Travelled Distance Estimation for GPS-Based Round Trips Car-Sharing Use Case

Angel J. Lopez[a,b], Ivana Semanjski[a], Dominique Gillis[a], Daniel Ochoa[b], Sidharta Gautama[a]

Traditional travel survey methods have been widely used for collecting information about urban mobility although Global Position System (GPS) has become an automatic option for collecting more precise data of the households since mid-1990s. Many studies on mobility patterns have focused on the GPS advantages leaving aside its issues such as the quality of the data collected. However, when it comes to extract the frequency of the trips and travelled distance, this technology faces some gaps due to the related issues such as signal reception and time-to-first-fix location that turns out in missing observations and respectively unrecognised or over-segmented trips. In this study, we focus on two aspects of GPS data for a car-mode, (i) measurement of the gaps in the travelled distance and (ii) estimation of the travelled distance and the factors that influence the GPS gaps. To asses that, GPS tracks are compared to a ground truth source. Additionally, the trips are analysed based on the land use (e.g. urban and rural areas) and length (e.g. short, medium and long trips). Results from 170 participants and more than a year of GPS-tracking show that around 9 % of the travelled distance is not captured by GPS and it affects more short trips than long ones. Moreover, we validate the importance of the time spent on the user activity and the land use as factors that influence the gaps in GPS.

a. Ghent University, Department of Telecommunications and Information Processing, Ghent, Belgium

e-mail: angel.lopez@ugent.be

b. Facultad de Ingeniería en Electricidad y Computación, Politécnica del Litoral, Guayaquil, Ecuador

e-mail: alopez@espol.edu.ec

1. INTRODUCTION

Traditional travel survey methods have been widely used in transportation research as a tool for collecting information at an individual or household level (e.g. description of demographics, travel patterns, trip purpose and mode choice) (Handy, 1996). Yet, respondents have the tendency to omit short stops such as the post office and ATM and when it comes to numerical answers, travel time is rounded to simple values like 10, 15, 30-minute interval. Similar happens to the travelled distance (Wagner, 1997). Nonetheless, since the introduction of the Global Positioning System (GPS) and its first adoption in transportation studies in mid-1990s, where Wagner (Wagner, 1997) reports one of the first studies that uses GPS for collecting information of 100 households through logger devices installed in their vehicles, it has been extensively used in combination with others datasets. Studies (Yalamanchili et al., 1999; Draijer et al., 2000; Wolf et al., 2000) have already demonstrated the possibility to use GPS data in Transportation research by capturing the characteristics of different types of trips. Later studies (Bolbol et al., 2012; Schuessler et al., 2015) evaluated the use of processed GPS data for both trip tracking and transportation mode detection without the support of questionnaires. Their results showed that trip identification deviates slightly from the census data, whereas

for mode detection it was not possible to distinguish between transportation modes with similar speed such as bus and car trips. Trip reporting and therefore travelled distance are challenging issues that can be achieved by detecting the transition between transportation modes although, if many transitions are detected for a single-mode trip, it turns out in over-reporting of trips and under-reporting of distance. In other studies (Zhang et al., 2013; Zheng et al., 2010), where the GPS trajectories are split using features such as speed and distance, it applies a good common-sense knowledge of the world describing that the start and end points of a walk segment may be changes of transportation mode. Liao et al. (2007) use the probabilistic approach to estimate those changes. Zheng et al. (Zheng et al., 2008) show that extracting trip frequencies is a challenging task despite his method's classifying transportation-mode with an accuracy of 76 %, the precision for trip counting is below 30 % due to the over-segmentation, which actually shows that many studies report the classification error as a proportion of miss-classification regardless the over-segmentation. A way to overcome that is by merging the consecutive trips with similar transportation mode using heuristics on the distance and time between trips or by smoothing the classification outcomes (Lopez at al., 2015).

However, signal reception and time-to-first-fix (TTFF) are well known issues of GPS, both of them affecting the reported distance (Flamm et al., 2007). Signal reception is mainly influent for external factors that can block or reflect the signal, it can lead with either *signal loss* due to poor satellite reception (e.g. underground travel, bridges and tunnels) or multi-path errors referred to as *urban canyoning errors* because they appear in urban canyons where the signal is reflected by buildings (Schuessler and Axhausen, 2008). TTFF also referred to as *cold / warm start* is the delay in getting the first observation when the GPS device has been off for a period of time, which turns out in missing observations at the beginning of the trip (Stopher et al., 2005).

In this study, we use tracking data from a fleet of shared cars to (i) assess the gaps present in the GPS-based distance and their possible effect in mobility studies, and (ii) to estimate the travelled distance and the relevant factors that influence the GPS gaps. To accomplish the previous points, we need a ground truth source that reports the driven kilometres when the car is used, therefore the odometer-sensor is chosen and its data is accessible through the Controller Area Network bus (CAN-bus) (I. S. O. Standard, 1993). CAN-bus is known as a protocol for high performance and high reliable serial communication links between electronic control units (e.g. sensors), and it is mainly used in the field of automotive and industrial control applications (Turski, 1994). CAN-bus has been used together with GPS data to estimate mobility parameters in off-road vehicles such as resistance force and wheel slip under different terrain conditions (Suvinen and Saarilahti, 2006), where a GPS logger was used to gather the ground speed and trajectory; and the gross power and rotational speed were extracted from the CAN-bus. Furthermore, given that GPS trajectories are spread across different type of land use, we use a Geographic Information System (GIS) to make a distinction between trips in urban and rural areas.

The remainder of this paper is organised as follows. Section 2 describes the datasets and methods for assessing GPS gaps on the car mode. Section 3 presents the case study and the outcomes of this research and Section 4 summarise the findings by drawing conclusions on it.

2. METHODOLOGY

The data from the present research is drawn from two sources (a) a car-sharing company named *Cambio* (Cambio Belgium a Car-sharing Company, 2016), which started its operation in Belgium in 2002 and at present (June 2016) it is available in 35 cities with 369 stations, 862 cars and more than 24,000 users. *Cambio* provided us with a dataset of the reservation details (e.g. distance, duration, start and ending times) based on CAN-bus data. (b) GPS data that are collected through loggers installed on selected cars from the car-sharing company.

2.1. Dataset Description

As one of the interests in this study is to identify the differences in GPS-based distance and the actually driven kilometres, we use the car odometer-sensor data as the ground truth. The access to such data is possible through the CAN-bus. A CAN-bus is known as a protocol for high performance and highly reliable serial communication links between electronic control units in the field of automotive and industrial control applications (Turski, 1994), e.g. it is typically used to control and automatically calibrate the engine performance in a vehicle. It was developed as a multi-master message broadcast system (Farsi et al., 1999) where each element on the network can send a message (e.g. temperature, state of charge) independently to the entire network, the bus priority being defined by the message identifier (Bosch, 1991). The dataset contains information about the car reservation including the total distance, reservation period (e.g. duration), identifiers (cars/client/reservation), start and ending times.

A limitation of this dataset concerns its granularity at the reservation level. Therefore, the reservation period represents the total duration of the reservation rather than the travel time. As an illustration, a reservation starts when the car is picked up from the car charging station and it ends when the car returns back to the charging station. Consequently, the reservation period contains not only the travel time, but also the time spent on the participant activities, e.g. doing shopping, visiting someone, etc. Nonetheless, the travelled distance and the starting time (the first

trip segment) are not affected by the aforementioned limitation.

GPS has been used in several studies (Section 1). It allows tracking targets through their geographical location, where a GPS logger is a device for collecting locations and other measurements such as speed, altitude, heading, accuracy and timestamp. It can have either a built-in or external antenna and the data can be stored in the internal memory for being downloaded later on, or sent to a centralised repository through the network. For collecting the data, a GPS logger GenLoc 53e (Erco and Gener *genlog53e*, 2016) is installed in the cars. The device can send the tracking data through the cellular network to a centralised system in which it is processed. The frequency of 1 Hz is used for collecting the GPS data and the logger only collects observations when the car is turned on (i.e. the GPS logger automatically starts when the car is turned on and stops when the car is turned off); therefore, it can occur that a single reservation includes more than one trip, which is expressed as follow:

$$R_i = \{ S_1^{(i)}, ..., S_m^{(i)} \} \tag{1}$$

where R_i represents the ith reservation, $S_j^{(i)}$ is the j th trip segment within a reservation R_i such as $S_j^{(i)} \in \{ S_1^{(i)}, ..., S_m^{(i)} \}$ and m is the number of trip segments per reservation.

2.2. Data Quality

To assess the quality of the GPS-based distance, we calculate the missing distance and the TTFF using the following formulas:

$$Missing\ distance_{prop} = \frac{\sum_{S_j^{(i)} \in R_i} d(S_j^{(i)})}{\sum_{i=1}^{N} d(R_i)} \tag{2}$$

$$Missing\ distance_{avg} = \frac{1}{N} \sum_{i=1}^{N} \left(d(R_i) - \sum_{S_j^{(i)} \in R_i} d(S_j^{(i)}) \right) \tag{3}$$

$$TTFF_{avg} = \frac{1}{N} \sum_{i=1}^{N} \left(t(S_1^{(i)}) - t(R_i) \right) \tag{4}$$

where $d()$ and $t()$ are functions to extract the distance and start time respectively, $S_1^{(i)}$ represents the first trip segment of a reservation R_j, and N is the number of observations.

Equation 2 is the proportion of the missing distance and it calculates the overall missing distance for all reservations. Nevertheless, it can also be applied to a set of trips based on some conditions such as land use and trip groups. However, this formulation faces two possible outcomes: when the outcome is positive, it represents an under-reported distance, whereas it is negative when the distance is over-reporting. This last scenario can be explained by the GPS accuracy because under certain conditions (e.g. tunnel, bridge, parking lots) it turns out in jumps around the same location. Those points can add an extra distance to the trip.

Equation 3 is the average missing distance and it is expressed in kilometres. Equation 4 is the average of time-to-first-fix; it calculates an average of time difference between the reservation start time and the timestamp of the first GPS location.

2.3. Land Use

The key research focus of this study is to assess the GPS gaps, where the land use provides an extra perspective to analyse those gaps in rural and urban areas, given that the GPS signal reception could be affected due to the high density of large structures (e.g. buildings, bridges). A GIS tool allows to identify the land use of a trip (Stopher et al., 2008), i.e. whether a trip was performed on a rural or urban area, in such a way that trips are matched within an administrative area (boundary area classified as rural or urban) using the origin and destination points. Consequently, a precise land use identification relies on the completeness of the geographical information for the administrative areas.

The administrative areas (in which the trips were performed) are extracted from ©OpenStreetMap contributors (OSM) (OpenStreetMap contributors, 2016). OSM is an open access platform for geospatial vector data and it is often considered complete and appropriate for planning studies in comparison to other commercial counterparts (Hochmair et al., 2013).

2.4. Regression Method

To explain gaps in the GPS data, the factors that influence data collection are identified through a linear regression. Linear regression methods are techniques for modelling the relationships between a scalar dependent variable and its explanatory or independent variables known as covariates. That relationship is modelled by an error term ε, a random variable that adds noise to the linear relationship between the dependent and independent variables (Neter et al., 1996). Therefore, the model is expressed as follows:

$$y = X\beta + \varepsilon \tag{5}$$

where y is the dependent variable or response variable, X represents the explanatory variables, β the regression coefficients or effects and ε is the error term.
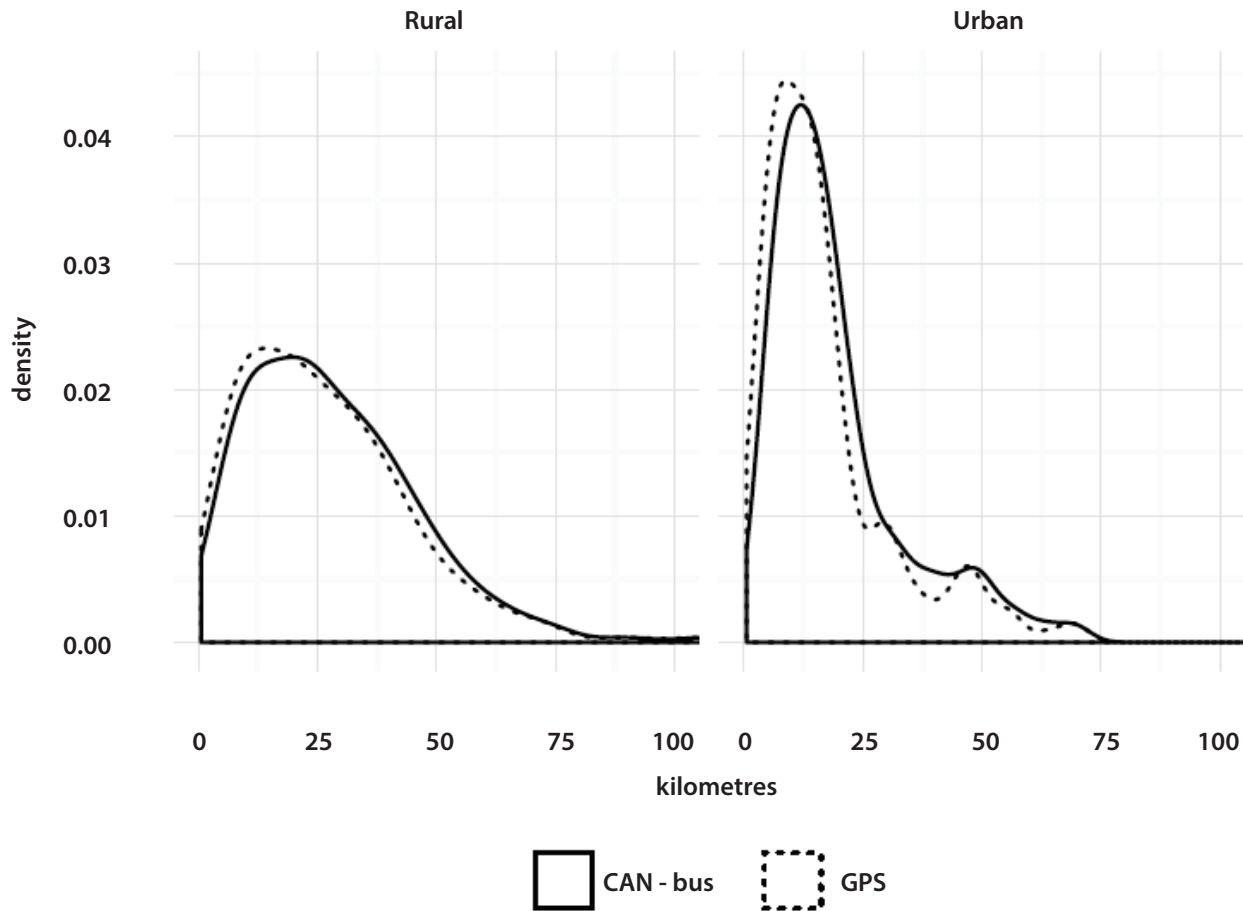

Figure 1.
Travelled distance density based on land use for GPS logger and CAN-bus data.

3. CASE STUDY

The data on this study are part of the Olympus project, a Flemish initiative to promote the introduction of electric vehicles in Belgium. It was a common project between suppliers, integrators and users of shared mobility, aiming at developing tools and systems to enhance multimodal travel behaviour. Therefore, the multimodal travel behaviour was monitored both from the vehicle perspective and from the personal perspective. From the vehicle perspective, GPS loggers were installed on shared electric cars, showing how these were used in terms of frequency, trip length, origins and destinations.

Data collection was performed between 2012 and 2013 in the Belgian cities Ghent, Antwerp and Leuven, involving 170 participants.

3.1. Descriptive Analysis

A summary of the travelled distance reported by the CAN-bus data and GPS logger is depicted in Table 1, where we can notice clear differences on the reported average distances.

Table 1.
A summary of the travelled distance (km).

Data	Min	median	mean	max
CAN-bus	2.0	19.0	23.9	106.0
GPS Logger	0.6	17.2	21.8	104.6

In order to obtain an analysis at different levels, we group the trips based on the travelled distance into three categories such as short, medium and long trips. The conditions for those groups are depicted in Table 2.

Table 2.
Group of trips based on the travelled distance.

Group	Description
Short trip	less than 10 km.
Middle trip	between 10 and 25 km.
Long trip	more than 25 km-

Those groups and the land use allow to identify which type of trips are mainly affected when it comes to GPS tracking.

Figure 1 shows the distance density on rural and urban areas based on CAN-bus and GPS data, where the GPS density curve is shifted to the left side with respect to CAN-bus in both rural and urban areas. This is another indication of missing distance based on different land use.

3.2. Distance Measurement Gaps

For the travelled distance by car, we consider the missing distance as the difference between the odometer distance (e.g. distance obtained from the CAN-bus data) and the logging distance (e.g. distance reported by the GPS logger) as it Equation (2) shows. It shows that on average 9 % of the travelled distance is not captured by the GPS logger (Figure 2). This means that for an average trip with distance 23.9 km the GPS will report on average 2.1 km less (Table 1).

From the groups, we calculate the travelled distance for both GPS and CAN-bus data. Figure 3 shows the missing data for short, medium and long trips within groups of different land used.
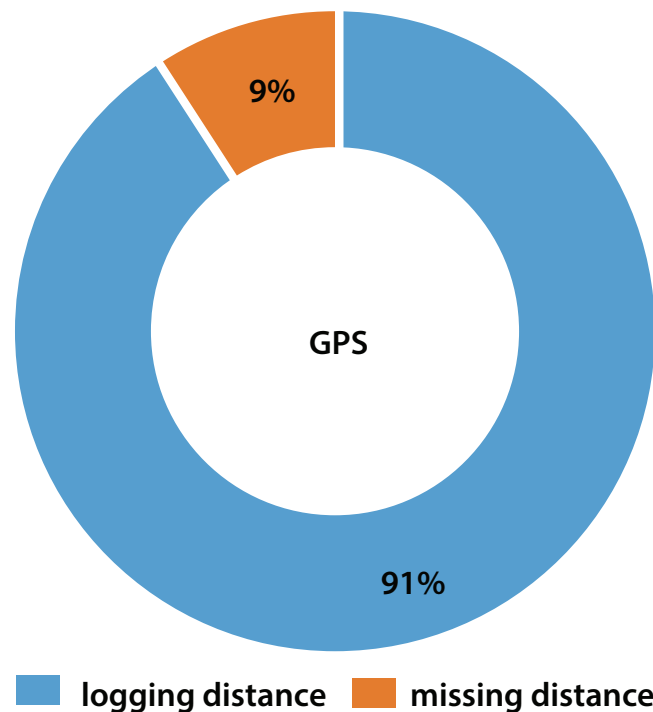

Figure 2.
Global missing distance for a car mode using GPS.

Table 3 shows that short and medium trips in urban areas are the ones that report high values of missing distance, logging around 81 % and 83 % of the travelled distance respectively. It also shows that long trips in both urban and rural areas perform better than the rest, logging 92% and 95% respectively. On average, trips on rural areas report less missing data than urban areas.

Table 3.
A summary of the GPS gaps based on the land use.

			TTLL (min)		Missing distance (km)	
	Group	**Trips**	**Median**	**Average**	**Average**	**Percentage**
Rural	Short	82	4.6	7.8	0.9	11.9%
	Medium	190	6.1	7.0	1.9	10.1%
	Long	292	5.6	8.3	2.0	4.6%
Urban	Short	161	5.5	16.5	1.4	19.3%
	Medium	326	5.9	10.8	2.9	16.9%
	Long	189	6.2	7.8	3.6	8.5%

To our understanding, part of the missing data could be related to the cold / warm start issue that is present on the GPS technology, where it is required a period of time before fixing the first location. Therefore, we assess TTFF with Equation (4) that uses the starting time of the trip.
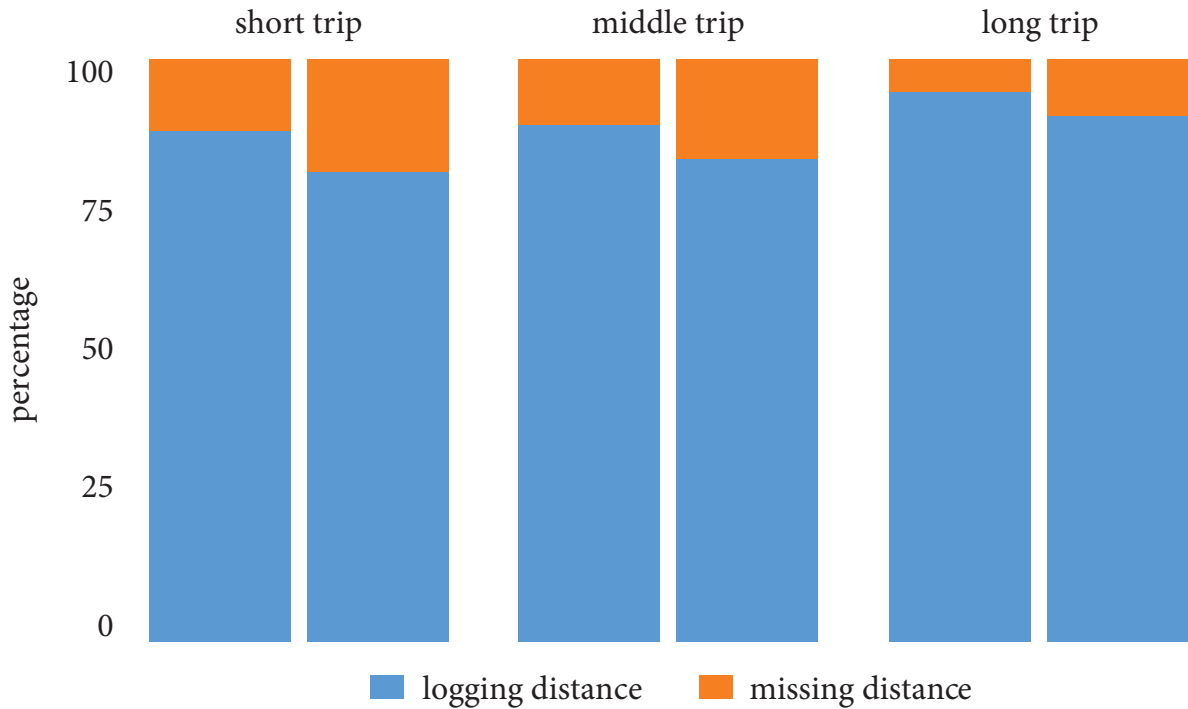

Figure 3.
Percentage of missing data for different group of trips, logging distance is the distance reported by the GPS logger and missing distance is the difference with respect to ground truth (CAN-bus).

3.3. Explanatory Factors

To explain the travelled distance a regression model is fitted to the GPS data, where the dependent variable is provided by the CAN-bus data as a travelled distance and from the GPS data we extract the covariates such as distance, duration, average speed and number of trips per reservation.

Others covariates such as land use are obtained using GIS, where the origin and destination points are used for classifying the trips within a rural or urban area. Finally, the time spent on the user activity (purpose of mobility) is calculated from the reservation time and the trip duration. A full description of the covariates is presented in Table 4.

A summary of the fitted models is depicted in Table 5, where the covariate coefficient is estimated using a regression model and its significant level is based on the p-value. Based on *model*

Table 4.
Model covariates.

Covariate	Description
Distance	Logging distance from the GPS data (km)
Duration	Travelled time for the logging distance (min)
Trip segments	Number of trip segments per reservation
Average speed	Average speed for the trip (km/h)
Time spent	Time spent on the user activity (e.g. shopping, other)
Time spent per trip	Average time spent per trip
Land use	Land use for the trip (e.g. urban, rural)

Table 5.
Models result.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: Travelled distance (km)						
Distance	0.976*** (0.013)	1.032*** (0.020)	1.044*** (0.021)	0.888*** (0.049)	0.905*** (0.017)	0.920*** (0.017)	0.930*** (0.016)
Duratiom		-0.047*** (0.013)	-0.061*** (0.016)	0.019 (0.024)			
Trip segments			0.250 (0.155)				
Average speed				7.180*** (2.232)	5.854*** (1.177)	4.701*** (1.247)	5.196*** (1.225)
Time spent					0.008*** (0.002)	0.003 (0.003)	
Time spent per trip						0.015*** (0.006)	0.019*** (0.004)
Urban area							1.062*** (0.388)
Constant	2.837*** (0.324)	3.698*** (0.402)	3.247*** (0.488)	0.535 (1.060)	0.457 (0.477)	0.709 (0.482)	-0.076 (0.565)
AIC	1967	1957	1956	1949	1928	1923	1916
BIC	1979	1972	1976	1968	1947	1946	1940
R²	0.944	0.946	0.946	0.948	0.950	0.951	0.952
Adjusted R²	0.944	0.946	0.946	0.947	0.950	0.951	0.952
Note:					*$p<0.1$:	; **$p<0.05$;	***$p<0.01$

Covariates: coefficient and coefficient error in brackets.
AIC: Akaike information criterion, BIC: Bayesian information criterion.

3 the number of *trip segments* does not seem to be significant. On the other hand, from *model 4* we can notice that the average speed is a good predictor and makes the *duration* less significant.

Time spent shows significant results on *model 5*. It means that the waiting time within trips (e.g. time when the car is parked) is an important factor for modelling the travelled distance because it adds more periods of cold / warm start to the GPS logger (i.e. a longer waiting time increases the chances of missing data). It is even more significant when it is combined with the *trip segments* as an average of time spent on a particular activity per *trip segment* (e.g. time spent over trip segments). Land use is another influential variable; its coefficient indicates that for each trip in the urban area it will add around a kilometre to the total distance. Based on the R^2 *model 7* explains better the travelled distance as a function of the covariates: distance, average speed, time spent per trip and land use.

4. CONCLUSION

In this paper, the gaps in the GPS-based distance using CAN-bus data as a ground truth are measured; likewise, the factors that influence those gaps were identified through regression models. It was found that 9% of the travelled distance is not captured for the GPS logger. This is important considering that many mobility studies are being conducted using GPS data, and their outcomes might underestimate the actual travelled distance. However, it will depend on the type of land use where the trips are carried out. In our findings, the rural area reports 6% of missing distance whereas the urban area reports about 13%. It is a clear indication that urban areas are more susceptible to issues related to the signal reception affected in around one kilometre to the reported distance (based on the *model 7*).

Moreover, the model also includes the cold / warm start as a function of the time spent on the user activity per trip, which provides an average waiting time between trips. This last factor should be considered when it comes to the trip reporting because in real situations round trips have more than a single trip and long periods of waiting time. Consequently, it adds a delay for getting the first valid location.

Part of the missing distance could be corrected by interpolating the missing GPS points within a trip. Using a road network is feasible to route (provide alternative trajectories between two points by a given transportation mode) and align the points to a valid location. However, it becomes complicated when the missing part is at the beginning of the trip because there is not any reference point to interpolate that part. This is the case of the cold / warm start.

Our findings contrast with another study (Vlassenroot et al., 2014) that reports 4% of the missing data (comparison among theoretic GPS points rather than travelled distance) for a driving mode using smartphones as tool for GPS data collection.

However, a smartphone does not have long periods of being off; hence, we could assume that TTFF does not have much influence on the data collection.

Future directions are focused on the GPS data quality for other transportation modes (e.g. walking, biking and public transportation) and also the comparison among the methods for collecting data such as passive and active logging.

ACKNOWLEDGMENT

REFERENCES

Bolbol, A., Cheng, T., Tsapakis, I. and Haworth, J., (2012), Inferring Hybrid Transportation Modes from Sparse GPS Data Using a Moving Window SVM Classification, Computers, Environment and Urban Systems, 36(6), pp. 526 – 537., http://dx.doi.org/10.1016/j.compenvurbsys.2012.06.001

Bosch, R., (1991), CAN Specification Version 2.0, Stuttgart: Bosch.

Cambio Belgium a Car-sharing Company, available at: http://www.cambio.be, [accessed 30 June 2016.].

Draijer, G., Kalfs, N. and Perdok, J., (2000), Global Positioning System as Data Collection Method for Travel Research, Transportation Research Record: Journal of the Transportation Research Board, 1719, pp. 147-153., http://dx.doi.org/10.3141/1719-19

Erco and Gener genlog53e, availabe at: http://www.ercogener.com, [accessed: 30 June 2016.].

Farsi, M., Ratcliff, K., Barbosa, M., Ratcliff, K. and Farsi, M., (1999), An Overview of Controller Area Network, Computing & Control Engineering, 10(3), pp. 113 – 120., http://dx.doi.org/10.1049/cce:19990304

Flamm, M., Jemelin, C. and Kaufmann, V., (2007), Combining Person Based GPS Tracking and Prompted Recall Interviews for a Comprehensive Investigation of Travel Behaviour Adaptation Processes During Life Course Transitions, Porc. 11th World Conference on Transportation Research, Berkeley, June 24 – 28.

Handy, S., (1996), Methodologies for Exploring the Link Between Urban Form and Travel Behavior, Transportation Research Part D: Transport and Environment,1(2), pp. 151-165., http://dx.doi.org/10.1016/S1361-9209(96)00010-7

Hochmair, H. H., Zielstra, D., Neis, P., Hartwig, P. N., Hochmair, H. and Zielstra, D., (2013), Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe, Proc. Transportation Research Board 92nd Annual Meeting, Washington DC, USA, January 13 – 17, 19(1), pp. 1 – 21.

I. S. O. Standard, "ISO 11898, 1993,", (1993), Road vehicles—interchange of digital information—Controller Area Network (CAN) for high-speed communication.

Liao, L., Patterson, D. J., Fox, D. and Kautz, H., (2007), Learning and Inferring Transportation Routines, Artificial Intelligence, 171(5-6), pp. 311-331., http://dx.doi.org/10.1016/j.artint.2007.01.006

Lopez, A. J., Ochoa, D. and Gautama, S., (2015), Detecting Changes of Transportation-Mode by Using Classification Data, Proc. 18th International Conference on Information Fusion (Fusion 2015). Washington, DC, USA, July 6 – 9, pp. 2078-2083., available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&} arnumber=7266810{ & }isnumber=7266535

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W., (1996), Applied linear statistical models, 4th edition, Chicago: Irwin.

OpenStreetMap contributors, availabe at: http://www.openstreetmap.org/copyright, [accessed 30 June 2016.].

Schuessler, N. and Axhausen, K., (2008), Identifying Trips and Activities and Their Characteristics from GPS Raw Data Without Further Information, Zürich: ETH., available at: http://e-collection.library.ethz.ch/eserv/eth:30471/eth-30471-01.pdf

Schuessler, N. and Axhausen, K., (2009), Processing Raw Data from Global Positioning Systems Without Additional Information, Transportation Research Record: Journal of the Transportation Research Board, 2105, pp. 28 – 36., http://dx.doi.org/10.3141/2105-04

Stopher, P., Clifford, E., Zhang, J. and FitzGerald, C., (2008), Deducing Mode and Purpose from GPS Data, Sydney, Australia: Institute of Transport and Logistics Studies.

Stopher, P., Jiang, Q. and FitzGerald, C., (2005), Processing GPS Data from Travel Surveys, 2nd International Colloqium on the Behavioural Foundations of Integrated Land-Use and Transportation Models: Frameworks, Models and Applications, Toronto, Canada, pp. 1-21.

Suvinen, A. and Saarilahti, M., (2006), Measuring the Mobility Parameters of Forwarders Using GPS and CAN Bus Techniques, Journal of Terramechanics, 43(2), pp. 237-252., http://dx.doi.org/10.1016/j.jterra.2005.12.005

Turski, K., (1994), A Global Time System for CAN Networks, Proc. 1st International CAN Conference, Mainz, Germany, availabe at: http://www.can-cia.org/fileadmin/resources/documents/proceedings/1994_turski.pdf

Vlassenroot, S., Gillis, D., Bellens, R. and Gautama, S., (2014), The Use of Smartphone Applications in the Collection of Travel Behaviour Data, International Journal of Intelligent Transportation Systems Research, 13(1), pp. 17-27., http://dx.doi.org/10.1007/s13177-013-0076-6

Wagner, D. P., (1997), Lexington Area Travel Data Collection Test: GPS for Personal Travel Surveys, Final Report, Columbus: Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, pp. 1-92.

Wolf, J. , Dr, D. and Guensler, R., (2000), Using GPS Data Loggers to Replace Travel Diaries in the Collection of Travel Data, Ph.D. dissertation, Georgia Institute of Technology.

Yalamanchili, L., Pendyala, R., Prabaharan, N. and Chakravarthy, P., (1999), Analysis of Global Positioning System-Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior, Transportation Research Record: Journal of the Transportation Research Board, 1660, pp. 58-65., http://dx.doi.org/10.3141/1660-08

Zhang, L., Qiang, M. and Yang, G., (2013), Mobility Transportation Mode Detection Based on Trajectory Segment, Journal of Computational Information Systems, 8, pp. 3279 – 3286.

Zheng, Y., Chen, Y., Li, Q., Xie, X. and Ma, W.-Y., (2010), Understanding Transportation Modes Based on GPS Data for Web Applications, ACM Transactions on the Web, 4(1), pp. 1-36., http://dx.doi.org/10.1145/1658373.1658374

Zheng, Y., Liu, L., Wang, L. and Xie, X., (2008), Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web, Proc. 17th International Conference on World Wide Web - WWW '08, Beijing, China, April 21 – 25, 49, pp. 247 – 256., http://dx.doi.org/10.1145/1367497.1367532