INFORMED VERSUS NON-INFORMED TAXI DRIVERS: AGENT-BASED SIMULATION FRAMEWORK FOR ASSESSING THEIR PERFORMANCE

Josep Maria Salanova Grau
Centre for Research and Technology Hellas - Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., 57001, Thermi, Thessaloniki, Greece
Tel: +302310498433 Fax: +302310498268; e-mail: jose@certh.gr

Luis Moreira-Matias
NEC Laboratories Europe
Kurfürsten-Anlage 36, D-69115 Heidelberg, Germany
Tel: +49 6221 4342 261 Fax: +49 6221 4342 155; e-mail: luis.moreira.matias@gmail.com

Amal Saadallah
Informatik LS VIII & The collaborative Research Center SFB 876, TU Dortmund
D-44227, Dortmund, Germany
Tel: +49 (0)231 755-6490 Fax: +49 (0)231 755-5105; e-mail: amal.saadallah@cs.tu-dortmund.de

Panagiotis Tzenos
Centre for Research and Technology Hellas - Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., 57001, Thermi, Thessaloniki, Greece
Tel: +302310498436 Fax: +302310498269; e-mail: ptzenos@certh.gr

Georgia Aifadopoulou
Centre for Research and Technology Hellas - Hellenic Institute of Transport
6th km Charilaou-Thermi Rd., 57001, Thermi, Thessaloniki, Greece
Tel: +302310498457 Fax: +302310498268; e-mail: gea@certh.gr

Emmanouil Chaniotakis
Technical University of Munich, Chair of Transportation Systems Engineering
Arcisstraße 21, Munich, Germany
Tel: +49 89 28910461; e-mail: m.chaniotakis@tum.de

Miquel Angel Estrada Romeu
Center for Innovation in Transport (CENIT), Barcelona Tech - Technical University of Catalonia (UPC), Barcelona, Spain
Tel: +34934137667 Fax: +34934137675 e-mail: miquel.estrada@upc.edu

Word count: 5,644 words text + 7 tables/figures x 250 words (each) = 7,394 words

Submission date: August 1st, 2017

ABSTRACT

Data driven research is becoming a standard in Transport. Recent advances in the Artificial Intelligence and Machine Learning related areas enable the possibility of automatically generating highly-accurate predictive analytics frameworks, under any context. Such frameworks can potentially provide unprecedented levels of information to all mobility actors regarding not only the current but also the future status of network variables – such as Origin-Destination flows. This fact elevates the decision support to a new standard, where operations can be optimized in real-time and in near-autonomous fashion. However, such advances also bring new questions: How much can a transport operator benefit from this? Is there a limit for the amount of information that all actors should have?

This paper aims to answer such questions by introducing an agent based model able to simulate the behavior of individual taxi drivers on their passenger-finding strategies. Multiple strategies are proposed and compared through exhaustive computer-aided simulations. The goal is to find how different drivers will benefit from the availability of accurate information about the future spatiotemporal demand distribution. The experiments were conducted using real-world operational data collected from a large scale taxi fleet operating in Thessaloniki, Greece. The obtained results illustrate different perspectives of the cost-benefit tradeoff on disseminating future demand-related information at different scales and ratios.

1. INTRODUCTION

The fast modernization communication and geolocation technologies created novel massive mobility data sources which have been massively deployed across transportation industry since the early 2000s: from mass transit operators to individual taxi drivers [1-2]. These technologies allows to monitor the operations of an entire fleet in real-time. Moreover, Machine Learning techniques can be built upon this massive data sources to provide insights about the future behaviors of the demand and the supply – especially in the taxi industry, where drivers can freely decide on different passenger-finding strategies [3].

Taxis are present in most cities around the world, accounting for a small but rather significant part of the daily trips of their citizens and consequently, the produced traffic. However, the abovementioned technological advances are changing this market drastically due to emergence of easier and cheaper ways to collect information. In the past, most taxis were owned and driven by professional drivers (with a drivers' license for operating a taxi), while, a good knowledge of the city streets was required and demonstrated -in many cases- through \ exams supervised by credited experts. Based on their experience, drivers had empirical information about where to find rides at each time of the day. The introduction of new technologies (navigators) and third party services (like Uber) allows *anyone* can become a taxi driver, independently of his level of expertise on this task. These technologies also disrupted traditional operations on on-demand transport networks, where old fashion dispatching centers (mostly call center for booking a ride) are slowly giving place to customer to customer (C2C) applications running on any smart phone.

The new scenario is ruled by the parties having access to the customers, which mostly prefer C2C direct deals through innovative web and mobile service instead of call centers. There are two main sets of internet-enabled software applications of this type which have been recently developed by third parties: (i) ridesharing and (ii) car hailing [4-5]. The first consists on offering social networks that allows users to share and join individual trips of each other (e.g. BlaBlaCar). The latter explores the concept of on-demand transport networks in a fully de-regularized fashion, where the hailed cars can either be professional drivers (e.g. taxi operators) or private vehicles (e.g. DiDi or Uber). In both cases, Artificial Intelligence (AI) and Machine Learning (ML) techniques are key for their successful penetration in the on-demand transport market by providing optimized relationships between supply and demand in a predictive fashion. Nowadays, the importance of getting accurate demand-related information to strive in this business sector is beyond question - independently of the actors providing it (e.g. dispatching centers, C2C software applications or AI-based agents running on those). Naturally, the explosive availability of these multiple types of information – especially predictive one – generates questions about its usage. After all, the question that all operators would like to answer is "How can we maximize the profit of my fleet/associates?"

On taxi modelling and particularly on the representation of their operations, a possible categorization can be made based on the three main operational statuses: (i) stand, (ii) hailing and (iii) dispatching. Each category may follow a different regulation, which can vary from market to market. In the stand mode taxis and users meet at predetermined meeting points, with taxis waiting for costumers, in the hailing mode, taxis circulate searching for a user, and users are waiting for the first empty taxi while in the dispatching mode, taxis services are provided by dispatching centers, which are responsible for matching available taxis with the demand for taxi services [6]. The first models of the taxi sector (1970-1990) were highly aggregated and focused in the profitability of the sector, mainly dealing with regulatory aspects [7, 8]. Their principal intention was to evaluate the necessity for regulation. Later studies focused on the operational characteristics of the taxi market, taking into account the spatial distribution of the demand and the supply for taxi services [9, 10]. During the last years various simulation models were developed but few were used for estimating the impact of having information about the demand in the performance of the taxi sector [11-13], especially dealing with how drivers having different information levels will perform. More recently, various studies in predicting taxi demand and rides have been released, but mostly focusing on the accuracy of the prediction itself and not on how and when this should be provided to the drivers.

This paper aims at analyzing the impact of the provision of demand-related information to taxi drivers focusing on how the information is shared -and not in the accuracy of the information itself. It analyzes how the performance of the drivers can change based on the information that they have available, the number of drivers having access to the same type of information, as well as the possible cannibalism effects that such availability can generate within [3].

The paper is structured as follows: A literature review is presented in section two, reviewing the taxi modeling and the how the provision of information to drivers affects their performance. The agent based model is described in section three, while the results are presented in section four. Finally, conclusions and future research guidelines are presented in the last section.

2. LITERATURE REVIEW

Many research works approached taxi-industry related problems under different perspectives. They contributed successfully to distinct applications designed for planning or organizing taxi services such as predicting taxi-demand [6,14-16], uncovering taxi-driver mobility intelligence [17-19], building intelligent taxi/passenger finding strategies [20-22] and designing recommender systems [3, 23-25]. A more extensive review on taxi modeling in general can be found in Salanova et al. [26].

The demand-forecasting problem tries to mitigate the demand/supply imbalance problem by providing accurate predictions of the spatiotemporal distribution of the service demand. [6] suggested a seasonal ARIMA model to handle the problem for short-term prediction. Recent works in [14-15] suggested time-sliding window ensemble schemas to provide real-time taxi demand forecasts. [16] used clustering to predict demand distributions with respect to contexts of time, weather, and taxi location. These works chose to put the light on the prediction task without assessing the impact of other factors contributing to the demand/supply imbalance such as drivers' mobility intelligence level and the dynamics of the operating traffic network. They focused on showing the predictive ability of their methods without assessing the impact of the provided prediction information on improving the efficiency or the profitability of taxi services. To the best of our knowledge, this is the first work to assess the potential impact of providing different levels of demand-related information using data-driven AI based technologies on different supply behaviors.

Other works put the light on taxi-drivers mobility intelligence. Liu et al [17] classify taxi drivers according to their income into two categories: top drivers and ordinary drivers. They observed that top drivers have the special proportion of operation zones, with an optimal balance between taxi demand and fluid traffic conditions, while ordinary drivers choose most of the times to operate in fixed spots with few demand variations. Ziebart et al [18] present a decision-modeling framework for probabilistic reasoning from observed context-sensitive actions. Based on 25 taxi drivers, the model is able to make decisions regarding intersections, route, and destination prediction given partially traveled route. Yuan et al [19] propose the T-Drive system that based on an historical GPS dataset of taxi-drivers trips, is able to compute the fastest path for a given destination and departure time. Most of these works chose to focus on studying the supply behavior independently from the existing demand information.

Recommender systems [3, 23-25] are more generic frameworks for organizing taxi operations. These systems aim to aid taxi drivers in choosing operational locations that maximize their utility function, which can be expressed in term of one or many variables such as time, profit and fuel consumption. In [23], the authors model the road network topology as a graph where the optimal route is given with respect to its profit by solving an optimization problem either approximately or in a closed form. [24] took it further to provide exact routes for a parking location on a subarea which would maximize the drivers short-term profit. In [25], a location-to-location graph model was adopted to capture the relation between a passenger drop-off location and the next passenger get-on location. In addition, the authors estimated the expected fare for a trip that starts at a given recommended location. A simulator that simulates the cruising behavior of taxies in the dataset and a virtual taxi that cruises based on their recommender system was developed to assess the impact of their system on taxis' profitability. Their simulation results indicate that their recommender system is still effective in terms of recommending more profitable cruising locations, although the statistics of the historical data that have been used could be different from real-time passenger requests. Moreira-Matias et al. [3] proposed an online taxi stand recommendation for taxi drivers. A fleet equipped with their framework can see its average waiting time reduced by 5% compared to its competitor. Their method takes advantage of the ubiquitous characteristics of the operational network and it assembles experience and knowledge of all the drivers with the real-time short-term prediction of the demand.

Nevertheless the relevant amount of related work on providing demand-related insights for taxi drivers, at the best of our knowledge, there is no relevant work in the related literature on quantifying the impact of shared demand related information on drivers adopting distinct driving strategies. Consequently, we believe that this study is a relevant contribution to encourage taxi fleets to adopt data-driven demand inference technologies - it encloses both a methodology and a potential tool for quantifying the overall benefit of a taxi-fleet from having access the demand related information at different extents.

3. CASE STUDY: THESSALONIKI

Thessaloniki is the second largest city in Greece and the largest of Northern Greece, with a total of more than 1 million citizens in its greater area, which covers a total of 1,500 km2. It has 665 inhabitants per km2 on average and more than 777,000 vehicles, including private cars, heavy vehicles and motorcycles [27]. Thessaloniki has roughly 2,000 taxis, 50% of which are under the same association which host the largest dispatching call center in the Balkan area, with 1,500-2,500 calls per day only in Thessaloniki, while the fleet executes a total of 7,500-15,000 trips per day. In order to assign customers to available vehicles, the fleet is being monitored continuously, generating a huge dataset with coordinates and status of the vehicles with a frequency of 6-10 seconds which allows following the trajectories of the trips with and without customer in detail.

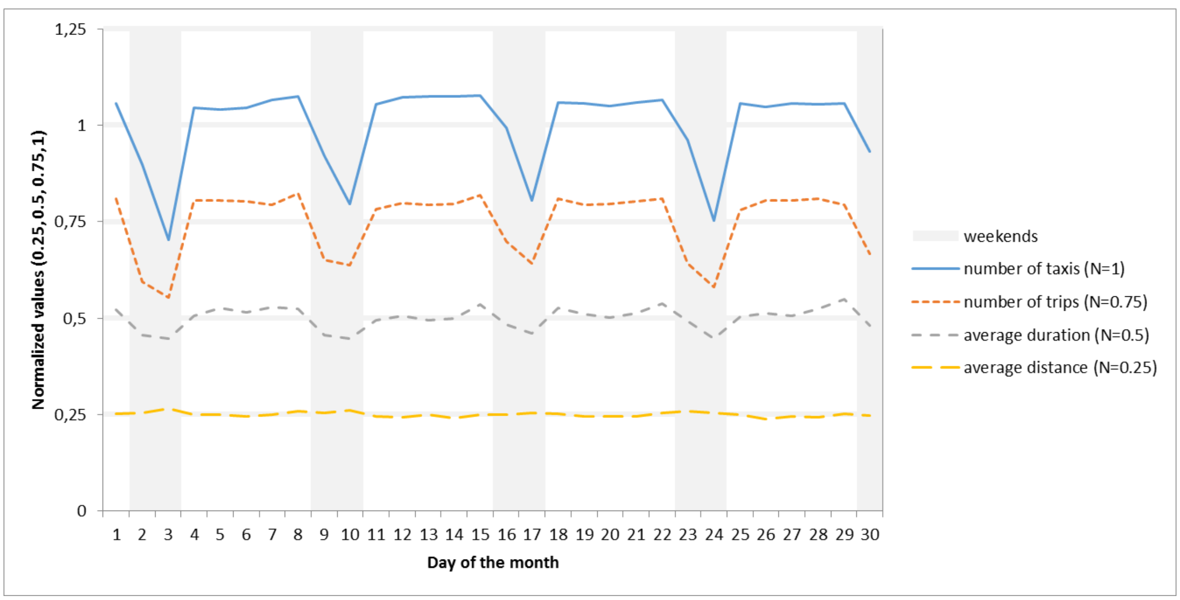

FIGURE 1 Number, total and average duration and distance of the taxi trips in Thessaloniki for one month normalized (N) to 1, 0.75, 0.5 and 0.25 respectively

The characteristics in terms of number, distance and duration of the taxi trips (Figure 1) present a low variance within the week, except between weekdays and weekends. Based on this low variability this research was conducted using operational data from a single day, representative of the weekdays. This dataset has a total of 8,675 passenger trips and ~5k empty trips provided by a fleet of 730 vehicles within the 80 zones defined by the taxi operator. The rides had a total duration of ~1.4k hours and covered a distance of ~28k kilometers. An analysis of the places where pick up and drop off of customers happen is provided in [28].

Thessaloniki is a longitudinal city, with sparse road network and a few high-hierarchy streets crossing the whole city, which difficult the provision of Public Transport services by means of mass transit (there is no metro in the city, only buses), so many citizens use taxis for satisfying their mobility needs. It can be observed a high concentration of locations where mostly pick up of customers happen, while the drop-off locations are sparser.

The accumulated demand and the number of taxis in the network for each time of the day are presented in figure 2 below.

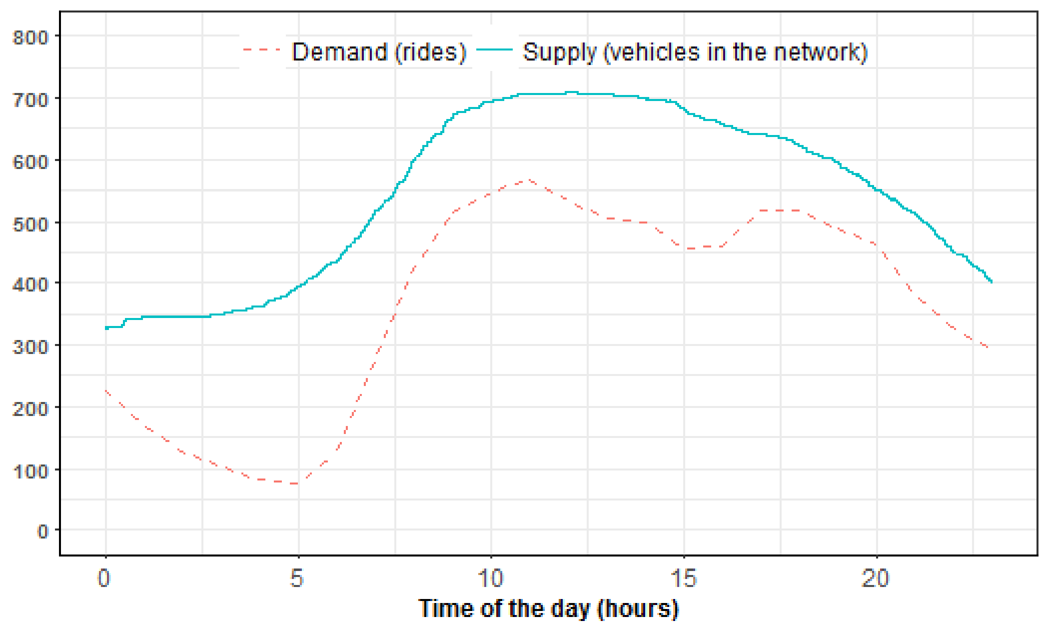

FIGURE 2 Demand and supply for taxi services in Thessaloniki.

The average number of rides per hour during the day is 361, but this is not homogenously distributed, presenting peaks (more than 1,000 rides per hour) at 13:00 and 20:00 and off peaks between 2:00 and 6:00 (35-75 rides per hour). The average number of vehicles in the network is always larger than 300, with its lowest level at night and its highest between 10:30 and 15:00 where 700 vehicles are active.

Figure 3 depicts the vehicles' activities. Each cab can be classified as empty (with a peak of 300-350 vehicles between 9:00 and 19:00), waiting at taxi stands (150 and 170 between 7:00 and 21:00) and occupied - circulating with a customer (100 and 150 between 9:00 and 20:00). There are two small periods were data is not available, early in the morning (around 0:30) and in the afternoon (around 18:30).

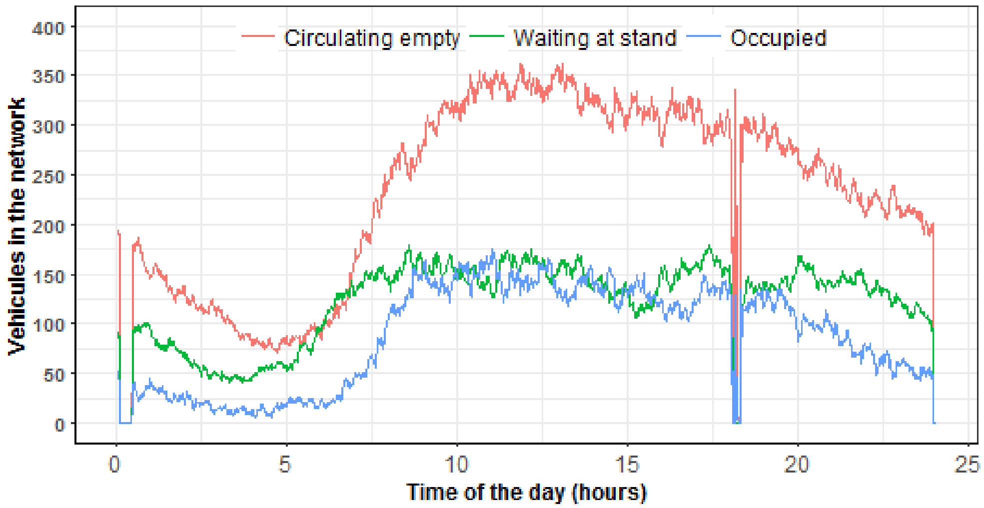

FIGURE 3 Distribution of vehicles (real data).

4. THE PROPOSED AGENT-BASED MODEL

A set of vehicles is provided with artificial intelligence rules for deciding which stand of a set of stands to visit looking for a ride. A set of customers appears at the stands following a distribution based on empirical data and willing to satisfy the need for traveling to another zone by picking up the first available taxi in the stand queue (if any) or waiting for a taxi to arrive to the zone forming a customer queue. The behavior of the vehicles in vacant mode is modelled following a data-driven fashion based on real data and information about demand. The trips are not modelled as such using a micro-simulation engine but exogenously modelled from empirical distributions obtained from real data.

The taxi drivers are assumed to constantly evaluate and follow strategies that maximize their performance. They are introduced in the simulation, using behavioral rules that apply when selecting the next action after delivering a customer, which can be a) stay at the same location b) go to another stand depending on their expectations (based on their experience, not in information available) to find a ride c) to cruise to nearby locations based on information about demand. The performance metric for the drivers is the number of rides per driver and indirectly the waiting time of the customers. Concerning information provision, four basic types are distinguished for the taxi drivers, 2 non- or partially-informed (conservative and empirical/experienced) and two informed (real time and prediction based information). Specifically:

- **Conservative**: drivers stay at the destination zone of the last customer, join the queue at the taxi stand and wait, based on a (First In First Out) FIFO policy .
- **Empiric/experienced**: drivers select a zone based on their experience, travel there after dropping off the last customer, join the queue at the visited taxi stand and wait for their turn.
- **Informed (real-time)**: drivers go to the most attractive zone, defining attractiveness of a zone as the ratio of the customers queue at that moment and the distance to that zone (taxi queueing information is not used since there will be no taxis waiting if there are customers in the queue).
- **Informed (prediction)**: in this case the drivers go to the most attractive zone, but instead of using the customers queue at that moment they use the prediction for taxi trips at the moment they will reach each zone.

Taxi customers are introduced in the simulation using a stochastic process modelled on real customer data for each respective simulation period. They do not have any other behavioral characteristic except for the destination (modelled deterministically based on the actual observed data). Since the trip duration is fixed the only performance metric for the customers is the waiting time, but they cannot affect it since it only depends on the decision of the drivers.

All trips are modelled to start from the centroid of each zone, with drivers and customers to be forming queues at taxi stands. Additionally, there is only one taxi stand per zone located at the centroid, where all taxis wait for a ride (there is neither hailing nor dispatching). This means that zones and stands are the same. In case of a ride starting and ending at the same zone the ride takes place by making the driver will leave the queue, be busy during the duration of the trip and return to the queue in the last place (in case the driver decides to stay at the stand). Finally, the start time of the recorded trips is considered as the time where the customer started waiting for a ride since is the time when he/she called the taxi dispatching center.

4.1. The simulation engine

The simulation accounts for the within day variations aiming at representing the impact of different information provision levels for different demand levels and different activities. On the driver side, the selected simulation period accounts also for the shifts of the taxi drivers as well as the information about the real trips executed that day (with and without customer). Given the approach adopted, the simulation takes place on a second by second basis (Figure 4). Customer and driver arrivals are introduced at each time step while obtained from shifts and rides data. Each actor introduced joins the respective queues of the zone in which they appear at the time they appear. Following, the queueing module matches customers and drivers at each zone, giving rides to drivers based on the FIFO rules.

The behavior module is comprised by operations for the estimation of the attractiveness of the zones performed on the number of empty trips to that zone in the same time interval while assigning taxi drivers to pertinent behavioral types given penetration rates. In every step the drivers are checking the queues table and choose either to stay at the queue or to move to another zone based on the attractiveness of the zones calculated as the ratio between the number of customers in the queue and the distance to each zone.

Finally, at each step of the simulation, the variables are updated, increasing by 1 the waiting time of customers and drivers at queues, reducing by 1 the countdown for drivers with customers and increasing by one the busy travel time of drivers with customers as well as the empty travel time of drivers circulating without customer. The shifts are checked and the drivers that end their shift leave the simulation (if they are traveling, with or without customer the end their travel and then leave the simulation).

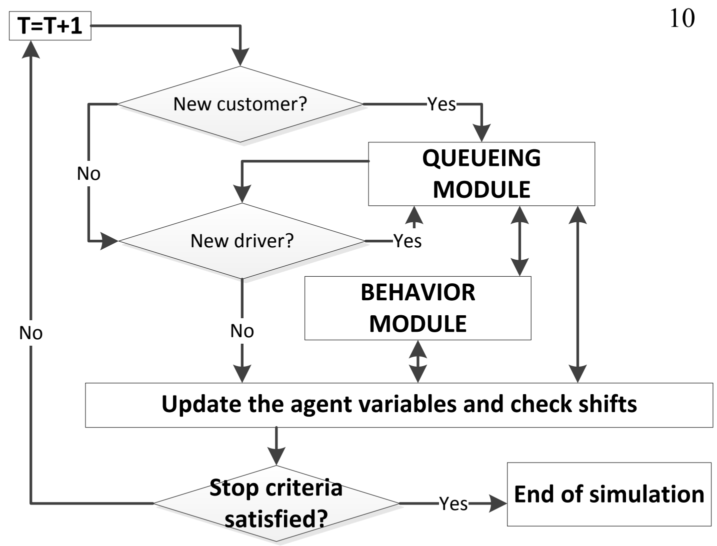

FIGURE 4 Simulation workflow.

The customer arrivals follow the distribution of the demand as recorded in the rides table, both in time and space while driver arrivals are based on the shifts table, also having space and time information. Taxis remain in the simulation until their shift ends, while customers remain until they arrive at their destination.

4.2. Case Settings

The zones have been defined by the taxi fleet operator using the historical database of rides. The trips collected are all provided by the same dispatching center, which manages 50% of the taxis in Thessaloniki. The way this is done is by having virtual queues in the zones and giving the calls to the first vehicle in the virtual line. The condition to be in the virtual queue of one zone is to be queuing in a physical rank inside the same zone.

The trip duration is obtained from real data, both for empty trips or trips with customer (there is no routing engine or road network defined). The fleet operator records the trajectory of each vehicles, so the distance and duration of each ride can be easily extracted and is used here aiming at making the simulation more realistic. The use of routing engines present a good fit with the distance of the trip, but they tend to underestimate travel time since the traffic status varies during the day [28].

The trips database consists of all trips executed by the taxi fleet, including empty trips when going from one zone to another. This database is split into three databases:

- Trips with customer: All trips with customer are recorded, containing customer id, start and end zone and timestamp as well as duration and length. These are used for generating the demand for taxi rides.
- Empty trips: All empty trips are recorded, containing driver id, start and end zone and timestamp as well as duration and length. These are used for defining the behavior of the experienced drivers when selecting the next zone/stand to wait for a customer.
- Drivers shift: The first and the last trip of the day are used for defining the shift of each driver, which is defined by driver id, start and end zone and timestamp.

5. EXPERIMENTS

This section describes in detail the computer-aided experiments carried out. Firstly, we describe the Scenarios that we evaluate and the metrics that we used to do it so. Then, a brief comparative statistics between the empirical spatiotemporal distribution of the taxi drivers vs. the one provided in the simulation is provided. Finally, we present and briefly discuss the obtained simulation results.

5.1. Scenarios and metrics

A total of 16 scenarios with different distribution of the drivers behavior were executed, four basic scenarios were all the drivers have the same behavior and twelve combined scenarios were the number of drivers following each one of the behavior varies. The scenarios were selected aiming at having a good representation of the possible combinations and therefore capturing correctly the trends and being able to plot them as a continuous variable.

We used to evaluation metrics: the non-satisfied demand and the rides per vehicle. The non-satisfied demand is important in scenarios where the non-informed vehicles dominate.

5.2. Validation

In order to validate the results of the model, the real distribution of taxis in the empirical scenario is presented in Figure 5 is compared to the same figures obtained from the simulation.

In general terms the distributions fit well, but the following issues can be observed:

- The empty vehicles from the simulation are always more than in the real dataset, especially at night. This is due to the fact that during the day some vehicles stop working for a few hours, especially during late-night and early-morning hours, but this cannot be easily obtained from the dataset, especially if they stop only for a few hours.

- The number of occupied vehicles in the simulation is always slightly lower than the real one, which is due to the false busy status declared by some taxis (by mistake) while waiting at the stands, which is corrected when the trips are extracted (very short trips both in time or distance are not considered) but it remains in the status of the vehicles.
- Finally, there is a lack of data between 18:00 and 18:45, which in the case of the real data is observed as a reduction in the number of vehicles of all classes while in the simulation is reflected as zero occupied taxis and a respective increase in the number of empty vehicles, which validates also the simulation results.

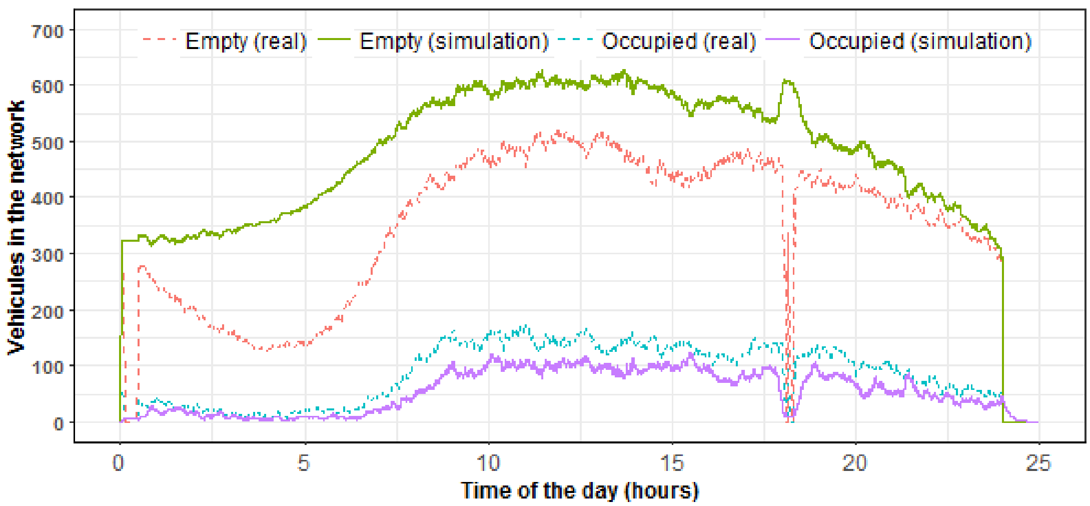

FIGURE 5 Distribution of vehicles (comparison between real and simulated data).

5.3. Results

The simulation results of the basic scenarios is a 60-64% of customers served and 7.1-7.5 rides per driver in the Conservative and Empirical scenarios respectively while in the informed scenarios all Customers are served, achieving 11.8 rides per driver. the number of rides of the non-informed scenarios is 35-40% lower with respect to the informed scenarios. Within the non-informed scenarios, the empirical performs slightly better but not as expected. It is important to note that in the empirical scenario the number of empty trips is only 20% of the empty trips contained in the real database, which means that in most cases the drivers stayed at the queue since there was no empirical behavior to follow at that time. In general the empty trips are harder to record than the trips with customer, which generates significant lacks of data for modeling the empirical behavior.

The results of the combined scenarios are presented in table 2, which contains the distribution of behaviors within the fleet (Conservative, Empirical, Informed in real time and informed about predictions), the metrics (Vehicles Distribution-VD and Rides per DriverRpD) and the percentage of customers served-CS.

TABLE 1 Simulation results (combined scenarios).

Simulation scenario	Metrics	Conservative	Empirical	Informed (real time)	Informed (predictions)	CS	Non-informed	Informed
5	VD	49%	48%	2%	2%	82%	96%	4%
	RpD	7.5	8.0	84.4	44.8		7.7	64.6
6	VD	48%	41%	10%	1%	100%	88%	12%
	RpD	8.0	7.9	42.9	30.3		8.0	41.6
7	VD	49%	39%	7%	5%	100%	88%	12%
	RpD	7.8	8.7	46.3	28.5		8.2	38.8
8	VD	49%	26%	15%	10%	100%	75%	25%
	RpD	8.7	9.9	17.5	23.6		9.1	19.9
9	VD	20%	30%	46%	4%	100%	50%	50%
	RpD	12.6	12.5	9.5	30.2		12.5	11.2
10	VD	22%	27%	24%	27%	100%	49%	51%
	RpD	11.9	12.7	9.0	13.6		12.3	11.4
11	VD	8%	23%	18%	51%	100%	31%	69%
	RpD	15.9	15.1	12.7	9.4		15.3	10.3
12	VD	9%	20%	61%	10%	100%	29%	71%
	RpD	15.0	17.3	7.1	27.0		16.6	10.0
13	VD	3%	5%	70%	21%	100%	8%	92%
	RpD	18.4	26.5	6.7	24.3		23.4	10.8
14	VD	3%	5%	47%	45%	100%	8%	92%
	RpD	26.1	27.5	7.3	14.0		26.9	10.6
15	VD	3%	5%	21%	71%	100%	8%	92%
	RpD	22.0	25.1	14.9	9.7		23.9	10.9
16	VD	2%	2%	52%	45%	100%	4%	96%
	RpD	30.8	31.2	7.7	15.1		31.0	11.2

In the combined scenarios, only in the first one there are non-satisfied customers (18%) since the number of informed vehicles is really small (less than 5%). Comparing the performance of each fleet in each scenario it can be concluded that the fleet performing better is the one with the less vehicles (either informed or not), but the number of rides is significantly higher in the informed fleets under the same conditions. It is observed that as the number of vehicles behaving in the same way grows the number of rides decrease due to increased competition, but in the same competitive environment the informed vehicles perform better, as it can be observed in Figure 6.

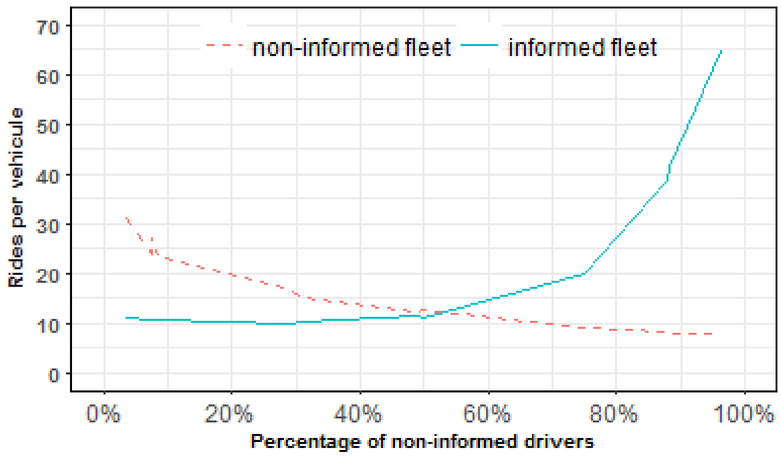

FIGURE 6 Rides per vehicle in informed and non-informed fleets.

The maximum rides per vehicles in the non-informed fleets is 30, even if they are only the 4% of the total number of vehicles, while when the informed vehicles are 4% this value rises up to 64. Oppositely, the informed fleet never gets less than 10.5 rides, even when they are the 96%, while the lowest performance for the non-informed fleets is below 8 rides. These trends show that the informed fleet performs significantly better under the same conditions, even when all vehicles have the same information, with a 25% better performance which is mostly based on the non-served trips of the 100% non-informed fleet. When the number of informed or non-informed vehicles is less than 5% of the total, the informed ones perform 110% better.

6. CONCLUSIONS

Sharing knowledge and information may improve the performance of the provision of taxi services for both the passengers and the drivers, but how this information is shared should be carefully defined. The provision of information to the whole fleet neutralizes its value and misleads the drivers. Therefore strategies on how this information is provided (i.e. to whom and to which extend) are needed. In this paper, initial insights on how informed drivers perform better than non-informed drivers have been presented, but when the informed drivers are the majority, the non-informed drivers perform better. Even in this situation, the drivers having access to predictions perform better than the ones having access only to real time information.

In similar conditions the informed drivers perform better than the non-informed ones, obtaining 100% more rides when they are few vehicles (64.6 versus 31 for the 4% scenario) and 50% when they are majority (11.2 versus 7.7 for the 96% scenario). It not clear how the drivers having access to predictions perform better than the ones having access to real-time information since the first performs better in most scenarios, but in some present a worst performance. In the simulation no commercial fleets were defined, but in a competitive environment, the informed fleets will perform better and reduce their empty distance/time by “stealing” rides from the other fleets.

As future work the authors will firstly focus in the calibration of the empirical scenario, for which better datasets (more days or even a representative day from a statistical point of view) are needed and advance machine learning processes applied. Second, next step of our research will be to analyze how the information should be provided to the drivers in order to avoid “cannibalism” when providing the same information to many drivers, which at the end is beneficial for the non-informed drivers. Initial ideas are to provide information only about some stands or only during a few intervals of the day to each driver as well as to provide a small ranking of the best choices letting them go for the second or third best choice and not always for the first. Finally, the performance of the informed drivers should be further analyzed, since in some cases the drivers having real-time information perform better and in other the ones having access to predictions perform better, without a clear pattern.

7. REFERENCES

[1] Castro, P.S., Zhang, D., Chen, C., Li, S., and Pan, G. (2013). From taxi GPS traces to social and community dynamics: A survey. ACM Computing Surveys (CSUR), 46(2), 17.
[2] Moreira-Matias L., Mendes-Moreira J., Sousa J.F. and Gama J.,: "On Improving Mass Transit Operations by using AVL-based Systems: A Survey". In: IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 1636-1653, July (2015)
[3] Moreira-Matias, L., Fernandes, R., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2012). An online recommendation system for the taxi stand choice problem (Poster). In Vehicular Networking Conference (VNC), 2012 IEEE (pp. 173-180). IEEE.
[4] Chan, N.D., and Shaheen, S.A. (2012). Ridesharing in north america: Past, present, and future. Transport Reviews, 32(1), 93-112.
[5] Chen, X.M., Zahiri, M., and Zhang, S. (2017). Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. Transportation Research Part C: Emerging Technologies, 76, 51-70.
[6] Li, X., Pan, G., Wu, Z., Qi, G., Li, S., Zhang, D., and Wang, Z. (2012). Prediction of urban human mobility using large-scale taxi traces and its applications. Frontiers of Computer Science, 6(1), 111-121.
[7] Douglas G. (1972). Price Regulation and optimal service standards. The taxicab Industry.
[8] Arnott R.(1996). Taxi Travel Should Be Subsidized. Journal of Urban Economics 40, 31-33.
[9] Yang H. and Wong S.C. (1998). A network model of urban taxi services. Transport Research B, Vol. 32, No. 4, pp 235-246.
[10] Wong S.C. and Yang H. (1998). Network Model of Urban Taxi Services. Improved Algorithm. Transportation Research Record 1623, 27-30.
[11] Kim, H., Oh. J.D., and Jayakrishnan, R. (2005). Effect of taxi information system on efficiency and quality of taxi services. Transportation Research Record, 1903:96-104.
[12] Song, Z.Q. and Tong, C.O. (2006). A simulation based dynamic model of taxi service. In Proceedings of DTA2006: First International Symposium on Dynamic Traffic Assignment.

[13] Song, Z.Q. (2006). A Simulation Based Dynamic Taxi Model. Master thesis at the University of Hong Kong.
[14] Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013). Predicting taxi–passenger demand using streaming data. IEEE Transactions on Intelligent Transportation Systems, 14(3), 1393-1402.
[15] Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013). On predicting the taxi-passenger demand: A real-time approach. In Portuguese Conference on Artificial Intelligence (pp. 54-65). Springer, Berlin, Heidelberg.
[16] Chang, H.W., Tai, Y.C., and Hsu, J.Y.J. (2009). Context-aware taxi demand hotspots prediction. International Journal of Business Intelligence and Data Mining, 5(1), 3-18.
[17] Liu, L., Andris, C., Bidderman, A., and Ratti, C. (2010). Revealing taxi drivers mobility intelligence through his trace. Movement-Aware Applications for Sustainable Mobility: Technologies and Approaches, 105-120.
[18] Ziebart, B.D., Maas, A.L., Dey, A.K., and Bagnell, J.A. (2008, September). Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. Proceedings of the 10th international conference on Ubiquitous computing (pp. 322-331). ACM
[19] Yuan, J., Zheng, Y., Xie, X., and Sun, G. (2013). T-drive: Enhancing driving directions with taxi drivers' intelligence. IEEE Trans. on Knowledge and Data Engineering, 25(1), 220-232.
[20] Lee, J., Shin, I., and Park, G.L. (2008). Analysis of the passenger pick-up pattern for taxi location recommendation. IEEE Network2121 Computing and Advanced Information Management. NCM'08. Fourth International Conference on (Vol. 1, pp. 199-204).
[21] Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., and Yang, Q. (2011, March). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In Pervasive Computing and Communications Workshops (PERCOM Workshops). IEEE International Conference on (pp. 63-68).
[22] Phithakkitnukoon, S., Veloso, M., Bento, C., Biderman, A., and Ratti, C. (2010). Taxi-Aware Map: Identifying and Predicting Vacant Taxis in the City. AmI.
[23] Ge, Y., Xiong, H., Tuzhilin, A., Xiao, K., Gruteser, M., and Pazzani, M. (2010). An energy-efficient mobile recommender system. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 899-908). ACM.
[24] Yuan, N.J., Zheng, Y., Zhang, L., and Xie, X. (2013). T-finder: A recommender system for finding passengers and vacant taxis. IEEE Transactions on Knowledge and Data Engineering, 25(10), 2390-2403.
[25] Hwang, R.H., Hsueh, Y.L., and Chen, Y.T. (2015). An effective taxi recommender system based on a spatio-temporal factor analysis model. Information Sciences, 314, 28-40
[26] Salanova J.M., Estrada M., Aifadopoulou G. and Mitsakis E., (2011). A review of the modeling of taxi services; Procedia and Social Behavioral Sciences 20 150-161.
[27] Mitsakis E., Stamos I., Salanova Grau J.M., Chrysohoou E., Aifadopoulou G. (2013) Urban Mobility Indicators for Thessaloniki, Journal of Traffic and Logistics Engineering (JTLE) (ISSN: 2301-3680), Vol. 1 No. 2. pp. 148 – 152.
[28] Salanova Grau J.M., Chaniotakis E., Toumbalidis J., Karanikolaos N, Aifantopoulou G. Big data for transportation analysis and trip generation. Transport Geography. Special issue "Big Data: a new opportunity for Transport Geography?" (under review)
[29] Salanova J.M., Estrada M.A. (2015) Agent based modelling for simulating taxi services. Procedia – Computer Sciences, Vol 52, pp 902-907.