), but it is subject to the web-server uptime. Because of its richness, GPS data is more precise, allowing to segment a transaction into trips and stops, giving us the full portrait of the situation. It is also possible to join the GPS dataset to the transactional dataset via unique identifiers to access information on the member (when available). This allows richer analyses (member-level metrics), but also allows a fairer assessment of commuting and symmetric trips. Without a unique member id, the OD data may be limited for symmetric trips in high-density free-floating zones.

6. Conclusion

This paper looked to compare multiple data streams on the case study of free-floating carsharing. More specifically, three main research questions were addressed. First, the trip type distribution for the GPS and the OD method has been ruled. While one-way trips correspond to the majority of the trips, a fair proportion (~20%) are two-way trips. Symmetric and commuting trips are also to be found, but in lesser proportion. Second, the change of these metrics was considered on a service-wide level. Some fluctuations have occurred, but further analyses should be done to effectively associate a cause to those changes. Finally, pros and cons of both methods were discussed. OD data is found to be a good proxy in general, with some well-specified limitations. Still, passive data streams are not telling the whole story, even when a full range of data collection methods are made available. While Communauto free-floating service is a FFcs service per se, the dual-mode setting of Communauto may distort trip types distribution (the pricing structure doesn't entice only short trips for members of both services).

On limitations, OD trip weighting is not a mentioned practice in the literature. Because OD data is subject to server downtimes, trip weighting could be an important factor. In our case, this process wasn't performed (downtimes are assumed to be at random), but it certainly as to be considered. Further directions should be closely considerate. Following the literature presented in section 2, a more thorough investigation of all employed hypotheses should be evaluated. This could give an opportunity to apply sensitivity analyses on past hypotheses. Routing based duration and energy-consumption methods were not explored in this paper and merit to be assessed. Also, trip activities should be integrated in one-way trips to split this trip type in other more precise categories. Finally, a study of the member as the unit of analysis should also be considered.

7. Acknowledgments

This work was possible thanks to the support and collaboration from Communauto. The research project is also funded by the Natural Science and Engineering Research Council of Canada (NSERC RDCPJ # 474642-14).

8. References

Boldrini, C., Bruno, R., & Laarabi, H. (2017). Car sharing through the data analysis lens. arXiv pre-print. Retrieved November 7, 2017 from arXiv database (1708.00497).

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, (2416), 37-47. https://doi.org/10.3141/2416-05

Ciociola, A., Cocca, M., Giordano, D., Mellia, M., Morichetta, A., Putina, A., & Salutari, F. (2017). UMAP: Urban Mobility Analysis Platform to Harvest Car Sharing Data. Telecommunication Networks Group. Retrieved November 7, 2017 from https://www.telematica.polito.it/sites/default/files/paper.pdf.

Heilig, M., Mallig, N., Schröder, O., Kagerbauer, M., & Vortisch, P. (2017). Implementation of free-floating and station-based carsharing in an agent-based travel demand model. Travel Behaviour and Society. Article in press. https://doi.org/10.1016/j.tbs.2017.02.002

Hui, Y., Ding, M., Zheng, K., & Lou, D. (2017). Observing Trip Chain Characteristics of Round-Trip Carsharing Users in China: A Case Study Based on GPS Data in Hangzhou City. Sustainability, 9(6), 949. https://doi.org/ 10.3390/su9060949

Kortum, K., Schönduwe, R., Stolte, B., & Bock, B. (2016). Free-Floating Carsharing: City-Specific Growth Rates and Success Factors. Transportation Research Procedia, 19, 328-340. https://doi.org/10.1016/j.trpro.2016.12.092

Leclerc, B., Trépanier, M., & Morency, C. (2013). Unraveling the travel behavior of carsharing members from global positioning system traces. Transportation Research Record: Journal of the Transportation Research Board, (2359), 59-67. https://doi.org/10.3141/2359-08

Lopez, A., Semanjski, I., Gillis, D., Ochoa, D., & Gautama, S. (2016). Travelled Distance Estimation for GPS-Based Round Trips Car-Sharing Use Case. Transactions on Maritime Science, 5(02), 121-129. https://doi.org/10.7225/toms.v05.n02.003

Millard-Ball, A., Murray, G., Ter Schure, J., Fox, C., and J. Burkhardt. Car-Sharing: Where and How It Succeeds. Transit Cooperative Research Program Report 108. Transportation Research Board of the National Academies, Washington, D.C., 2005.

Shaheen, S. A., & Cohen, A. P. (2013). Carsharing and personal vehicle services: worldwide market developments and emerging trends. International Journal of Sustainable Transportation, 7(1), 5-34. https://doi.org/10.1080/15568318.2012.660103

Shaheen, S., & Cohen, A. (2016). Innovative mobility carsharing outlook - Winter 2016. Transportation Sustainability Research Center - University of California, Berkeley. Retrieved October 10, 2016 from http://innovativemobility.org/wp-content/uploads/2016/02/Innovative-Mobility-Industry-Outlook_World-2016-Final.pdf.

Schmöller, S., Weikl, S., Müller, J., & Bogenberger, K. (2014). Empirical data analysis of free-floating carsharing systems. 93rd Annual Meeting of the Transportation Research Board. Paper no. 14-4410.

Schulte, F., & Voß, S. (2015). Decision support for environmental-friendly vehicle relocations in free-floating car sharing systems: The case of car2go. Procedia CIRP, 30, 275-280. https://doi.org/10.1016/j.procir.2015.02.090

Tyndall, J. (2017). Where no cars go: Free-floating carshare and inequality of access. International Journal of Sustainable Transportation, 11(6), 433-442. https://doi.org/10.1080/15568318.2016.1266425

Wang, H., Fu, L., Zhou, Y., & Li, H. (2008). Modelling of the fuel consumption for passenger cars regarding driving characteristics. Transportation Research Part D: Transport and Environment, 13(7), 479-482. https://doi.org/10.1016/j.trd.2008.09.002

Wang, X., MacKenzie, D., & Cui, Z. (2017). Complement or Competitior? Comparing car2go and Transit Travel Times, Prices, and Usage Patterns in Seattle. 96th Annual Meeting of the Transportation Research Board. Paper no. 17-06234.

Wielinski, G. (2014). Analyse de l'implantation de l'autopartage en libre-service intégral à Montréal (Master's thesis, Polytechnique Montreal). Retrieved November 8, 2017 from https://publications.polymtl.ca/1664/.

Wielinski, G., Trépanier, M., & Morency, C. (2015). What About Free-Floating Carsharing? A Look at the Montreal, Canada, Case. Transportation Research Record: Journal of the Transportation Research Board, (2536), 28-36. https://doi.org/10.3141/2536-04

Wielinski, G., Trépanier, M., & Morency, C. (2017). Electric and hybrid car use in a free-floating carsharing system. International Journal of Sustainable Transportation, 11(3), 161-169. https://doi.org/10.1080/15568318.2016.1220653

Willing, C., Klemmer, K., Brandt, T., & Neumann, D. (2017). Moving in time and space–Location intelligence for carsharing decision support. Decision Support Systems, 99, 75-85. https://doi.org/10.1016/j.dss.2017.05.005

Wolf, J., Guensler, R., & Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. Transportation Research Record: Journal of the Transportation Research Board, (1768), 125-134. https://doi.org/10.3141/1768-15