



Proceedings of the 20th International Conference of
Hong Kong Society for Transportation Studies (HKSTS)

Urban Transport Analytics

12-14 December 2015 • Hong Kong

Edited by

Janny M.Y. LEUNG C.H. CHENG

Sylvia Y. HE Yong-Hong KUO

Hong Kong Society for Transportation Studies



Printing and Receiving Orders:
Organizing Committee
The 20th International Conference of
Hong Kong Society for Transportation Studies (HKSTS)
c/o Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
6/F, William M.W.Mong Engineering Building
The Chinese University of Hong Kong
Shatin, New Territories
Hong Kong SAR
China

December, 2015

Urban Transport Analytics
Proceedings of the 20th International Conference of
Hong Kong Society for Transportation Studies (HKSTS)
12-14 December 2015

Copyright © 2015
Hong Kong Society for Transportation Studies Limited
Website: <http://www.hksts.org/conf.htm>
All rights reserved

No part of this publication may be reproduced, transmitted,
transcribed, stored in a retrieval system, or translated into any
language, in any form or by any means without written
authorization from Hong Kong Society for Transportation
Studies Limited.

ISBN: 978-988-15814-4-0



INVESTIGATING THE MOBILITY HABITS OF ELECTRIC BIKE OWNERS THROUGH GPS DATA

Angel J. LOPEZ ^{a,c}, Paola ASTEGIANO ^b, Chris M.J. TAMPÈRE ^b, Sidharta GAUTAMA ^a and Carolien BECKX ^d

^a *Department of Telecommunications and Information Processing, Ghent University, St-Pietersnieuwstraat 41, B-9000 Ghent, Belgium*

Email: angel.lopez@ugent.be, sidharta.gautama@ugent.be

^b *L-Mob Leuven Mobility Research Center, KU Leuven, Belgium*

Email: paola.astegiano@kuleuven.be, chris.tampere@kuleuven.be

^c *Facultad de Ingeniería Mecánica y Ciencias de la Producción, Escuela Superior Politécnica del Litoral, ESPOL,*

Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador

Email: alopez@fiec.espol.edu.ec

^d *Transport and Mobility Unit, VITO NV, Belgium*

Email: carolien.beckx@vito.be

ABSTRACT

This paper investigates the mobility habits of electric bike owners as well as their preferred routes. Through a GPS tracking campaign conducted in the city of Ghent (Belgium) we analyze the mobility habits (travel distance, time spent, speed) during the week of some e-bike users. Moreover, we propose the results of our map matching, based on the Hausdorff criterion, and preliminary results on the route choice of our sample. We strongly believe that investigating the behavior of electric bikes' owners can help us in better understanding how to incentivize the use of this mode of transport. First results show that the trips with a higher travel distance are performed during the working days. It could be easily correlated with the daily commuting trips (home-work). Moreover, the results of our map-matching highlight how 61% of the trips are performed using the shortest path.

Keywords: electric bike, GPS tracking, mobility habits, map matching

1. INTRODUCTION AND LITERATURE REVIEW

In recent decades, the promotion of non-motorized modes of transport is increasing as part of more sustainable eco-mobility vision. In this area e-bikes are gaining more and more popularity. Because of their higher speeds (compared to the ordinary bike) and longer reach, they extend the capabilities of normal cycling and could be an attractive alternative to the car. In the near future they could become the best way to incentivize a shift from car to bicycle in order to reduce road congestion, traffic-related air pollution, road accidents and infrastructure costs. Currently, the world's leader e-bike market is China (MacArthur et al. 2014) but according to (Fishman & Cherry 2015) in the last few years, a positive trend of the e-bike market share was also observed in the Western Europe where Germany and the Netherlands cover together the 66% of this market.

In (Weinert et al. 2007) and (Cherry & Cervero 2007) the first investigation on how and why e-bikes developed so quickly in the eastern countries has been performed, providing important insights to policy makers in China and abroad. They showed how timely regulatory policy can influence the purchase choice of millions, incentivizing the use of a new mode of transport introduced in the market. Considering control strategies that limit the number of stops for this mode would also be an additional way to increase electric bike use, through signal coordination or grade separated intersections, thus decreasing the travel time of the electric bikes.

In this framework another important role is played by U.S. and North America. Their markets are still behind China and Europe, but a strong group of researchers (MacArthur et al. 2014), (Dill & Rose 2012), (Popovich et al. 2014) is investigating which factors influence purchase decisions in these countries and, with a comparison between ordinary bikes and e-bikes, trying to understand whether e-





bikes can effectively address barriers to bicycling and therefore encourage more sustainable mobility. Their results suggest that e-bike users cycle more often and to more distant locations. Moreover, e-bikes allow people with physical limitations to cycle thanks to electric assist.

This paper, unlike the aforementioned studies, investigates the mobility habits of e-bike owners through a GPS tracking campaign instead of employing just a travel survey. In the field of analysis of GPS data, groundbreaking studies (Yalamanchili et al. 1999), (Draijer et al. 2000), (Wolf et al. 2000) aimed both at understanding whether it was possible to use GPS technology in travel studies and in evaluating the effectiveness of GPS in capturing different kind of trips. Over the years, there has been an increasing interest in the use of Global Position System (GPS) devices as a mean to measure people's journeys. GPS devices can collect and record data on points during a person's travel, providing the latitude, longitude (and altitude, if desired), the time, the speed, etc. However, these devices cannot collect data on the mode or on the purpose of the travel, both of which are essential components for transport planning purposes. Compared to the work of (Stopher et al. 2008) in which the authors developed a set of heuristics rules for determining the mode of transport, in our case the device is installed directly on the e-bike because our main focus is understanding the behavior of e-bike owners detecting their movements. Instead, compared to (Schuessler & Axhausen 2009), in which a post processing procedure that needs no other input than the most basic GPS raw data is shown, we are mainly interested in the route choice of our sample. In this research, a travel diary is also available, which will become fundamental in future research when the analysis of the trips performed with another mode of transport will be necessary in order to study the impact of the electric bike on the activity pattern of each user. Another field, strictly connected with managing GPS data, is related with the map matching methods. Unlike the work of (Bierlaire et al. 2013) in which the authors propose a probabilistic map-matching in order to overcome problems due to the poor quality of the GPS data coming from smartphones, we can take advantage of the good quality of the data coming from our devices and match them on the network provided by Open Street Map. Moreover, we are able to check the goodness of our map matching using the Hausdorff criterion (Huttenlocher et al. 1993).

The data used for this study comes from the SPRINT project (Astegiano et al. 2015). It was launched at the beginning of 2014 with the aim of helping the Flemish Government in better understanding how to improve the use of the electric bikes in Belgium. The goals of this project are both (i) the creation of a bicycle layer for the Flanders multimodal traffic model in order to ensure that (e-)bike remains a fundamental strategic mobility policy in Flanders and (ii) getting an answer for questions like: "Does promoting the e-bikes require a different approach compared to the ordinary bikes?", "Who are current e-bike owners?", "Which kind of infrastructure do they prefer?". In order to support the shift from car to more sustainable modes of transport it becomes indispensable to understand whether, for example, new cycle paths that are currently being built everywhere in Flanders would also be suitable for e-bikes.

Summarizing, the aim of this paper is to use the GPS data from the SPRINT project in order to provide statistical results related with the travel distance, the travel time and the speed of our users but also understanding whether and how their behaviour changes between their working days and their weekend. Furthermore, we provide the results of our map matching together with a preliminary analysis of the route choice of our sample, in which we research whether or not the topological shortest path is the preferred alternative.

The remainder of the paper is organized as follows: we introduce before the methodological part; we then go through our case of study; finally we present results and conclusions.

2. METHODOLOGY

The data set employed for our study has been obtained from (Astegiano et al. 2015) and contains information about socio-demographic characteristics and journey characteristics of electric bike owners during one year. This paper focuses on the analysis of two weeks (for a total of 439 trips) GPS data coming from this data set in which location points were collected using GPS data loggers installed directly on the participants' e-bike. The data logger reports a location point every five





seconds and stores it as a *data frame* into its internal memory. A data frame includes the following attributes: latitude, longitude, speed, accuracy, heading and timestamp.

2.1 Processing raw GPS data

In order to perform a complete data analysis, the individual data frames of the loggers were merged together into a single file; the process follows these four steps: (i) first, we verify the checksum of data frame to ensure its integrity, the data integrity may be compromised either due to memory error (storage device) or while the data are transferred to another storage device; (ii) second, a parse step identifies the data frame attributes and transforms numeric values into a right format (e.g. latitude, longitude, date and time); (iii) third, a validation step excludes the data frames that are marked as invalid by the data logger, because they do not provide any information neither location coordinates nor other attributes. The data logger marks the data frame as invalid when it is not possible to have a position although the date and time information are available; (iv) and finally, the valid data frame are merged into a single file.

2.2 Trip segmentation

A data exploration regarding the time lag among locations showed that most of the reported locations hold a lag within the 240 seconds. Considering the dwell time criterion, i.e. the minimum time difference between two GPS points after which it is assumed that an activity took place (Schuessler & Axhausen 2008), it provides us an indication for segmenting individual trips. Thus, we set a dwell time of 240 seconds (4 minutes) for segmenting location points into trips (i.e. a lag bigger than 240 seconds is considered as a new trip).

Using the spatial features of *PostGIS*, the location points that belong to a specific trip are aggregated into a single line (trip segment) with start and end points. It allows us to calculate summary statistics such as travel distance, time spent and average speed per trip. Besides these statistics, we use the trip segments for assessing the Routing and Map-matching stage.

2.3 Routing

After the trip segmentation stage we obtain trip segments (lines). It allows us to identify easily the start/end point of each trip; these points are used for routing the shortest path and alternative paths as well.

We rely on Google Maps API for computing the shortest path between each origin and destination pair. Moreover, up two additional paths were calculated for the same O/D pair. These two alternatives are more realistic than the traditional shortest path (Dreyfus 1969) because they consider elements such as travelled mode (bicycling), road direction, bicycle paths, closed roads and preferred streets.

The authors chose this routing approach because the map coverage for Belgium fulfills the details such as Map Tiles, Geocoding Traffic Layer, Driving Directions, Biking Directions, and Walking Directions. Readers are invited to find out further information regarding Google Maps API¹ and Map Coverage Details² on the websites.

The API output is a text file based upon JSON format (JavaScript Object Notation) that includes features as the path, distance and time estimation. Suggested paths are matched among user's trips in order to check which one was taken by the user.

2.4 Map Matching

The aim of this paper is investigating the mobility behavior of the e-bike owners. Therefore, it is necessary to realize what are the preferred routes for electric bike owners and the trips for which the e-bike is the most used mode of transport. To accomplish it, we performed a map-matching between the participants' trips and the suggested paths (from Google) where Hausdorff distance approach (Huttenlocher et al. 1993) has assessed in order to identify which alternative matched better to the real trip.

¹ Google Map API, <https://developers.google.com/maps/documentation/javascript/directions>

² Map Coverage Details, <https://developers.google.com/maps/coverage>





2.4.1 Hausdorff distance

Given two finite point set $A = \{a_1, \dots, a_p\}$ and $B = \{b_1, \dots, b_q\}$, the Hausdorff distance is defined as

$$H(A,B) = \max(h(A,B), h(B,A)) \quad (1)$$

where $h(A,B)$ ranks each point of A based on its nearest point of B and uses the most mismatched point of A , and $\| \cdot \|$ is some underlying norm on the points of A and B .

$$h(A,B) = \max_{a \in A} \min_{b \in B} \| a - b \| \quad (2)$$

Thus, it is a measure of the degree of mismatch between two trip segments.

3. CASE STUDY

The data set employed for this study contains information about 439 trips performed by electric bike in the city of Ghent (Belgium). In the next subsections we firstly show some descriptive statistics (speed, travel distance, travel time), then we investigate how people behave differently during working days and weekend, finally we propose the results of our map-matching algorithm.

3.1 Trip features

In this section, we analyze the trip features such as *distance*, *duration* and *speed*. The units of measurement are *meters*, *minutes*, and *kilometers per hour* respectively. A summary of the aforementioned features is given by the descriptive statistics that are depicted on the table below.

Table 1. Descriptive Statistics

Variable	Trips	Mean	Std. Error	Std. Dev.	Max.	Min.	Total
Distance	439	5356.2	381.9	8001.1	47753.8	0.0	2351350.7
Duration	439	16.3	1.0	21.2	166.8	0.0	7168.3
Speed	439	14.6	0.5	9.7	45.9	0.1	--

Table 1 shows the average measure of a representative trip in our data set: distance of 5km, duration of 16 minutes and a speed of 15 km/h. Not satisfied by these results (we expected a higher average distance) we have further investigated these features.

Using the Ward Hierarchical Clustering (Ward 1963) we clustered the data set in 4 groups based upon the travel distance.

Table 2. Cluster groups based upon the travel distance

Group	Percentage	Trips	Avg. distance	Avg. duration	Avg. speed
1	47.4	208	228.7	2.5	4.8
2	23.2	102	2989.8	13.1	15.4
3	15.9	70	9152.3	28.9	19.9
4	13.4	59	23019.8	55.7	26.6

Despite the average distance of 5km, Table 2 shows two groups for which the distance is bigger than 9km. The number of the short trips strongly influences the final average. We deduce how groups 3 and 4 include mainly the working trips.

We finally represent these trip features with the kernel density distributions (Figure 1). The mean value is represented as a dashed line.



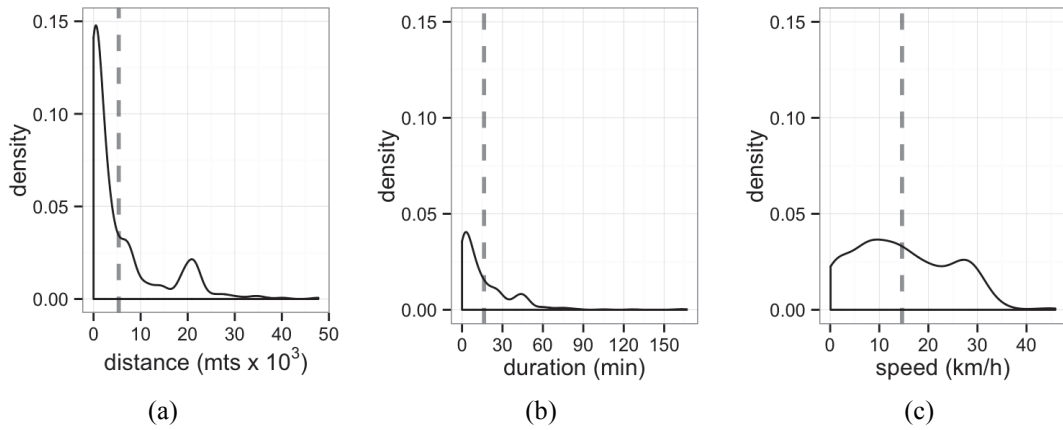


Figure 1. Density distributions

3.2 Trip estimation

In this subsection, we focus on how our users behave differently during working days and weekend. Analyzing in details our groups, we can notice, from the distance distribution (Fig. 2), that the longest trips (about 20km) are performed during the working days.

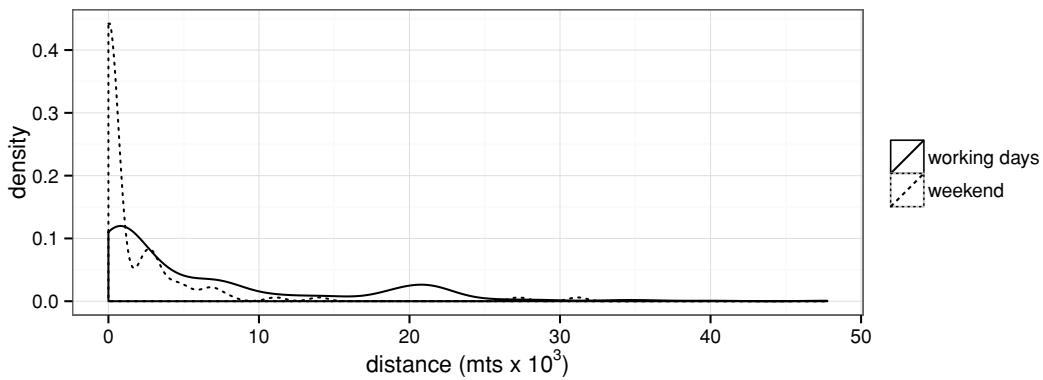


Figure 2. Travel distance by week

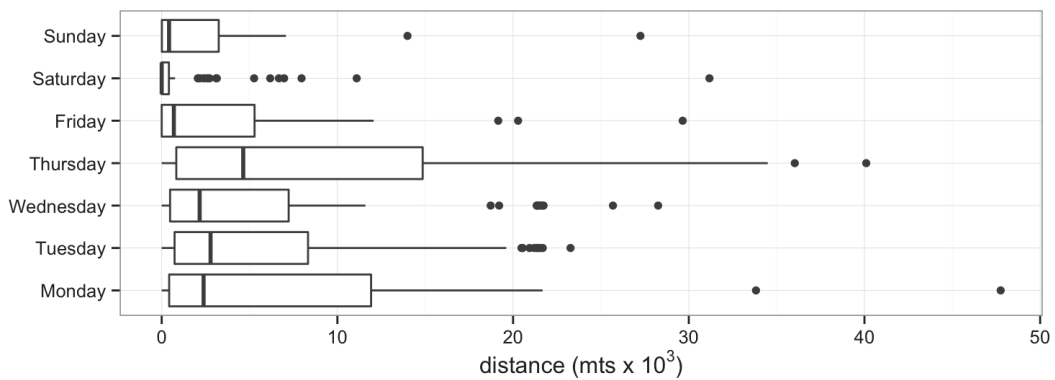


Figure 3. Travel distance and its variance

Fig. 2 shows the travelled distance by day while Fig. 3 the average travelled distance with its variance for each day of the week.





Moreover, a local polynomial regression is fitted on the trip features for estimating the number of trips (Figure 4) and speed (Figure 5) during the day.

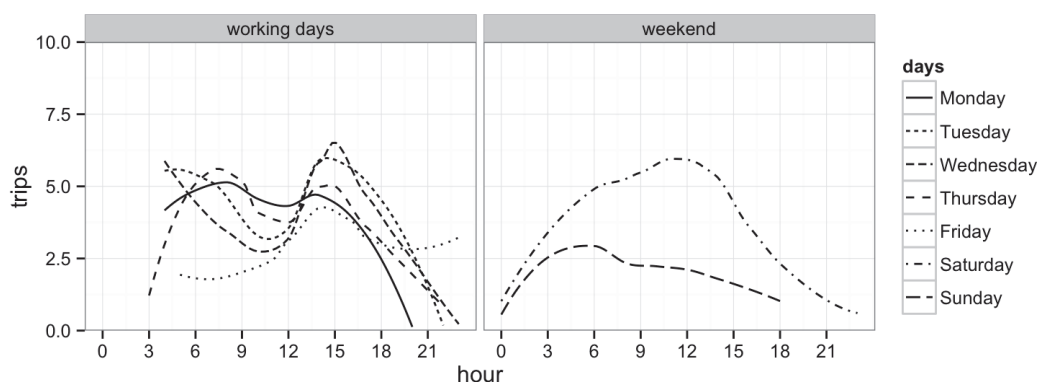


Figure 4. Trip estimation by hour

The results depicted in Figure 4 show how the trips are widely spread in the morning (from 05:00 to 10:00) rather than in the afternoon (from 14:00-17:00), although the second range is grouped a higher number of trips. Additionally, we can recognize how some trips extend during the night especially on Friday and weekend.

Focusing on the peak hours (morning peak 07:00-10:00 and evening peak 16:00-19:00 Belgian standards), the participants, as expectable, prefer cycling during the off-peak; they even ride faster during the off-peak than the on-peak (Table 3).

Table 3. Cluster groups based upon the travel distance per week

Peak	Group	Percentage	Trips	Avg. distance	Avg. duration	Avg. speed
off peak	1	33.7	148	196.3	2.2	4.6
off peak	2	11.8	52	3006.5	13.0	15.5
off peak	3	9.3	41	8963.6	28.1	20.2
off peak	4	10.5	46	23371.1	54.9	27.2
morning peak	1	8.9	39	303.2	3.3	4.7
morning peak	2	7.7	34	2868.5	14.2	13.5
morning peak	3	3.6	16	10281.3	33.8	19.1
morning peak	4	1.8	8	20017.5	47.3	25.8
evening peak	1	4.8	21	318.7	3.1	5.6
evening peak	2	3.6	16	3192.9	10.9	19.1
evening peak	3	3.0	13	8357.9	25.6	20.0
evening peak	4	1.1	5	24591.5	76.7	22.6

Analyzing the speed in details, it is lower during the first half of the day than in the second one. It reaches its maximum value between 8:00 and 9:00 whereas it keeps increasing during the evening. During the weekends, instead, the speed reaches its maximum value between 14:00 and 15:00 (Fig.5).



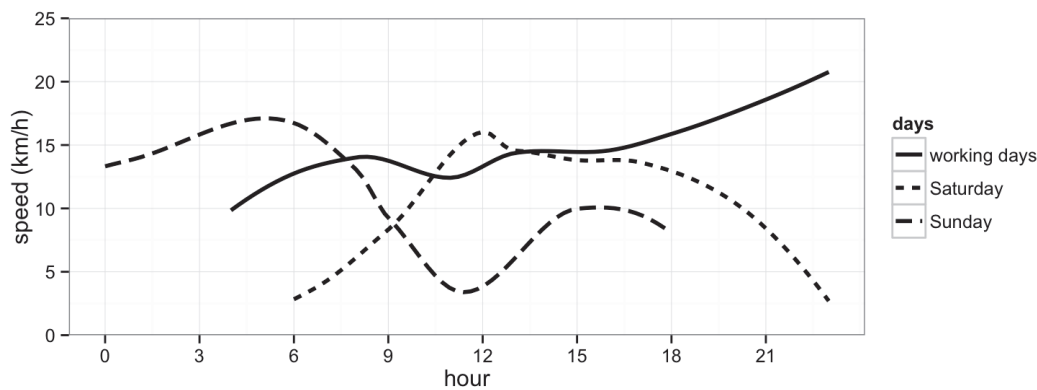


Figure 5. Speed estimation by hour

3.3 Map Matching

In order to understand which are the preferred routes for e-bike owners, we performed a map-matching algorithm based on the Hausdorff distance criterion between the participants' trips and the suggested paths by Google. For each O/D pairs, we provide three different route alternatives and we analyze which ones have been chosen by our users.

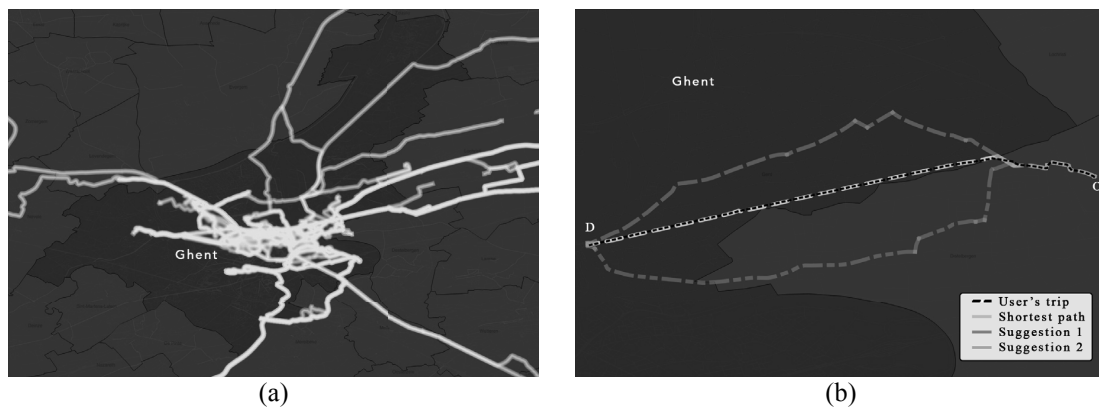


Figure 6. (a) Trip density map; (b) Example of routing alternative paths

In Figure 6a the density map of our sample is shown, while in Figure 6b we show an example of our approach for one O/D pair. In green, orange and light blue (the shortest path) the route alternatives suggested by Google are shown while with the dashed line that one chosen by our user (in this case it coincide with the shortest path). Using the Hausdorff distance as criterion, we have 65% of the trips that match one of the alternatives proposed by Google. Moreover, considering the whole sample, the 75% of the trips are performed using the shortest path, even if this result is influenced by the fact that for a subset of O/D pairs Google provide only one alternative (the shortest path). If we consider only the O/D pairs for which Google provide at least two alternatives, the 61% of them coincide with the shortest path.

4. CONCLUSIONS

In this paper, we analyze GPS data from 439 trips performed with an electric bike in the city of Ghent. The average distance for short trips is about 5 km while for longer trips is up to 30 km. Looking at the average speed we noticed how the longer trips are correlated with a higher average speed. This can be related with the most confident cyclists. Going in details of the different subgroups we found that 13% of the trips have an average travel distance of 22.8 km and the majority of them are performed during the working days. It could be easily correlated with the daily commuting trips (home-work).



Therefore, it leads us to suppose that commuting trips are appealing for being ridden by e-bike and consequently car's usage could consistently decrease for this kind of trips. In the next future, this analysis will be supported by the presence of the travel diary that could be better explained this kind of relation. The results of our map-matching algorithm show how 61% of the trips are performed using the shortest path. Future research directions will investigate which kind of attributes affect people for which the preferred route is not the shortest path.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge financial support by project 3.3 in the SPRINT Policy Research Centre on Traffic Safety, and by the Vlaamse Overheid, SBO project "Urban Logistics and Mobility" (IWT-140433).

6. REFERENCES

- Astegiano, P., Tampère, C.M.J. & Beckx, C., (2015). A preliminary analysis over the factors related with the possession of an electric bike. In *Transp. Res. Procedia*, 10, pp. 393–402
10.1016/j.trpro.2015.09.089
- Bierlaire, M., Chen, J. & Newman, J., (2013). A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies*, 26, pp.78–98.
- Cherry, C. & Cervero, R., (2007). Use characteristics and mode choice behavior of electric bike users in China. *Transport Policy*, 14(3), pp.247–257.
- Dill, J. & Rose, G., (2012). Electric Bikes and Transportation Policy. *Transportation Research Record: Journal of the Transportation Research Board*, 2314(15 mi), pp.1–6.
- Draijer, G., Kalfs, N. & 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.
- Dreyfus, S.E., (1969). An Appraisal of Some Shortest-Path Algorithms. *Operations Research*, 17(3), pp.395–412.
- Fishman, E. & Cherry, C., (2015). E-bikes in the Mainstream: Reviewing a Decade of Research. *Transport Reviews*, pp.1–20.
- Huttenlocher, D.P., Klanderman, G. a. & Rucklidge, W.J., (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9), pp.850–863.
- MacArthur, J., Dill, J. & Person, M., (2014). Electric Bikes in North America. *Transportation Research Record: Journal of the Transportation Research Board*, 2468, pp.123–130.
- Popovich, N. et al., (2014). Experiences of electric bicycle users in the Sacramento, California area. *Travel Behaviour and Society*, 1(2), pp.37–44.
- Schuessler, N. & Axhausen, K., (2008). Identifying trips and activities and their characteristics from GPS raw data without further information.
- Schuessler, N. & 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.
- Stopher, P.U. of S. et al., (2008). Deducing mode and purpose from GPS data. *Institute of Transport and Logistics Studies*.
- Ward, J.H., (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), p.236.
- Weinert, J., Ma, C. & Cherry, C., (2007). The transition to electric bikes in China: History and key reasons for rapid growth. *Transportation*, 34(3), pp.301–318.
- Wolf, J., Dr, D. & Guensler, R., (2000). *Using GPS data loggers to replace travel diaries in the collection of travel data*. Georgia Institute of Technology.
- Yalamanchili, L. et al., (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.

