Understanding taxi travel demand patterns through Floating Car Data

Nuzzolo, A., Comi, A., Papa, E. and Polimeni, A.

This is an author's accepted manuscript of a chapter published in Nathanail, E. and Karakikes, I.D. (eds.) Data analytics: Paving the way to sustainable urban mobility: Proceedings of 4th Conference on Sustainable Urban Mobility (CSUM2018), 24 - 25 May, Skiathos Island, Greece, Springer. pp. 445-452.

The final authenticated publication is available at Springer via:

https://dx.doi.org/10.1007/978-3-030-02305-8_54

Understanding taxi travel demand patterns through Floating Car Data

Agostino Nuzzolo[1[0000-0002-9685-4849]] Antonio Comi[1[0000-0001-6784-1638]], Enrica Papa[2[0000-0003-2922-2443]] and Antonio Polimeni[1[0000-0002-5104-7373]]

[1]University of Rome Tor Vergata, Department of Enterprise Engineering, Via del Politecnico 1, 00133 Rome, Italy
[2]University of Westminster, United Kingdom
{comi; nuzzolo; polimeni}@ing.uniroma2.it
e.papa@westminster.ac.uk

Abstract. This paper analyses the current structure of taxi service use in Rome, processing taxi Floating Car Data (FCD). The methodology used to pass from the original data to data useful for the demand analyses is described. Further, the patterns of within-day and day-to-day service demand are reported, considering the origin, the destination and other characteristics of the trips (e.g. travel time). The analyses reported in the paper can help the definition of space-temporal characteristics of future Shared Autonomous Electrical Vehicles (SAEVs) demand in mobility scenarios.

Keywords: Taxi Demand, Travel Demand, Floating Car Data.

1 Introduction

Improving air quality and reduce congestion in urban areas has been a priority in many countries. In central areas, regulatory traffic control has been introduced, usually integrated with improvement of public transport provision. These solutions are effective, but more than 20 years has passed from the introduction of those schemes and the technological context is changing. One main technological change refers to the rising of Automated Electric Vehicles (AEVs) that have the potential to affect substantially congestion, energy use and emissions in central areas. These AEVs will facilitate the sharing mobility development, with expected relevant effects on urban livability. One issue in designing future SAEVs service is the identification of the number of shared autonomous vehicles and the number and the localization of stations for the pick-up and drop-off of shared vehicles and the feasibility of the service [1]. In order to help the definition of space-temporal characteristics of the SAEVs demand in future mobility scenarios, the analyses of current characteristics of car sharing and taxi services can be of great help. For such purpose, this paper analyses the current taxi services in Rome, using FCD collected in February 2014, from a sample of 310 taxi drivers out of a total population of about 7,700 taxi drivers operating in the municipality.

The paper is composed as follows. Section 2 reports a literature review, section 3 presents the study area where the taxi service is analyzed. Section 4 reports the methodology adopted in the analysis, considering the data and the approach. Section 5 contains the main obtained results. Finally, section 6 reports some conclusions.

2 State of the art

Several studies have been carried out on the analysis of taxi services, thanks to the innovative technologies. One of them concerns Berlin’s taxi services [2] and it’s based on data collected from on-board GPS devices while operating taxis. In particular, the study analyses the travel behavior and vehicle supply of the Berlin taxi market using floating car data (FCD) for one week each in 2013 and 2014. Regarding spatial analysis, the study shows that most taxi trips take place either within the city center or from/to Tegel Airport, the most important Berlin’s airport. Another study [3] concerns Washington, DC where the authors investigate the relationship between taxi pick-ups and drop-offs from GPS data considering land use and travel data for each zone in the city. Also in this case, the presence of three civil airports in the study area influences the taxi demand. The purpose is the development of a model to relate taxi demand with land use and accessibility. Harbin City ([4]) is a further example for taxi data analysis, in particular the city was divided into traffic zones and the pick-up and drop-off locations are identified to build the origin-destination matrix of the trips. Other considered input data are travel distance, time and average speed in occupied and non-occupied status. The aim is to analyze travel demand distributions and the estimation of an entropy-maximizing model to estimate the traffic distribution. To reveal the travel patterns in Shangai, Liu et al. [5], analyze the trips of about 6,000 taxi. This study aims to identify spatial interactions among areas of the city and identify the structure of the travel flows. The data collected from taxi trips are complex, contains geographical and temporal components and, in some cases, could contains other trip information. Consequently, can be hard to use queries to perform analyses (e.g., trip distance, trip distribution). Ferreira et al. ([6]) propose a possible solution of this computational hard problem using the trips data of New York City taxi, developing a model able to support visual exploration of big data related to origin-destination taxi trips. A similar analysis is conducted using FCD collected in Beijing and containing the taxi trips [7]. The data processed deals with the taxi stay location and the taxi operations. In order to analyze the travel length, the distribution of trip distance and the spatial distribution of the taxi, [8] use a dataset containing the trajectory data of 11,880 taxis in Beijing. The analysis is conducted considering both the travels with customer and the empty travels. The variables analyzed in [9] are the displacement of each trip the duration of each trip, the time interval between successive trips by the same taxi. The data are from five different cities and consider only the trips with customer.

3 The study area

The municipality of Rome extends over an area of 1,283.70 Km^2, with the 22.20% destined to urban activities. The area is divided in six zones according to the PGTU (General urban traffic plan). Four of the six zones are inside the main road ring (GRA). The fifth zone is outside the GRA and includes urban perimeters of some relevance. The last zone is located in the west part of the city.
The population reaches 2.9 million of inhabitants, which daily generate 4.7 million of trips, while trips generated from outside Rome are about 800,000. Trips from outside is continuously growing (250,000 units more than 2004), confirming the population's tendency to go live outside Rome.

4 Methodology of work

4.1 Supporting data

GPS data used in this case study belong to CRAWDAD dataset 2014 [10]. The GPS position of each taxi is logged every 10 seconds and it has been possible to build a database of historical GPS traces through around 27 thousand GPS positions recorded per day, therefore 756 thousand for the entire month of February 2014. Each entry includes:

- ID (the taxi identifier),
- Date (date the record is logged),
- Timestamp (time the record is logged),
- Coordinates (geographical location: latitude and longitude).

4.2 Approach

The possible status of taxicabs was aggregated in the following categories:

1. to customer (driving to or waiting for a customer),
2. with customer (driving with a customer on board),
3. at rank (standing at a taxi rank),
4. outside rank (idle but not at a rank, for example returning to a rank).

The analysis of taxi demand from GPS traces, is composed of several phases:

1. computation of distance between two successive GPS positions recorded (if the distance travelled by the taxi driver in 2 minutes is less than 10 meters, it has been considered the vehicle non-moving, otherwise the driver is traveling),
2. individuation of the origin *o* and the destination *d* of each trip,
3. computation of progressive distance (to evaluate the length of a trip).

Other calculated values are the traveling time (for each trip) and the waiting time of a taxi in a position (if the waiting time in a position is greater than 2 minutes we suppose

that the position is a destination).
The activity *with customer* is the taxicabs status needed for this demand study. In order to individuate this kind of status, it's important to recognize all the times the vehicle is at a rank. A taxi cab is considered steady at a rank when the time previously calculated is more than 2 minutes and the coordinates are quite near one of 100 rank of the city. Hence, it's possible to establish the number of times a vehicle is steady at a rank and instead when it's steady because the driver is dealing with a customer (the costumer is entering or is leaving). Consequently, it can be found the trips with the activity *with customer*.
Once the data processing is applied, it has been possible to build a database based on the vehicles' status *with customer*, for the studied period. Each entry of the database contains:

- length of the trip,
- travel time,
- origin and destination of the trip,
- waiting time.

Finally, the origin-destination matrix based on the vehicles' status *with customer* is built.

5 Main results of the analyses

5.1 Temporal patterns of taxi demand

A detailed distribution of taxi demand (the number of requests submitted per day and the number of requests submitted per hour) over the two weeks are presented in Fig. 1 and Fig. 2. For the demand side, the following happens:

- taxi demand during weekdays (Monday – Friday) is generally higher than that of weekend (Saturday and Sunday),
- taxi demand during weekdays follows a clear pattern: a major peak from 8 am to 9 am, smaller peaks during the afternoon (the first peak at 3 pm and the second one at 6 pm) and very low taxi demand from 12 am throughout the night,
- Monday is demand-wise the busiest day,
- Saturday night records a peak around 11 pm and significant taxi demand throughout the night.

Besides Saturday night, the demand for taxi services during weekends is low.

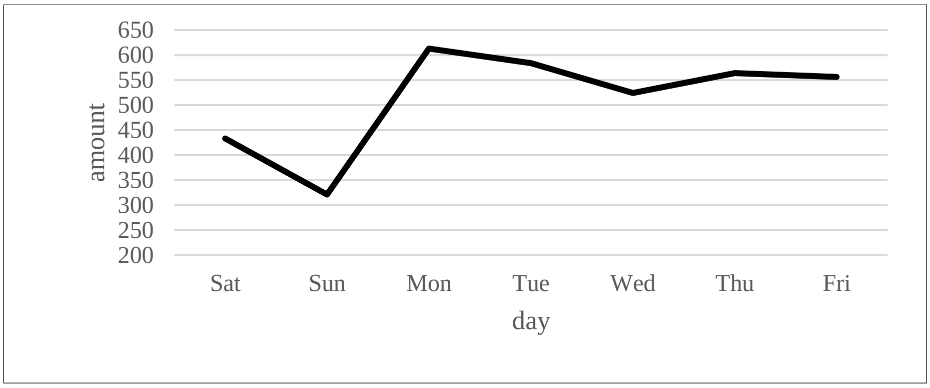


Fig. 1. Request submissions per day

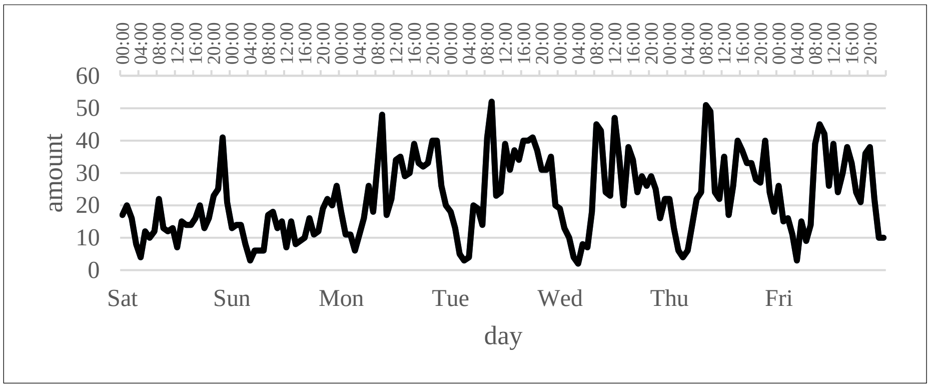


Fig. 2. Request submission per hour

5.2 Trip distances and travel time

The average trip distances, calculated adding the beeline distances between two consecutive geographical points (the points are recorded every 10 seconds), are at most weekday hours between five and fifteen kilometers (Fig. 3). However, the weekday morning (about 8 am), with average distances around twenty-five kilometers, form a notable exception. The day with the highest average trip distance is, on the other hand, Wednesday, with roughly 30 kilometers. Relating to the average travel time (Fig. 4), it range from 7 to 31 minutes. The maximum average travel time is recorded Monday, between 8 am and 12 am, another peak is notable on Friday around 8 pm. The time analysis can be useful in defining the use of SAEV, assuming that the time that a user spend traveling with a taxi is comparable with the time that he/she would spend using the SAEV.

As to the distribution of trip distances (Fig. 5), there are few trips of less than one kilometer, whereas a trip distance between two and five kilometers is the most common. Longer distances are less and less likely: not even twenty percent of all trips are long between ten and twenty kilometers. On the other hand, trips longer than twenty kilometers are about the 13%.

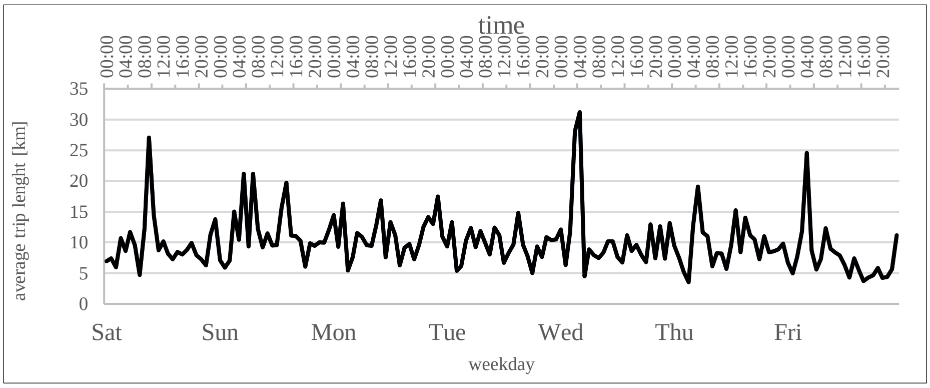


Fig. 3. Average trip distance at different times

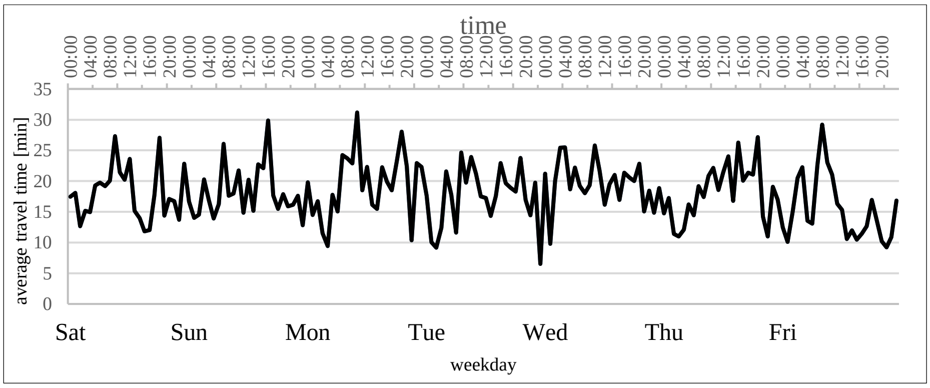


Fig. 4. Average travel time

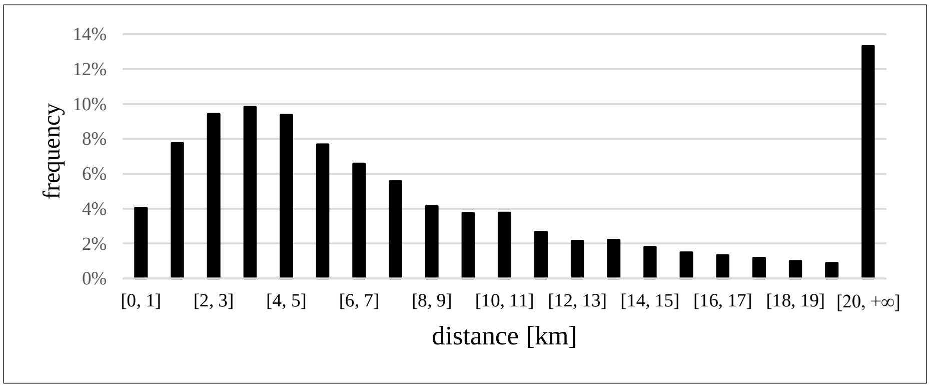
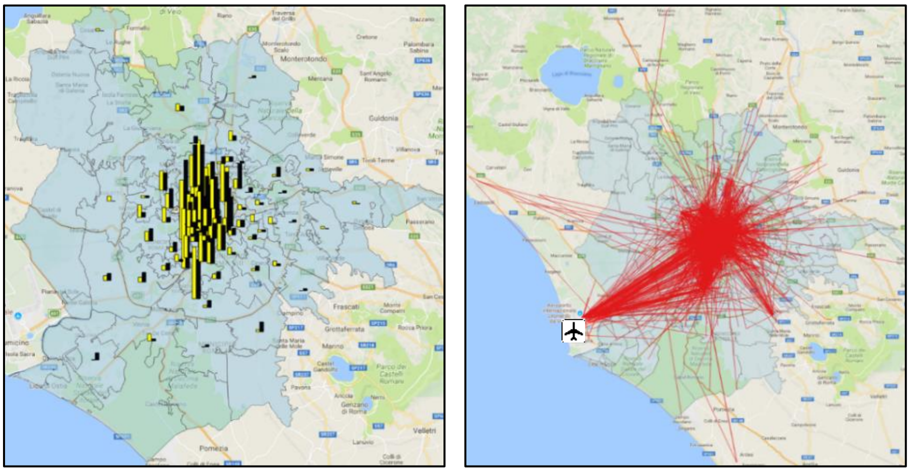


Fig. 5. Trip distance distribution

5.3 Location-based taxi demand

Origins and destination of taxi trips are generally spread all over the city (Fig. 6a). The majority commences and ends within the inner railway ring and in the EUR district, one of the most important financial and tertiary area of the city. Moreover, the data suggest that about the 15% of trips are from the city to the outside and vice versa. Fig. 6b highlights these trips, remarking that they are mainly from/to Fiumicino Airport, the biggest Rome Airport.



a) Origins and destinations per zone b) Trip visualization

Fig. 6. Trips and origin destination relations (Map source: Google Streets)

6 Conclusions

In this paper, we presented an analysis of FCD related to the taxi trips in Rome. The data were collected in February 2014, from a sample of 310 taxi (on a total of about 7,700 taxi operating in the municipality). The data processing suggests that there is generally a demand peak on workday mornings and a several lower peaks over a longer time in the afternoon. From analyzed data, about 30% of the trips are taken in the morning from 5am to 8am while the remaining 70% are distributed over the rest of the day with two afternoon peaks, the first one at 4pm and the second one at 6pm. On weekends the demand peaks shift towards the night and are generally lower than weekdays. On Saturday night, about 40% of the total number of trips occurs, with a peak at 11pm. The remaining trips (60%) are distributed equally over the course of the day.

As far as spatial analysis is concerned, the majority taxi trips begins and ends within the Inner Railway Circle, the trips from/to the outside (mainly from/to Fiumicino airport) are about the 15%. As to the distribution of trip distances, there are few trips of less than 1 km, whereas a trip distance between 2 and 5 km is the most common. Longer distances are less and less likely: not even 20% of all trips are long between 10 and 20 km. On the other hand, trips longer than 20 km are about the 13%.

This analysis of Rome can be useful for the future SAEV services design, also using the taxi data (space-temporal distribution, number of trips, travel time, distance) as a proxy to try to define one component of the future demand for SAEV. Indeed, a first element to design the sharing service is the definition of the areas where the vehicles can be picked-up or dropped-off. To do this, can be selected as potential areas those where are registered high values of demand for taxi. Besides, the travel time evaluated with taxi data can be used to size the fleet of shared vehicles. It is noted that the actual zones of Rome covered with the vehicle sharing services are those with high taxi demand (central zones).
Further analyses are also in progress to use these data, with the aim to develop and test a model framework for forecasting taxi travel demand in relation to socio-demographic characteristic of different zones of the city.

Acknowledgments

The authors want to thank Luis Moreira-Matias for the help in data retrieval and Claudia Proietti for the support in data elaboration.

References

1. Nuzzolo A., Persia L., Comi A., Polimeni A., Shared Autonomous Electrical Vehicles and Urban Mobility: a vision for Rome in 2035. 4th Conference On Sustainable Urban Mobility, 2018.
2. Bischoff J., Maciejewski M., Sohr A.: Analysis of Berlin's taxi services by exploring GPS traces. In: Proc. Int'l Conf. Models and Technologies for Intelligent Transportation System (2015).
3. Yang, Z., Franz, M.L., Zhu, S., Mahmoudi, J., Nasri, A., Zhang, L.: Analysis of Washington, DC taxi demand using GPS and land-use data. Journal of Transport Geography, 66, 35-44 (2018).
4. Tang J., Liu F., Wang Y., Wang H.: Uncovering urban human mobility from large scale taxi GPS data. Physica A: Statistical Mechanics and its Applications, 438, 140-153 (2015).
5. Liu X., Gong L., Gong Y., Liu Y.: Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43, 78-90 (2015).
6. Ferreira N., Poco J., Vo H. T., Freire J., Silva C. T.: Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips. IEEE Trans. Vis. Comput. Graph., 19(12), 2149–2158 (2013).
7. Jianqin Z., Peiyuan Q., Yingchao D., Mingyi D., Feng L.: A space-time visualization analysis method for taxi operation in Beijing. Journal of Visual Languages & Computing, 31(A), 1-8 (2015).
8. Cai H., Zhan X., Zhu J., Jia X., Chiu A. S.F., Xu M.: Understanding taxi travel patterns. Physica A: Statistical Mechanics and its Applications, 457, 590-597 (2016).
9. Wang W., Pan L., Yuan N., Zhang S., Liu D.: A comparative analysis of intra-city human mobility by taxi. Physica A: Statistical Mechanics and its Applications, 420, 134-147(2015).
10. Bracciale L., Bonola M., Loreti P., Bianchi G., Amici R., Rabuffi A.: CRAWDAD dataset roma/taxi, https://crawdad.org/roma/taxi/20140717, last accessed 2018/02/02.