The Potential of Wi-Fi Data to Estimate Bus Passenger Mobility

Léa Fabre*[1], Caroline Bayart[2], Patrick Bonnel[3], Nicolas Mony[4]

[1] PhD student, Université Lumière Lyon 2, France

[2] Senior lecturer, Université Claude Bernard Lyon 1, France

[3] Professor HDR, Université Lumière Lyon 2, France

[4] Project manager, Explain, France

SHORT SUMMARY

Last decades have been marked by several socio-economic transformations which have a strong impact on individual mobility behaviors. Using technologies such as Wi-Fi allows to gather passive mobility data, useful for ensuring the sustainable development of transport infrastructures. However, some work still needs to be done to address the limits of Wi-Fi sensors. Our research presents interesting solutions for sorting the transmitted signals and estimating Origin-Destination matrices. With a partitioning algorithm, it is possible to automatically distinguish passengers to get transit ridership flow and O-D matrices. The originality of the paper consists in comparing the results with those of other data sources, and in proposing a methodology that can be reproduced. The findings show that the algorithm is efficient and transferable. They provide concrete and replicable solutions to transport operators for understanding travel demand and managing the quality of service.

Keywords: Big data, clustering, public transport, trajectory reconstruction, Wi-Fi/Bluetooth sensors

1. INTRODUCTION

Last decades have been marked by several socio-economic transformations, as well as the emergence of new transport modes. All these changes have a strong and direct impact on individual mobility behaviors which are less and less regular. Hence, transport planners are constantly in search for accurate data, to feed transport engineering models to ensure a smarter and environmentally friendly mobility (Bonnel, 2004). The use of technologies such as Wi-Fi and Bluetooth sensors (Hidayat et al., 2018; Ji et al., 2017; Nitti et al., 2020; Pu et al., 2021) appears as a promising opportunity to collect mobility data. Compared to traditional surveys, Wi-Fi and Bluetooth sensors are easy to implement, and the detection of connected objects is totally passive (Nitti et al., 2020). Data gathering and data processing are thus faster and less expensive, allowing large amounts of mobility data to be obtained (Traunmueller et al., 2017). However, this technology presents some limits: the sensors do not capture strictly all the objects present in the vehicle and some passengers do not hold any connected objects, while others have several (Nitti et al., 2020). Also, the main challenge of Wi-Fi sensors is to identify the signals coming from passengers among the huge number of detected signals. Most studies (Araghi et al., 2012; Dunlap et al., 2016; Ji et al., 2017; Kurkcu and Ozbay, 2017; Myrvoll et al., 2017; Nitti et al., 2020) overcome this problem using threshold values to eliminate "unwanted signals". These values are context-specific, which makes it difficult to replicate. Very few recent studies refer to alternative methods (partitioning) to separate passengers from non-passengers (Afshari et al. 2019, Pu et al. 2021). These studies, as most of the ones previously cited, only consider data collected on a single vehicle and on only a few trips. They do not allow for a complete representation of mobility on the network, neither for temporal variations. To address these limits, this study aims at identifying methods to select signals coming from bus passengers which are no longer dependent on threshold values. That is to say, to develop a method which could be replicable on other bus lines and other agglomerations. Hence, it uses a partitioning algorithm to select signals coming from bus passengers. The results are confronted with several sources of real data commonly used in transport planning (optical counts and O-D onboard survey). The developed method is first applied to data collected on a small period of time, then we increase the perimeter and the length of the data collection, to prove the replicability of the method. The objective is to confirm the quality of mobility data gathered continuously in public transport by Wi-Fi sensors, validate the use of clustering algorithms to get O-D matrices from these passive data, and show the replicability of this method.

2. METHODOLOGY

As there is a lot of noise in the data collected with Wi-Fi sensors, some pre-processing has been carried out. First, in line with the literature, some variables were computed. These variables can be related to a signal or an object. Then, some objects considered as "non-valid" are removed. It concerns objects with a travel distance or duration null or greater than the maximum observed on the network. At this point, the size of the database is already considerably reduced.
To select signals emitted by bus passengers we apply a clustering algorithm which is not context dependent. The K-means algorithm was chosen for its simplicity and computation speed. It allows to split observations of a set of points $x_1, x_2, \dots, x_n$, in k clusters $S_1, S_2, \dots, S_k$, so that the distance between points of a same cluster is minimized:

$$min \sum_{i=1}^{k} \sum_{x_j \epsilon S_i} d(x_j, \mu_i) \qquad (1)$$

with μ_i the barycentre of the points in S_i.
Before applying the K-means algorithm, we processed a principal component analysis (PCA) on all the variables computed, from which we kept the smaller number of PCs providing at least 80% of the total variance. These PCs were used as input variables for the K-means algorithm. To find the best appropriate number of clusters expected, we implemented the elbow method. This led to a number of 10 clusters.
Once the clusters are formed, some are defined as passengers signal and others as unwanted signals, based on the properties of the clusters. Clusters identified as "non-passengers" are then eliminated, and the Origin-Destination matrix is finally built from the passenger signals clusters.

3. RESULTS AND DISCUSSION

The data collection sensor used in this study is called "Laflowbox". It measures mobility thanks to an electromagnetic wave sensor which detects connected objects (smartphone, smartwatch…). Two data collections were used for this study. The first one took place on the 11th of October 2018, on the line T3 of the TEOR (Transport Est Ouest Rouennais) network, in Rouen, France. The second lasted three months from November 2019 to January 2020, on lines T1, T2 and T3. To compare Wi-Fi results with ground truth data, we used reliable behavioral data coming from a face-to-face Origin-Destination on-board survey and optical counts coming from cameras positioned in buses.

Clustering

By applying the K-means algorithm to our dataset, we get 10 clusters of detected tracks. Almost 70% of the total variance is explained by the intergroups variance, so the clustering method used separates the tracks quite well. Based on the characteristics of the different groups identified by the clustering (**table 1**), we can classify the tracks as coming from “passengers” or “non-passengers”.

Table 1: Main characteristics of each cluster

Clusters	Observations	Number of detections per track	Track speed (m/s)	Distance from track barycenter std deviation (m)	Track signal strength std deviation (dBa)	Distance from next signal std deviation (m)	Time lapse from next signal std deviation (s)
1	10.8%	8	2.02	24	1.66	1209	32582
2	1.6%	78	3.97	383	4.65	307	14011
3	4.7%	5	4.15	62	2.08	2508	70467
4	50.9%	8	1.34	13	1.45	60	9987
5	10.5%	6	7.50	30	1.47	294	22757
6	7.5%	6	2.96	45	1.80	528	165416
7	1.8%	11	2.82	105	13.46	469	31069
8	0.2%	240	4.09	611	5.51	190	7787
9	7.7%	13	4.37	240	3.99	656	35650
10	4.3%	16	4.74	616	5.11	876	41380

Clusters 2, 8, 9 and 10 seem to be composed of signals emitted from connected objects belonging to bus passengers. We can then build the Origin-Destination matrix from the tracks of the selected clusters and compare it to ground-truth data, to determine if the methodology leads to relevant results.

Comparison to ground truth data

For comparison with the first data collection campaign, we dispose, of optical counts and of an O-D onboard survey implemented on the same day on all vehicles circulating on line T3. We look at several indicators to evaluate the correspondence between two data sources. Comparing Wi-Fi data and optical counts, the R-squared values are all above 0.70 for the distribution of trip starting/ending stops and above 0.80 for the bus loads. This reflects a good fit between the two data sources. **Figure 1** shows the relative bus loads for line T3.

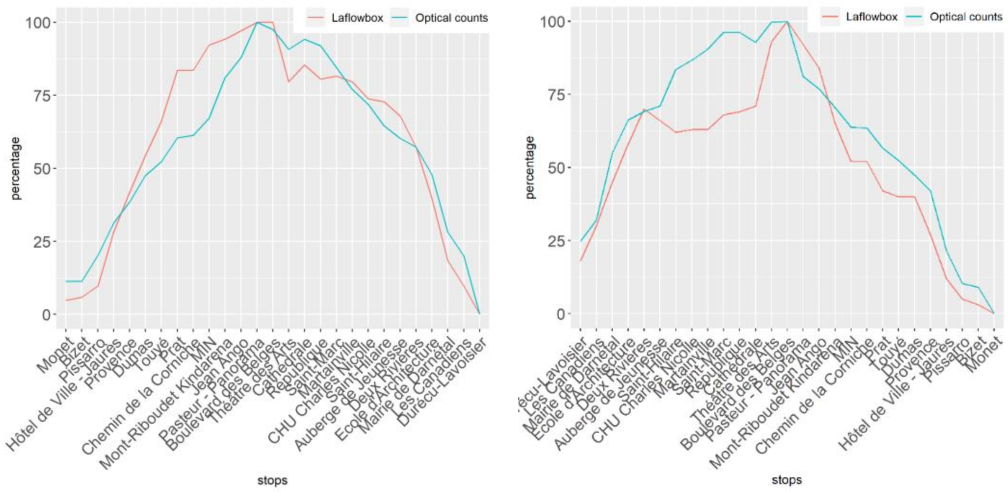

Figure 1: (a) Relative bus load – T3 direction 1 - 2018; (b) Relative bus load – T3 direction 2 - 2018.

However, at some stops, a noticeable difference appears. The observation of each O-D pair computed from Wi-Fi data and the onboard survey suggests that these differences mainly come from an excess of small trips originating or ending in the central zone.

Replicability of the method

In order to validate the replicability of the methodology to a larger scale, we analyze data from the second data collection campaign. Here again, we dispose of optical counts for comparison. We will use the O-D onboard survey implemented in 2018 to compare O-D pairs. R^2 values for the comparison between Wi-Fi data and optical counts are given in **Table 2**. And bus loads are represented in **Figure 2**. Here again, trips derived from Wi-Fi data collected in buses match quite well optical counts.

Table 2: R^2 values between Wi-Fi data and Optical counts 2019/2020.

	Direction 1			Direction 2		
	Boardings	Alightings	Bus load	Boardings	Alightings	Bus load
T1	0.73	0.59	0.95	0.73	0.77	0.97
T2	0.29	0.50	0.79	0.62	0.49	0.79
T3	0.68	0.63	0.91	0.81	0.55	0.90

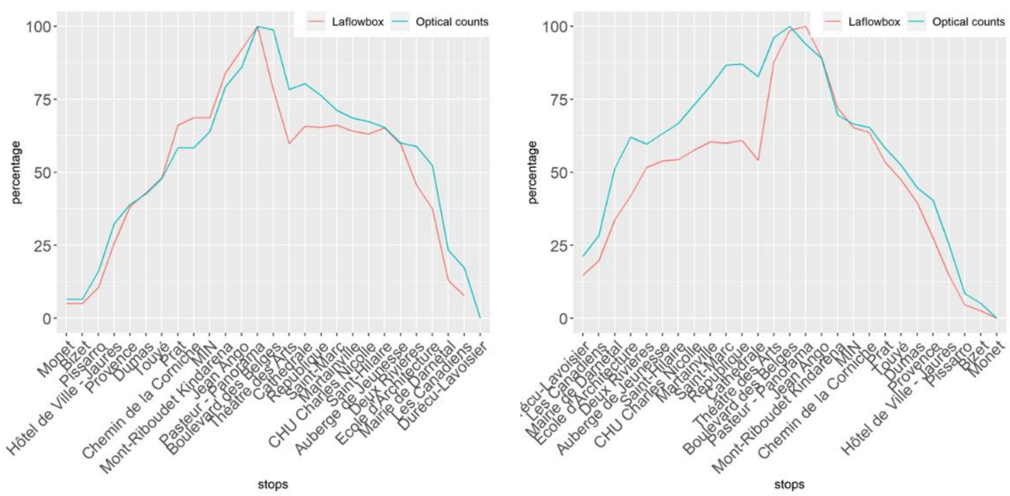

Figure 2: (a) Relative bus load – T3 direction 1 – 2019/2020; (b) Relative bus load – T3 direction 2 – 2019/2020.

Hence, we have shown that the methodology developed for a Wi-Fi dataset collected with a lot of sensors placed in vehicles circulating on a single line on a single day of 2018 can be applied to another dataset. The observation of each O-D pair computed from Wi-Fi data and the onboard survey shows the same trends as the one observed in 2018, i.e. the excess of small trips. Several hypotheses can be mentioned to explain these differences. They concern the Wi-Fi sensor (antenna coverage) but also the O-D survey (underestimation of small trips). Some adjustment to the method could be suggested too, like placing several sensors in the same bus, or improving the clustering algorithm.

4. CONCLUSIONS

In recent years, Wi-Fi sensors installed in city buses emerged as a new way of collecting mobility data to understand the travel behavior of inhabitants. As a passive method, it requires no effort from passengers and represents a marginal cost for the public authority. This study uses the K-means algorithm to automatically identify the Wi-Fi signals of connected objects belonging to bus passengers and build an Origin-Destination matrix. This automatic process avoids the commonly used step of evaluating threshold values for each new experiment and makes the method replicable to other bus lines and other agglomerations. Also, by comparing Wi-Fi data collected in buses with other real data sources, this study showed the reliability of Wi-Fi data to understand mobility behaviors. Transposing this comparison to a second dataset, collected on a larger perimeter and a longer period, the study also shows the replicability of the method. The quality of mobility data gathered continuously in public transport by Wi-Fi sensors is confirmed, as well as the relevance of using clustering algorithms to automatically get O-D matrices from these passive data. By making it free of temporal and spatial considerations, the methodology used in this study is an interesting answer for ensuring the sustainable development of transport infrastructures.

ACKNOWLEDGMENT

This research was conducted as part of a research agreement between Explain, the Urban Planning, Economics and Transport Laboratory (LAET) and the Actuarial and Financial Sciences Laboratory (LSAF). The authors acknowledge them for the financial support. The Métropole Rouen Normandie is also thanked for providing the data.

REFERENCES

Afshari, H.H., Jalali, S., Ghods, A.H., Raahemi, B., 2019. An Intelligent Traffic Management System Based on the Wi-Fi and Bluetooth Sensing and Data Clustering, in: Arai, K., Bhatia, R., Kapoor, S. (Eds.), Proceedings of the Future Technologies Conference (FTC) 2018, Advances in Intelligent Systems and Computing. Springer International Publishing, Cham, pp. 298–312. https://doi.org/10.1007/978-3-030-02686-8_24

Araghi, B.N., Christensen, L.T., Krishnan, R., Lahrmann, H., 2012. Application of Bluetooth Technology for Mode-Specific Travel Time Estimation on Arterial Roads, in: Proceedings from the Annual Transport Conference at Aalborg University. https://doi.org/10.5278/ojs.td.v1i1.5644

Bonnel, P., 2004. Prévoir la demande de transport.

Bonnel, P., Bayart, C., Smith, B., 2015. Workshop Synthesis: Comparing and Combining Survey Modes. Transp. Res. Procedia, Transport Survey Methods: Embracing Behavioural and Technological Changes Selected contributions from the 10th International Conference on Transport Survey Methods 16-21 November 2014, Leura, Australia 11, 108–117. https://doi.org/10.1016/j.trpro.2015.12.010

Bonnel, P., Munizaga, M.A., 2018. Transport survey methods-in the era of big data facing new and old challenges. *Transp. Res. Procedia* 32, 1–15.

Dunlap, M., Li, Z., Henrickson, K., Wang, Y., 2016. Estimation of origin and destination information from Bluetooth and Wi-Fi sensing for transit. *Transp. Res. Rec*. 2595, 11–17. https://doi.org/10.3141/2595-02

Ji, Y., Zhao, J., Zhang, Z., Du, Y., 2017. Estimating Bus Loads and OD Flows Using Location-Stamped Farebox and Wi-Fi Signal Data. *J. Adv. Transp*. 2017, 1–10. https://doi.org/10.1155/2017/6374858

Kurkcu, A., Ozbay, K., 2017. Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors: *Transp. Res. Rec.* https://doi.org/10.3141/2644-09

Kyritsis, D., 2017. The identification of road modality and occupancy patterns by Wi-Fi monitoring sensors as a way to support the "Smart Cities" concept: Application at the city centre of Dordrecht.

Myrvoll, T.A., Håkegård, J.E., Matsui, T., Septier, F., 2017. Counting public transport passenger using WiFi signatures of mobile devices, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6. https://doi.org/10.1109/ITSC.2017.8317687

Nitti, M., Pinna, F., Pintor, L., Pilloni, V., Barabino, B., 2020. iABACUS: A Wi-Fi-Based Automatic Bus Passenger Counting System. *Energies 13,* 1446. https://doi.org/10.3390/en13061446

Traunmueller, M., Johnson, N., Malik, A., Kontokosta, C.E., 2017. Digital Traces: Modeling Urban Mobility using Wifi Probe Data. Proc. 6th Int. Workshop Urban Comput. ACM KDD 2017 Halifax NS Can. 9.