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

Angel J. Lopez*†, Ivana Semanjski*, Dominique Gillis*, Daniel Ochoa† and Sidharta Gautama*
*Department of Telecomunications and Information Processing
Ghent University
St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium
Emails: {angel.lopez, ivana.semanjski, dominique.gillis, sidharta.gautama} @ugent.be
†Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, ESPOL
Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
Emails: {alopez,dochoa} @espol.edu.ec

***Abstract*—Traditional travel survey methods have been widely used for collecting information about urban mobility although, since middle of the 90's Global Position System (GPS) has become an automatic option for collecting more precise data of the households. But how good is the collected data? many studies on mobility patterns have focused on the GPS advantages and leaving aside its issues. However, when it comes to extract the frequency of the trips and travelled distance this technology faces some gaps due to 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, middle 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 the GPS and it affects more to 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 on GPS.**

***Keywords*—Data quality; travelled distance; CAN-BUS data; GPS data**

I. INTRODUCTION AND RELATED WORK

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) [1]. Yet, respondents have the tendency to omit short stops, such as post office and ATM and when it comes to numerical answers, travel time is rounded to simple values like 10, 15, 30 minutes interval. Likewise it happens to the travelled distance [2].

Nonetheless, with the introduction of the Global Position System (GPS) and its first adoption on transportation studies in the middle of the 90's, where Wagner [2] reports one of the first studies that uses GPS for collecting information of 100 households through logger devices installed in their vehicles, since that it has been extensively used in combination with others datasets. Studies [3], [4], and [5] have already demonstrated the possibility to use GPS data in Transportation research by capturing the characteristics of different types of trips. Later studies, such as [6] and [7] 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, for instances 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 studies [8] and [9], 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 walk segment may be changes of transportation mode and Liao et al. [10] uses an probabilistic approach to estimate those changes. Zheng et al. [11] shows that extracting trip frequencies is a challenging task, despite his method classifies 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 [12].

However, signal reception and time-to-first-fix (TTFF) are well known issues of GPS, both of them affecting the reported distance [13]. 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 aka *urban canyoning errors* because they appear in urban canyons where the signal is reflected by buildings [14]. TTFF aka *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 [15].

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) [16]. 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 [17]. 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 [18], 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 II describes the datasets and methods for assessing GPS gaps on the car mode. Section III presents the case study and the outcomes of this research and Section IV summarise the findings by drawing conclusions on it.

II. Methodology

The data from the present research is drawn from two sources (a) a car-sharing company named *Cambio* [19] that opened its operation in Belgium in 2002, and at this time (June 2016), it is available in 35 cities with 369 stations, 862 cars and more than 24.000 users. *Cambio* provided us a dataset with 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.

A. 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 a ground truth. The access to that such a data is possible through the CAN-bus. A CAN-bus is known as a protocol for high performance and high reliable serial communication links between electronic control units in the field of automotive and industrial control applications [17], for example 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 [20] where each element on the network can send a message (e.g., temperature, state of charge) independently to the entire network, being the bus priority defined by the message identifier [21]. 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 to its granularity, which is 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 for instances time doing shopping, time visiting someone, etc. Nonetheless, the travelled distance and the starting time (first trip segment) are not affected for the aforementioned limitation.

GPS has been used in several studies (Section I) and 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 it to a centralised repository through the network. For collecting the data, a GPS logger *GenLoc 53e* [22] 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. A 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 jth 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.

B. Data quality

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

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

$$\text{Missing distance}_{\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \Big(d(R_i) - \sum_{S_j^{(i)} \in R_i} d(S_j^{(i)}) \Big) \tag{3}$$

$$\text{TTFF}_{\text{avg}} = \frac{1}{N} \sum_{i=1}^{N} \Big(t(S_1^{(i)}) - t(R_i) \Big) \tag{4}$$

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_i, and N is the number of observations.

Equation (2) is the proportion of 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) 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 and (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.

C. Land use

A 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 for the high density of large structures (e.g., buildings, bridges). A GIS tool allows to identify the land use of a trip [23] (i.e., whether a trip was performed on either rural or urban area) in such a way that trips are matched within an administrative area (boundary area classified as rural or urban) through 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) [24]. 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 [25].

D. Regression method

To explain the gaps in GPS data, the factors that influence the 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 aka covariates, that relationship is modelled through a error term ε, a random variable that adds noise to the linear relationship between the dependent and independent variables [26], therefore the model is expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \quad (5)$$

where $\mathbf{y}$ is the dependent variable or response variable, $\mathbf{X}$ represents the explanatory variables, $\boldsymbol{\beta}$ the regression coefficients or effects and ε is the error term.

III. 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 as from the personal perspective. From the vehicle perspective, GPS loggers were installed

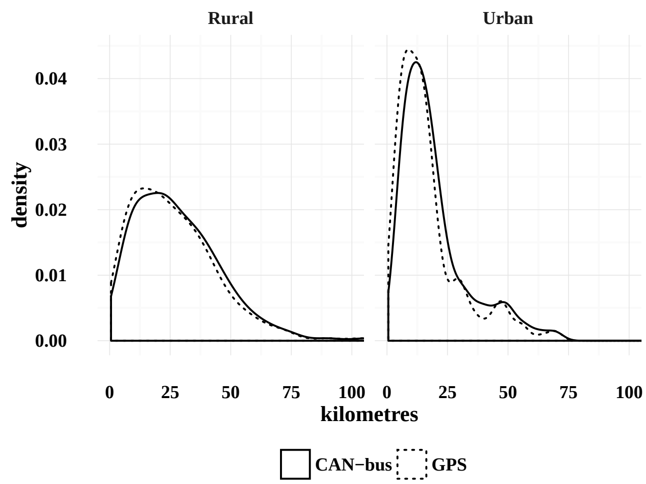

Figure 1. Travelled distance density based on land use for GPS logger and CAN-bus data

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.

A. Descriptive analysis

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

TABLE I. 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 get an analysis at different levels, we group the trips based on the travelled distance into three categories, such as short, middle and long trips. Table II shows the conditions for those groups.

TABLE II. 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.

The missing distance for the GPS logging can be noticed in Fig. 1, where the GPS density curve is shifted to the left side with respect to CAN-bus curve in both rural and urban areas, this is another indication of missing distance based on the land use.

B. 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 shows in (2), it turns out that on average 9% of the travelled distance is not captured by the GPS logger (Fig. 2). It means that for an average trip with distance 23.9 km the GPS will report on average 2.1 km less (Table I).

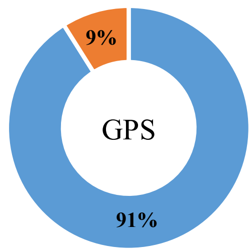

Figure 2. Global missing distance for a car mode using GPS

From the groups, we calculate the travelled distance for both GPS and CAN-bus data. Fig. 3 shows the missing data for short, middle and long trips within groups of different land used, in which, urban areas affect more to the short and middle distance trips (around 81% and 83% respectively) than the long trips (Table III). Besides, the GPS logging performs better for long trips in both urban and rural areas, logging 92% and 95% respectively.

TABLE III. A SUMMARY OF THE GPS GAPS BASED ON THE LAND USE

			TTFF (min)		Missing distance (km)	
	Group	Trips	Median	Mean	Mean	Percentage
Rural	Short	82	4.6	7.8	0.9	11.9%
	Middle	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%
	Middle	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 the TTFF using (4), which makes use of the starting time as an argument.

C. 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 is 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. Table IV shows a full description of the covariates.

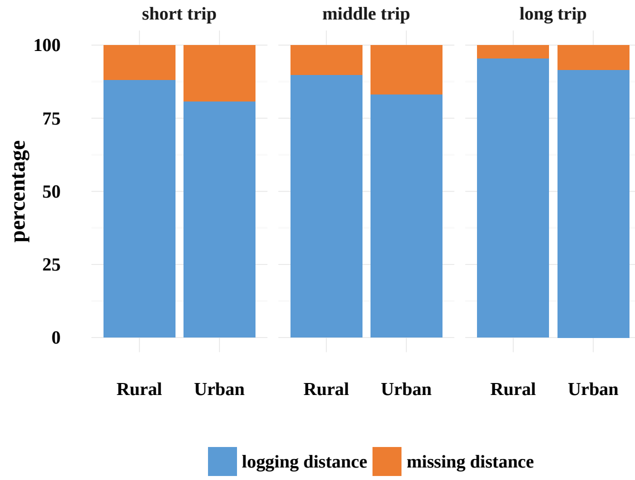

Figure 3. Percentage of missing data for different group of trips

TABLE IV. 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)
Time spent per trip	Average time spent per trip
Land use	Land use for the trip (e.g., urban, rural)

A summary of the fitted models is depicted in Table V, where the covariate coefficient is estimated using a regression model and its significant level is based on the p-value. Based on *model 3*, the number of *trip segments* seem to be not 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 parking) is important factor for modelling the travelled distance because it adds more periods of cold/warm start to the GPS logger (i.e., a higher waiting time increases the chances of missing data). And 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 segments* (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.

IV. CONCLUSION

In this paper, we measure the gaps in the GPS-based distance using CAN-bus data as a ground truth, likewise the factors that influence those gaps were identified through regression models. It was found that on averages 9% of the travelled distance is not captured for the GPS logger, this is important, considering that many mobility studies are

 ISBN: 978-1-61208-510-4

TABLE V. MODELS RESULT

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

being conducted using GPS data, and their outcomes might underestimate the actual travelled distance. Although, 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 about 13%, which is a clear indication that urban areas are more susceptible to issues related to the signal reception, affecting 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 also contrast with other study [27] that reports 4% of 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 is not having a big 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 differences across the collection methods, for example, passive and active logging.

Acknowledgment

The authors thank their respective institutions, Ghent University, Escuela Superior Politécnica del Litoral (ESPOL), and The National Secretary of Higher Education, Science, Technology and Innovation of Ecuador (SENESCYT) for its support through the scholarship program Open Call 2012-I.

References

[1] S. Handy, “Methodologies for exploring the link between urban form and travel behavior,” *Transportation Research Part D: Transport and Environment*, vol. 1, no. 2, pp. 151–165, 1996.

[2] D. P. Wagner, “Lexington area travel data collection test: GPS for personal travel surveys,” *Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus*, pp. 1–92, 1997.

[3] L. Yalamanchili, R. Pendyala, N. Prabaharan, and P. Chakravarthy, “Analysis of Global Positioning System-Based Data Collection Methods for Capturing Multistop Trip-Chaining Behavior,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1660, pp. 58–65, jan 1999. [Online]. Available: http://trrjournalonline.trb.org/doi/10.3141/1660-08

[4] G. Draijer, N. Kalfs, and J. Perdok, "Global Positioning System as Data Collection Method for Travel Research," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1719, pp. 147–153, jan 2000. [Online]. Available: http://trrjournalonline.trb.org/doi/10.3141/1719-19

[5] J. Wolf, D. Dr, and R. Guensler, "Using GPS data loggers to replace travel diaries in the collection of travel data," Ph.D. dissertation, 2000.

[6] A. Bolbol, T. Cheng, I. Tsapakis, and J. Haworth, "Inferring hybrid transportation modes from sparse GPS data using a moving window SVM classification," *Computers, Environment and Urban Systems*, vol. 36, no. 6, pp. 526–537, nov 2012. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0198971512000543

[7] N. Schuessler, K. Axhausen, and N. Schüssler, "Processing Raw Data from Global Positioning Systems Without Additional Information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2105, pp. 28–36, dec 2009. [Online]. Available: http://trrjournalonline.trb.org/doi/10.3141/2105-04

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

[9] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, "Understanding transportation modes based on GPS data for web applications," *ACM Transactions on the Web*, vol. 4, no. 1, pp. 1–36, 2010.

[10] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, "Learning and inferring transportation routines," *Artificial Intelligence*, vol. 171, no. 5-6, pp. 311–331, apr 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0004370207000380

[11] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw gps data for geographic applications on the web," in *Proceeding of the 17th international conference on World Wide Web - WWW '08*, no. 49. New York, New York, USA: ACM Press, 2008, p. 247. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1367497.1367532

[12] A. J. Lopez, D. Ochoa, and S. Gautama, "Detecting Changes of Transportation-Mode by Using Classification Data," in *18th International Conference on Information Fusion (Fusion 2015)*. Washington, DC: Information Fusion (Fusion), 2015, pp. 2078–2083. [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp={&}arnumber=7266810{&}isnumber=7266535

[13] M. Flamm, C. Jemelin, and V. Kaufmann, "Combining person based GPS tracking and prompted recall interviews for a comprehensive investigation of travel behaviour adaptation processes during life course transitions," in *11th World Conference on Transportation Research*, Berkeley, 2007.

[14] N. Schuessler and K. Axhausen, "Identifying trips and activities and their characteristics from GPS raw data without further information," Tech. Rep., 2008. [Online]. Available: http://e-collection.library.ethz.ch/view/eth:30471

[15] P. Stopher, Q. Jiang, and C. FitzGerald, "Processing GPS Data from Travel Surveys," in *2nd international colloqium on the behavioural foundations of integrated land-use and transportation models: frameworks, models and applications*, Toronto, 2005, pp. 1–21.

[16] I. S. O. Standard, "Iso 11898, 1993," *Road vehicles–interchange of digital information–Controller Area Network (CAN) for high-speed communication*, 1993.

[17] K. Turski, "A global time system for CAN networks," in *Proceedings of the 1st International CAN Conference*, vol. 13, no. 3, Mainz, 1994, pp. 31–36.

[18] A. Suvinen and M. Saarilahti, "Measuring the mobility parameters of forwarders using GPS and CAN bus techniques," *Journal of Terramechanics*, vol. 43, no. 2, pp. 237–252, 2006.

[19] "Cambio Belgium a car-sharing company," http://www.cambio.be, retrieved: July, 2016.

[20] M. Farsi, K. Ratcliff, M. Barbosa, K. Ratcliff, and M. Farsi, "An overview of Controller Area Network," *Computing & Control Engineering*, vol. 10, no. June, pp. 113–120, 1999. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/cce{_}19990304

[21] R. Bosch, "CAN Specification Version 2.0," *Rober Bousch GmbH, Postfach*, vol. 300240, p. 72, 1991.

[22] "Erco and Gener genlog53e," http://www.ercogener.com, retrieved: July, 2016.

[23] P. Stopher, E. Clifford, J. Zhang, and C. FitzGerald, "Deducing mode and purpose from GPS data," *Institute of Transport and Logistics Studies*, 2008.

[24] "OpenStreetMap contributors," http://www.openstreetmap.org/copyright, retrieved: July, 2016.

[25] H. H. Hochmair, D. Zielstra, P. Neis, and P. N. Hartwig H. Hochmair, Dennis Zielstra, "Assessing the Completeness of Bicycle Trail and Designated Lane Features in OpenStreetMap for the United States and Europe," *Proceedings of the Transportation Research Board 92nd Annual Meeting*, vol. 19, no. 1, pp. 1–21, 2013.

[26] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied linear statistical models*. Irwin Chicago, 1996, vol. 4.

[27] S. Vlassenroot, D. Gillis, R. Bellens, and S. Gautama, "The Use of Smartphone Applications in the Collection of Travel Behaviour Data," *International Journal of Intelligent Transportation Systems Research*, pp. 17–27, 2014. [Online]. Available: http://link.springer.com/10.1007/s13177-013-0076-6