Journal of Traffic and Transportation Management
Volume 1, No. 1 (2019) pp. 27-36

Modeling framework for supporting taxi policy making

Josep Maria Salanova Grau [a] *, Miquel Estrada [b]

[a] *Centre for Research and Technology Hellas/Hellenic Institute of Transport, Thessaloniki, Greece, 57001*
[b] *Technical University of Catalonia, Barcelona, Spain, 08034*

Abstract

Taxi services account for a significant part of the daily trips in most cities around the world. These services are regulated by a central authority, which usually monitors the performance of the taxi services provision and defines the policies applied to the taxi sector. In order to support policy makers, fleet managers and individual taxi drivers, there is a need for developing models to understand the behavior of these markets. Most of the models developed for analyzing the taxi market are based on econometric measurements and do not account for the spatial distribution of both taxi demand and supply. Only few simulation models are able to better understand the operational characteristics of the taxi market. This paper presents a framework for the development of taxi models both aggregated and simulation-based. It is aimed at assessing policy makers, taxi fleet managers and individual drivers in the definition of the optimum operation mode and the number of vehicles.

Keywords: *taxi modeling, agent-based modeling, modeling framework*

1. Introduction

More than 80% of the world's population is expected to live in urban settlements by 2030; its mobility needs will be among the main challenges faced in the near future and addressing them will require both conventional and innovative alternatives [1]. This concentration of population creates large cities with high densities and mobility needs, which in part are covered by taxis. Taxi services are present in most of cities around the world, accounting for a significant part of the daily trips. There are three basic organizational and operational modes: stand, hailing and dispatching. In the stand mode taxis and users meet at predetermined locations, called taxi stands or ranks, where a first-in-first-out (FIFO) system applies for both the users' and the drivers' queue. In the hailing mode, taxis circulate looking for a user, and users wait for the first vacant taxi. Finally, in the dispatching market, taxi services are coordinated by dispatching centers, which are responsible for matching available taxi services with the demand of users' trips. Taxi markets are usually composed by the three operation modes, but when the demand for taxi services is small, the stand and dispatching modes are the most usual ones. On the other hand, the hailing model is usually found in cities with high population densities where a Business District zone concentrates a high percentage of the daily trips.

The taxi market is an important one, especially in the cities, where the mobility share ranges between 1% and 15%. This market was strongly regulated during the last century, but during the last decade this sector have been significantly changed by technological advances, not only with regards to the technology itself, but also having an impact to operational and organizational issues. At the end of the 20th century most taxi drivers were affiliated to dispatching centers aiming at receiving rides from customers' calls, being this the only communication channel between the drivers and the customers, while their knowledge about the network and the demand habits were crucial in order to be in the right place at the right time. Today, thanks to the penetration of mobile phones and smart devices, the communication between drivers and customers can be direct, without the need of the dispatching centers. This opportunity has been taken by third parties, which provides disruptive services putting in contact taxi drivers and customers in a more fashion way, more attractive for the users, who rapidly adopted it. This new scenario replaced the traditional dispatching centers by third parties running an app and allowed for new business models, such as the one launched by UBER, where taxi drivers are not needed anymore. The reaction of the traditional players was slow, but after a few years they are also being transformed and offering apps to their drivers and customers.

Various models have been developed for evaluating the costs of the provision of taxi services in terms of waiting time of customers and income of taxi drivers, providing policy makers

* Corresponding author. Tel.: +302310498433
Fax: +302310498269; E-mail: jose@certh.gr

DOI: 10.5383/JTTM.03.01.004

with methodologies for estimating the optimum number of licenses for each demand level and city (in terms of size, network geometry and congestion level) as well as identifying the best market regulation and vehicle operation mode for each city. These models depend on a set of variables that determine the performance indicators and the operation cost of the whole fleet and can be classified into aggregated and simulation models. Various formulations have been presented for estimating these variables mostly for the dispatching market, resulting in the calculation of the generalized system cost and the respective optimum fleet generating the minimum unitary system cost, while a few simulation models have been developed for the stand and dispatching markets. This paper aims at providing a framework and guidelines for modeling taxi markets based on the research done by the authors, which started with analyzing the state of the art in 2011 and developing the first models and framework in 2013. In 2015 the authors published the application of both models to the city of Barcelona. During the years 2016-2018, the authors had various publications from the application of the models, related to taxi policy making, taxi operational modes, dispatching strategies and information sharing.

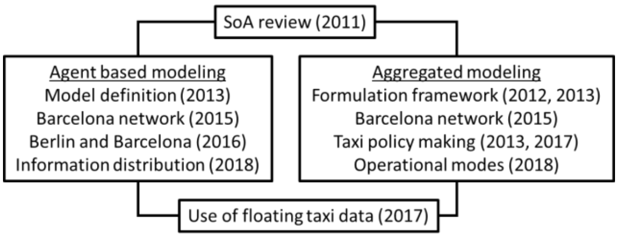


Fig. 1. Outline of the research done by the authors

The paper is structured as follows: the work flow of the research done by the authors that support this publication is presented below. A continuation, a brief literature review on aggregated and simulation modeling is described in chapter two, while the modeling framework for both model types is developed in chapter 3. Finally, conclusions are presented in the final chapter.

2. Literature review

2.1. Aggregated modeling

Many studies have been published in relation to the taxi sector focusing on its profitability using aggregated models based on continuous variables for analyzing regulation issues. The first models were basically econometric models focusing on the profitability of the taxi sector; they mostly used aggregated values for representing all the variables of the taxi services. Later studies presented more granular models, taking into account the spatial dimension of the taxi market as well as traffic congestion or travel demand elasticity. The third group of taxi models is the simulation-based ones, dealing with discrete-event models of dispatching or stands markets for reproducing the complex multi-agent system of the taxi services provision.

The first model developed for evaluating the performance of the taxi services was developed by [2]. He presented an aggregated model using economic relationships of the goods and services sectors. [3] used the full price demand function proposed by Douglas, adding an index of the full prices of all other goods, while introducing the value of time of customers and the waiting time in the demand assumptions. [4] developed a theoretical model in a regulated market where dispatch and airport cabstand are the primary modes of operation. [5] provided a model of a taxi stand, assuming Poisson customer arrivals, negative exponential service time and FIFO queue rule. [6] redefined the demand proposed by [2] and its relation to the supply. They assumed uniform demand within the day, which decreases as waiting time increases. [7] analyzed the shadow cost of taxis in the first best solution, proposing subsidization for covering these costs in the vacant trips of taxis.

Yang and Wong developed the first equilibrium model in 1997, taking into account the spatial distribution of demand and supply in the city using traffic assignment models. The sophisticated models presented by [8] – [11] are able to take the spatial distribution of demand and supply in the city into account, using traffic assignment models and are able to account for congestion, elasticity of demand, different customer classes, external congestion and non-linear costs. Recently, [12] presented a mixed model combining the different taxi operation modes. Most authors developed models for analyzing the effects of regulations in the taxi market, such as [13] and [14], proposing mathematical formulations for calculating demand and supply and simulating different types of markets and regulation schemes.

A detailed review of the aggregated and equilibrium models of taxi services can be found in [15].

2.2. Simulation modeling

[16] was the first that used the word agent-based model for designing computational models able to simulate actions and interactions of autonomous entities. [17] examined the link between the agent-based modeling and the current transportation problems and presented the definitions of agent and agent-based modeling as well as their attributes and structure. His work concluded with a study presenting various applications of agent-based modeling to the traditional transportation theory.

[18] and [19] continued the work started by Kikuchi and extended the review of the use of agent-based in the transportation field. Various agent-based models have been developed afterwards, but only a few of them for modelling taxi services, the main contribution of which is the possibility of obtaining more detailed results about the optimum number of vehicles and the operation mode while analyzing the behavior of the taxi drivers and users with regards to the impact of the provision of information. [20] developed an agent-based models for investigating the relation between the number of taxis and the performance in the dispatching market, concluding that the waiting time is relatively insensitive to changes in demand but highly sensitive to changes in the number of taxi cabs. [21] developed a discrete event method able to simulate dispatching taxi services, the authors obtaining a linear relation between total distance and fleet size; [22] developed a simulation-based stand taxi services and proved that the use of information technologies could improve the quality of the service by 20%.

[23] and [24] presented dynamic taxi stand demand models, highlighting the limitations of traditional aggregated models, which are time-dependent patterns, imperfect information, learning convergence and non-equilibrium in taxi market due to regulation. The authors also tested the effects of Advanced Transport Information Systems (ATIS) in this specific market. [25] presented a dispatching architecture for the increase of the customers' satisfaction by concurrently dispatching multiple taxis to the same number of customers in the same geographical region. [26] developed a discrete-event

simulation model for answering the "what-if" questions of the taxi market, supporting that the mathematical models are out of reach for such a complex stochastic multi-agent system composed of taxis, users and the network. [27] presented an event-based simulation stand taxi model for analyzing the customer-searching behavior of individual drivers and its influence on the performance of the system.
[28] developed a time-dependent agent-based taxi simulation model and tested it with various customer patterns in order to provide policy-related guidelines for improving the service performance when the demand pattern is asymmetric. [29] proposed a massive multiagent simulation platform for investigating interactions among taxis and customers, incorporating real-world driver's behaviors validated in a real-world case study. [30] developed an agent-based model for assessing the performance of various operation modes and applied to the Sioux Falls network. [31] developed an agent-based model for obtaining the optimum number of vehicles and applied to the city of Barcelona using real-world data.
[32] presented a large-scale microscopic simulation of taxi services build using floating car data of Berlin and Barcelona for proving the performance of various dispatching and fleet management techniques. [33] studied the impact of the provision of information to taxi drivers by developing an agent-based model for the city of Thessaloniki in Greece, proving that the informed drivers perform significantly better, especially when they are a minority.

3. Framework for modeling the taxi sector

3.1. Aggregated modeling

3.1.1. Problem formulation

The first step is to identify the variables that will be used for formulating the problem to optimize (maximize or minimize depending on the objective function defined), including the decision variables. Once all the variables have been identified and linked to each other (when possible), the costs function of each stakeholder should be defined using monetary or time costs and aggregated into the objective function. Constraints may be defined in continuation depending on the problem to be solved. Usually the most important decision variable of the model is the number of vehicles per unit of area and surface but others may be used. It is important to list and explain all the hypothesis used during the formulation of the problem (e.g. homogeneous demand).

3.1.2. Estimation of variables

The second step is to minimize the number of independent variables by obtaining formulations that relate them. Various formulations can be found in [34] and [35]. Geometric relations, physical, statistical and probabilistic formulations may be used. For example, the average in-vehicle travel time can be estimated as the quotient of the average in-vehicle distance of one trip and the cruising speed of the taxi vehicle; at the same time the average in-vehicle distance can be calculated by considering the region as a square of side   and estimating the expected distance between two random points within the region applying a correcting factor (r) depending on the layout of the network. The average distance travelled and access/waiting time will depend on the operation mode. Finally, the formulation of the external costs may be calculated using the congestion created by taxis or the additional emissions.

3.1.3. Optimum fleet size calculation

The optimum fleet size (first best solution) is obtained by deriving the cost formulation and finding the fleet size for which the derived formulation is equal to zero, while the fleet size related to the second best solution is obtained by finding the fleet value that has a zero cost (and therefore zero benefit) for the drivers. A third solution may be provided by estimating the fleet size that will generate a concrete benefit for the taxi drivers, which tries to reflect the benefit of the license holders. The cost functions are second degree polynomials.
Once the optimum formulations are obtained is important to group them properly aiming at identifying the components of the formula. For example, in [34] the first part of the formulation for the optimum fleet size is the minimum fleet size for serving all trips while the second term is the extra fleet needed for providing a better LoS to customers, maintaining a satisfactory profit to taxi drivers. The constant value of this extra fleet is directly proportional to the Value of Time (VoT) and inversely proportional to the cruising speed and hourly cost, which means that high VoT of taxi customers implies a higher extra fleet in order to reduce waiting time; high hourly operating cost implies fewer taxis for reducing the vacant distance and time of taxis; higher speeds are related to smaller taxi fleets due to the higher performance of the vehicles. Figure 2 shows this behavior for different demand levels.

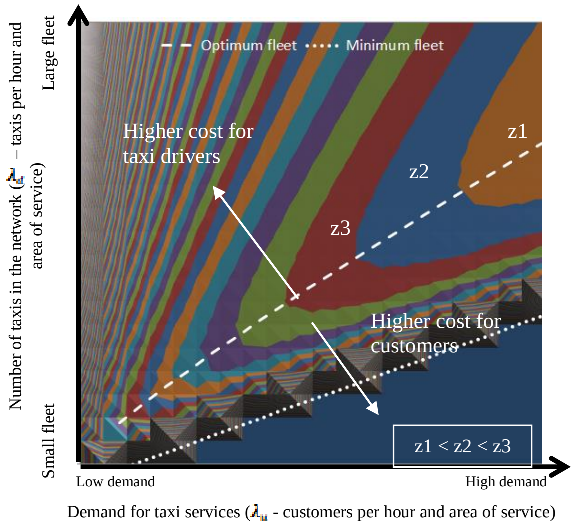


Fig. 2. System unitary cost (z) of each demand and supply configurations for the dispatching market. Source: [34]

Figure 3 presents a section of Figure 2, where the demand for taxi services is fixed. Each demand level will define a unique section, which has a fleet size M for which the total unitary cost is minimum (first best solution). It can be observed that for this supply size the cost of the drivers is positive, which means that the benefits are negative.
The second-best solution (N) is the point where the costs of the taxi drivers are zero, which implies a small increase in the waiting time in relation to the first best solution. Taking into account the benefit of the taxi license due to the exploitation of this asset, a third point (T) could be identified, where the benefits for the license holders are equal to the expected benefit (B), which is independent of the operation mode.

The expected benefit can be calculated as the opportunity cost of the taxi license value in comparison to a more secure investment or as the depreciation cost related to the purchase of a taxi license. A detailed discussion and methodology for the estimation of the expected benefit is presented in [36].

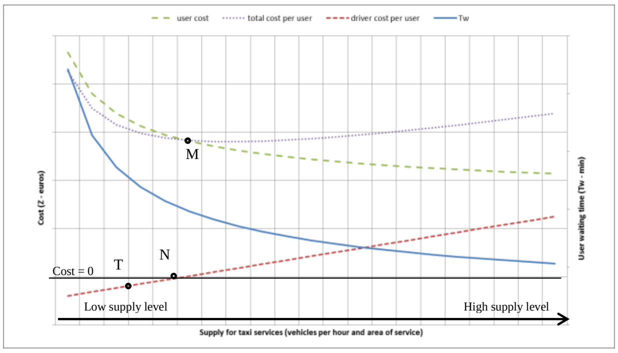


Fig. 3. Waiting time (Tw), customer, driver and system unitary costs of different demand levels for the dispatching market. Source: [34]

3.1.4. Results

The results may be presented in different ways, depending on the scope of the study and the proposed formulation and decision variables. For example, [34] presents the system costs and the costs of each individual actor for different fleet sized, identifying for the Barcelona taxi market both the first- and second-best solutions.

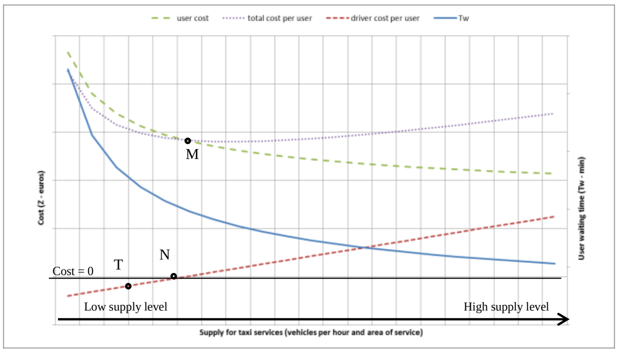


Fig. 4. Waiting time, driver benefits and unitary costs for each fleet size obtained by the aggregated model (dispatching operation mode). Source: [34]

As observed in Figure 4, the application of the dispatching mode in Barcelona shows that the number of taxis per hour and km2 should be higher than 29 since the waiting time for smaller fleets is very high due to the low number of free taxis. The strict minimum fleet size is 28 taxis/hour*km2, which means that with this taxi fleet size, the number of demanded customer hours are equal to the offered vehicle hours. It can also be observed that the maximum number of taxis with positive benefits is 38 taxis/hour*km2 since more taxis than this value will generate losses to the drivers. Therefore, 38 is the second-best solution. The optimum number of taxis taking into account the system costs is 44 – 46 taxis/hour*km2 (first best solution). First best solution can be profitable for taxi drivers if each taxi trip is subsidized by the state by an amount of 3-4 euros per trip. Table 1 summarizes the results of the hailing, dispatching and stand model applications to the city of Barcelona. In the hailing market, the first- and second-best solutions coincide due to the inclusion of congestion externalities in the system costs formulation, so there is no need for subsidizing the taxi market.

Table 1. Results of the application of the aggregated hailing model in the city of Barcelona. Source: [34]

First best solution	34 – 38 taxis/hour*km^2
Subsidy	No subsidization needed
Second best solution	36 taxis/hour*km^2

Table 2. Results of the application of the aggregated dispatching model in the city of Barcelona. Source: [34]

First best solution	44 – 46 taxis/hour*km^2
Subsidy	3 – 4 euros per trip
Second best solution	38 taxis/hour*km^2

Table 3. Results of the application of the aggregated stand model in the city of Barcelona. Source: [34]

First best solution	52 – 54 taxis/hour*km^2
Subsidy	10 euros per trip
Second best solution	37 taxis/hour*km^2

The results can be also related to the fare system of the taxi market. If the demand in considered to be elastic, the effects of different fares on the demand or the drivers' benefit can be estimated and linked to the fleet size, as done in [37] (figure 5).

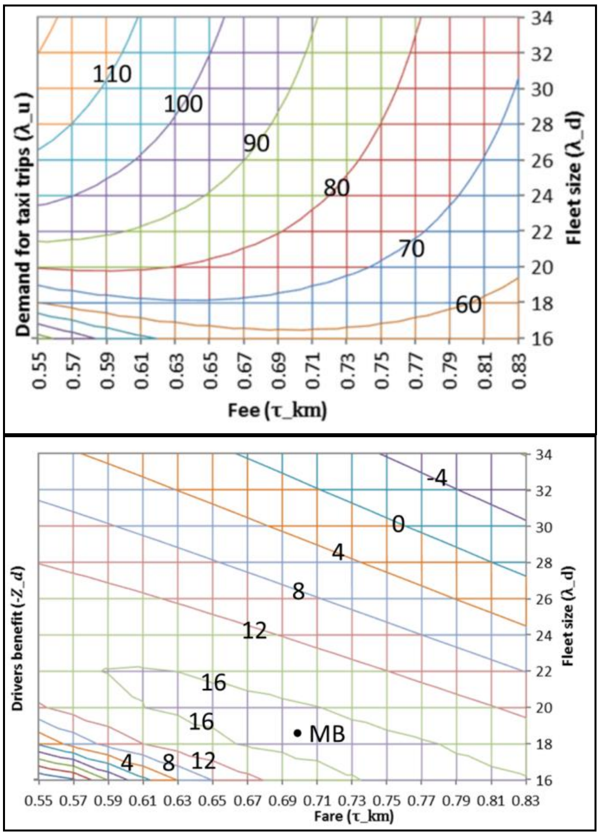


Fig. 5. Effect of fares and fleet size on demand for taxi trips. Source: [37]

Applying the results to various intervals of the day allows for analyzing the impact of different shifts policies. [37] developed a methodology for obtaining the best shifts distribution in Barcelona in order to minimize the systems costs along the day (figure 6) quantifying the impact of the under/over supply during the day in the waiting time of customers and the benefits of drivers for different taxi operation modes (figure 7). In addition, by adding two consecutive 4h shifts, the drivers can work 8h consecutive hours 80% of the days, therefore only once a week they are forced to work in two shifts, still being able to use the "rest" period between the shifts for attending personal issues.

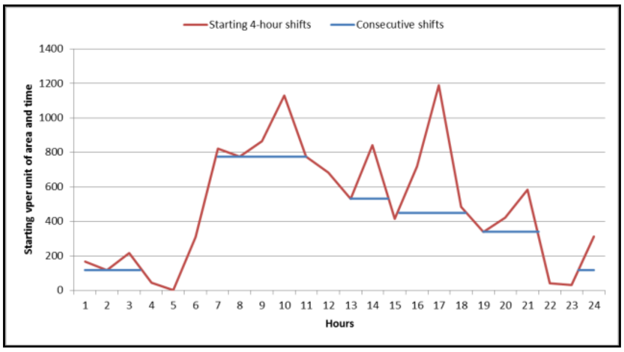


Fig. 6. Grouping of pairs of 4-hour shifts. Source: [37]

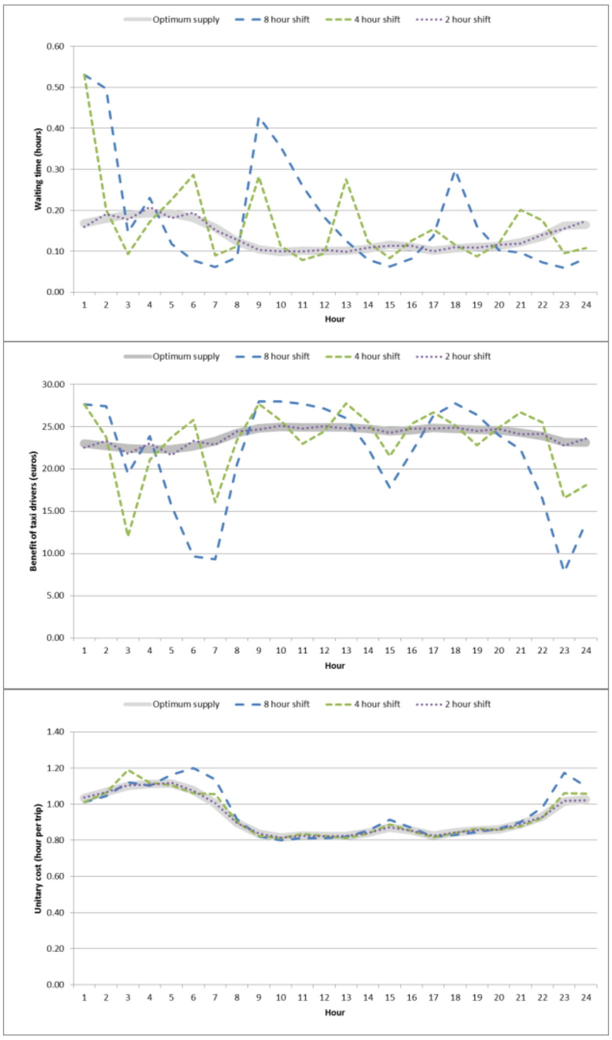


Fig. 7. Customer waiting time, income and unitary cost for the various proposed policies. Source: [37]

3.2. Simulation modeling

A 4-step framework is proposed for the simulation of how information can be shared in the taxi sector, namely data collection and analysis, scenario set up, agent-based model development and scenarios assessment.

3.2.1. Data collection and analysis

The first step is to collect and analyze the available data, most of the times generated by the individual taxis and available at the taxi associations. The data may be generated in various formats and contain different information. The most common format is Floating Car Data (FCD), where the locations of the vehicles are logged following a spatial or temporal interval (figure 8 down). FCD usually contains UTM coordinates, timestamp, orientation, status (empty or occupied), stand id (if waiting at a stand) as well as vehicle and driver identities. Regarding the spatial and temporal intervals, the triggers actuating the log of the data can be a pre-defined distance or time, but also an event, such as picking-up or dropping-off a customer.

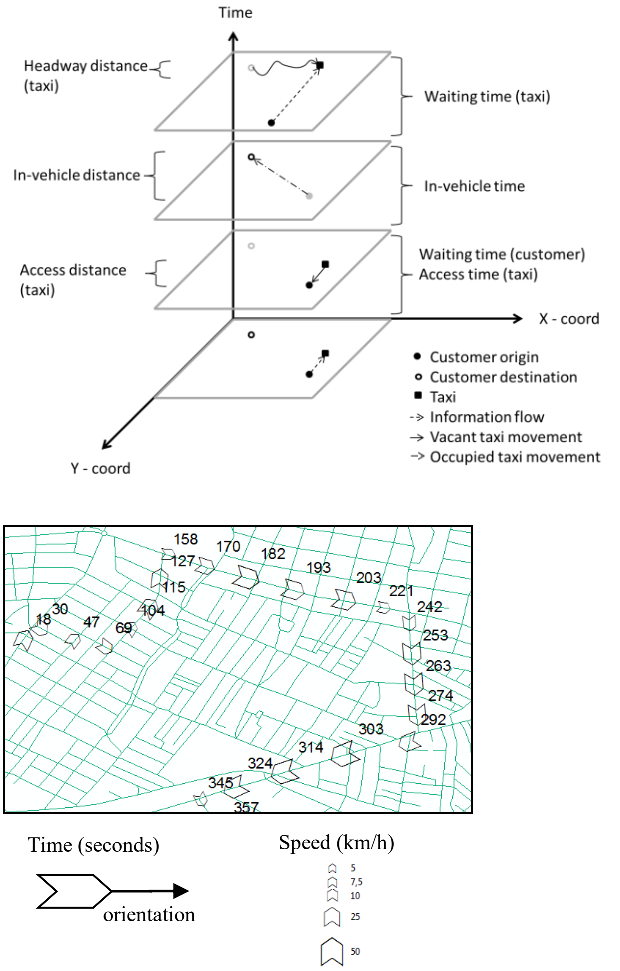


Fig. 8. Space-time diagram of the dispatching market. [30] (up). Floating car data sample. [38] (down)

Once the data are available, it should be analyzed in order to understand how the market works by estimating various indicators, including city related variables such as number of vehicle-kilometers and vehicle-hours, number of total trips, total costs of the system (drivers costs and users costs); taxi driver related variables such as circulating time and distance (total, occupied and vacant), taxi occupancy and vacant taxi headway as well as earnings and benefits; taxi customers related variables such as waiting time, travel time and cost of trip. All the variables should be mapped to the operation mode (hail, stand and dispatch) both temporally and spatially (figure 8 up).

These indicators should be generated for large series of days in order to identify the differences between weekdays and weekends as well as to define the characteristics of the typical day that will be used in the simulation. Performance indicators should be also defined, such as number of trips per driver, empty and occupied distance and time per driver as well as average speed. Once all the indicators have been defined and the calculation method defined, a representative distribution should be identified to characterize each of them, which will be the input of the simulation. These distributions should be validated towards the functional characteristics of the network, such as the peak hours.

Exogenous variables, such as the total demand (absolute value, temporal distribution and geographical distribution), supply (number of vehicles, temporal distribution and mode distribution), network geometry and links characteristics and taxi fare policy (fixed fare and distance/time-based charge rates) should be also collected and analyzed for the city in which the model will be developed (figure 9).

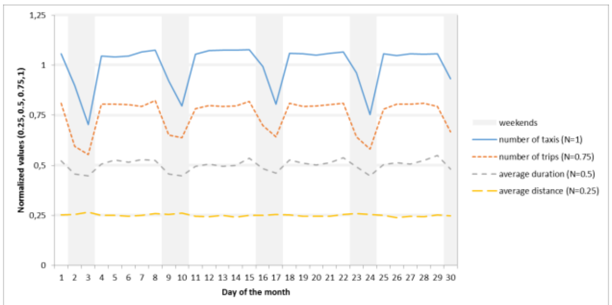


Fig. 9. Number, total and average duration and distance of the taxi trips in Thessaloniki for one month normalized (N) to 1, 0.75, 0.5 and 0.25 respectively. Source: [39]

3.2.2. Scenario set-up

In a second phase, the data needed to feed the simulation tool should be collected and preprocessed. Road segments and intersections should be modeled in GIS including their functional characteristics, such as free flow speed, volume-delay functions, length and many others. At this point, various approaches can be followed:

- To model the impacts of the taxis in the network by defining the travel time in each section, based on the number of taxis. This is the most challenging and resource-consuming approach, since the car traffic may be modeled together with a routing engine.
- To use the network as a background where the drivers circulate without having impact on its characteristics, meaning that the travel times will be pre-defined. In this approach there is no need for modeling private vehicles traffic, but a routing engine is also needed.
- Network performance, distances and travel times between zones are estimated from the collected data and are not modelled. This is the simplest approach since there is no need for modeling the private vehicles traffic flows neither developing a routing engine.

Once the network geometry and functionalities have been defined, the operative issues should be addressed, including the type of taxi market operation (hail, stand, dispatch or a combination), the shifts and the rules of the simulation agents. The two main agents to be simulated are the taxi drivers and the taxi customers. Taxi drivers and customers will be provided with decision capabilities based on various levels of information availability aiming at maximizing an objective function. For the taxi drivers, this function may mainly include income or earnings but also productivity metrics, such as occupied time/distance versus empty time/distance, while the customers' main metrics is time. These metrics can be shared by the agents in a system optimum approach, were drivers will be also trying to minimize customers' waiting time, or independent from each other, in a more user optimum approach. Even within the taxi drivers, there can be groups of collaborating agents aimed at maximizing the group revenues in front of the other groups. Customers can also be provided with collaborative capabilities.

- Behavior rules of the agents should follow the patterns extracted from the data. For example, the dataset containing the empty trips of the drivers is fundamental for understanding how they behave when looking for a customer, if they just wait at the stands or circulate in specific zones. Information availability is modeled at this stage. Various driver behaviors can be considered, mainly grouped in conservative, empirical and informed, as developed below:
- Conservative drivers stay at the last customer's destination zone and join the existing queue. This kind of driver may be a low-experienced one having low (or none) knowledge about the spatial and temporal demand profile.
- Experienced drivers select a zone based on their experience and go there. They balance the distance and travel time to the new zone versus the ride they expect to find, defined by both waiting time and income.
- Informed drivers may be supported by real-time information about taxi service demand, including future predictions. They decide to visit other zones based on the current demand (as experienced by the drivers during the last time interval) or on available predictions.

3.2.3. Agent-based model development

The third step is to develop the modules of the simulation engine. These can be classified into the following categories:

3.2.3.1. Demand and supply generation

Taxis and customers will be fitted into the network following the distributions identified during the data analysis process. Customers' origin, destination and appearance time can be stochastically determined by the identified distributions or deterministically introduced from the observed data. Drivers starting time, location as well as shift duration can be also extracted from the data and modeled. Locations can be specific (x and y coordinates, stand identity) or less detailed (zone). If a network is needed for calculating the routes it should be defined as a directed graph composed of links and nodes. The links should have the following characteristics: Length; Free flow travel time; Volume-delay parameters; Flow of vehicles; Capacity. The demand can be elastic or inelastic and should be characterized by demand density functions. If it is elastic, these functions should contain all variables that alter the demand elasticity.

3.2.3.2. Flowchart diagram

Once all components and modules have been addressed, the flowchart between all processes should be defined, emphasizing on their relations and the temporal dimension in which they are executed sequentially. Depending on the available components and agents, the connections between the modules may differ. Modules can be dedicated to demand and supply generation, vehicle movement and decision making, vehicle routing and navigation, vehicle-user meeting, drop-off, learning/experience enrichment, trip charging module, traffic update module among many others, always depending on the goal of the simulation and the available data. Various flowcharts are provided in figure 10.

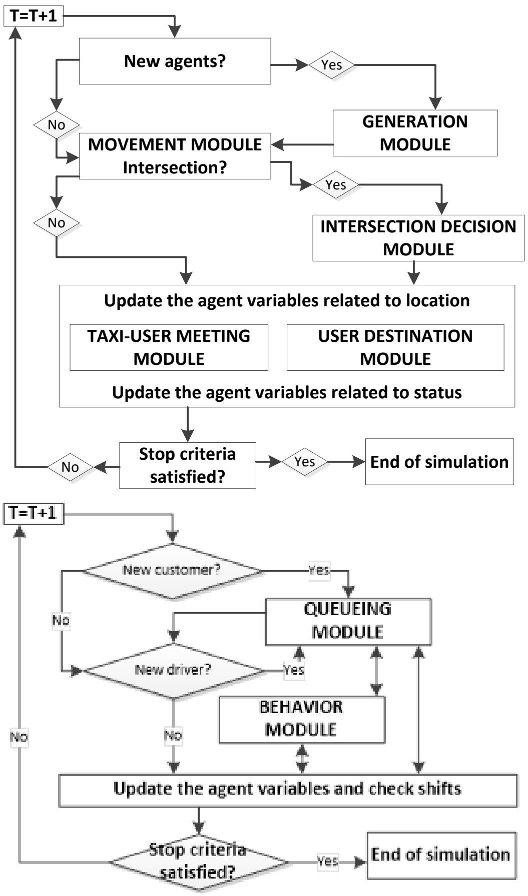


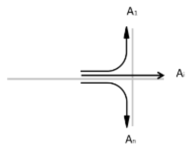
Fig. 10. Flow chart diagram of [31] (up) and [39] (down).

3.2.4 Routing and decision making

If a background network is provided, a routing engine is needed to define the path between two locations. Professional drivers take into account other parameters in comparison to normal drivers, and they have a good knowledge of the network (even including real time information provided by other colleagues about bottlenecks). Therefore, route choice models can be used in order to understand how taxi drivers plan their routes. En-route decision making may be also modeled, providing re-routing options when congestion appears. During the empty kilometers, the routing component may be totally different, since the drivers are looking for a customer instead of going to a concrete location.

Speed of taxis within the network can be estimated as equal to the speed of the private vehicles, which it is a function of the geometric and functional characteristics of the road section as well as of the congestion level. Volume-delay functions can be used, such as the link performance function proposed by 20 and adopted by the Bureau of Public Roads. If taxis can run along dedicated lanes (bus or HOV lanes), this can be taken into account in the simulation by differentiating the travel time or the speed in the set of links where this infrastructure is available and using free flow speed.

When looking for a customer, the experience of drivers can be modeled by using a roulette rule, where every decision concerning where taxi should drive towards is based on the attractiveness of each road section or a concrete zone. Experience and learning process can be modeled by updating the attractiveness of each zone or road section during the simulation. This can be done also at driver-level, hence each driver will have its own learning experience. An illustration is provided in figure 11.



$$P(A_i) = \frac{w(A_i)}{\sum_{j=1}^{n} w(A_j)}$$

$w(A_i)$ is the attractiveness of link

$$next\ link = A_i\ if\ \sum_{j=1}^{i-1} P(A_j) < Rn \leq \sum_{j=1}^{i-1} P(A_j) + P(A_i)$$

Fig. 11. Intersection numeration (up left), location of an agent (up right) and roulette for the intersection decision procedure (down). Source: [31]

3.2.5 Simulation run and scenarios assessment

Once the model has been completely defined, the scenario simulation can be run and the outputs of the model collected. Various runs are needed with different random seeds for the creation of the stochastic variables, using the average outputs obtained. However, users should check their distribution in order to eliminate the outliers and obtain average output distributions with small variability. If the variability of the outputs is high, the model should be checked and sensitivity analysis done in order to identify the variables and parameters responsible for this variability. In that case, we should try to better estimate them, or select other periods with less variability. The sensitivity analysis is also useful for identifying the critical variables, parameters, components and modules that highly affect the model's sensitiveness, and put more effort in their modeling and/or estimation.

3.2.4.1 Validation

The outputs of the model should be validated towards exogenous variables of the trips, such as travel time, distance or cost as well as towards exogenous variables of the customers, such as waiting time, which may be collected and processed independently of the data used. In the validation process the parameters of the model should be calibrated in order to achieve a good fit with the exogenous measurements, such as the ones related to routing preferences and driving behavior of the drivers or to the charging system. Perfect validation may not be possible due to the assumptions done in the simulation model, such as constant speed during the whole simulation, which will end up to overestimation and underestimation of travel times when the network is more or less congested respectively in comparison to the congestion level used for defining the speed in each road section. Usually validation towards distances provide better fit than towards travel time, since it is less dependent on the congestion level (see figure 12).

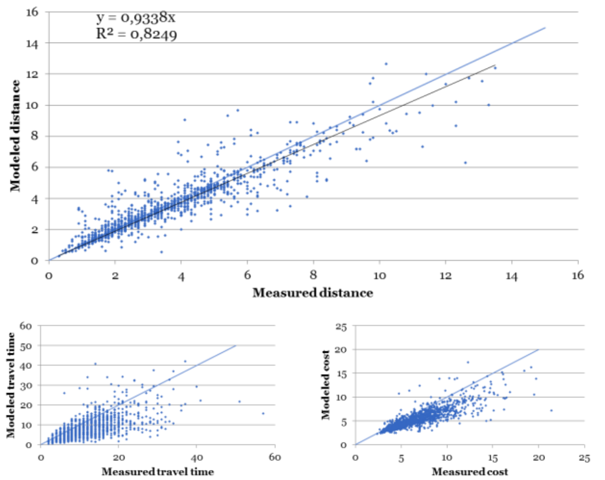


Fig. 12. Validation towards distance (up), travel time (down) and cost (right). Source: [31]

Validation can be also done with regard to the number of vehicles in each state during the simulation. It allows a better tackling of the dynamics of the system as well as daily variations in demand and supply for taxi trips. Lacks in the monitoring data or lack of capabilities of the model for simulating the breaks of the drivers between shifts make this kind of validation more difficult. Finally, validation can be done using exogenous data sources, like statistical data collected and estimated by third parties, such as taxi organizations or public authorities related to the taxi sector. In this case economic variables like incomes or expenses can be used together with more operational ones, such as the working hours and shifts based on regulation (figure 13).

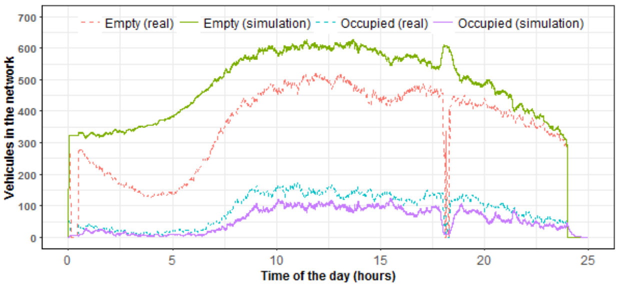


Fig. 13. Validation towards amount of vehicles in each status. Source: [39]

3.2.4.1 Results

Once the model has been validated, the indicators defined earlier in the preparation of the model can be calculated from the results generated by the simulation runs. It is important to note that the whole model was built for estimating these indicators, which means that other set of indicators may not be as accurate as the original ones. Three illustrative examples are provided below:

- Estimation of the rides per driver based on the behavior and access to information and predictions in Thessaloniki (using one day of real data containing trajectories and status of all vehicles).

From figure 14 (up), the difference in the number of rides per driver for informed drivers versus non-informed can be observed. When the informed drivers are minority, they outperform up to 100% better than the non-informed ones under the same conditions. This ride increment percentage is reduced to 25% when the informed drivers are majority, which is still a good result.

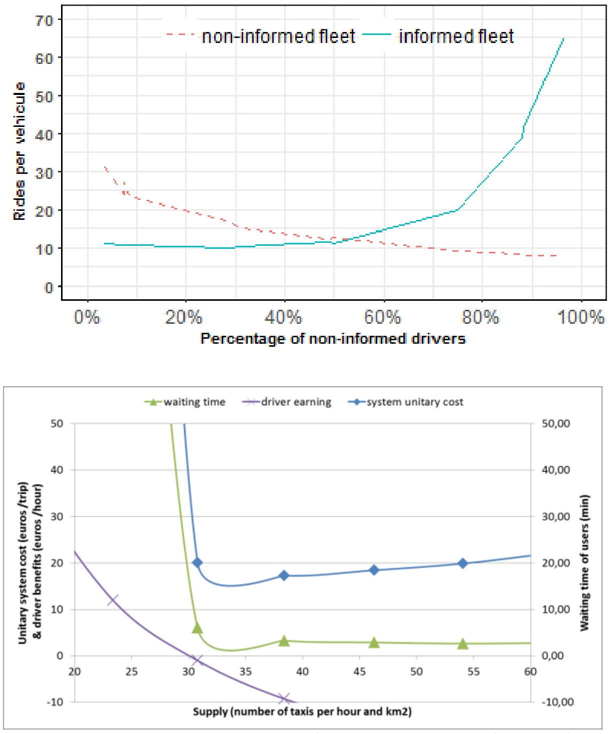


Fig. 14. Rides per vehicle in informed and non-informed fleets. Source: [38] / Waiting time, driver benefits and unitary costs for each fleet size. Source: [31]

- Estimation of the best operation taxi operation mode in the Sioux Falls network (using generated data).

Table 4. Simulation Results of the Three Operation Models. Source: [30]

	Dispatching	Stand	Hailing
Optimum fleet for system cost (vehicles)	8	10	14
Average occupied distance (m)	6.100	4.900	3.500
Average vacant distance (m)	4.200	330	14.000
Average occupied time (min)	25	20	14,5
Average vacant moving time (min)	35	16	90
Income (euros/h)	67	54	38
Driver unitary earnings (euros/hour)	32,5	30	-8
Ratio occupied/vacant moving time	0.65 (35-40%)	1.3 (55-60%)	0.16 (10-15%)
Rate occupied/vacant distance	1.5 (60%)	15 (94%)	0.25 (20%)
Average user waiting time (sec)	136	65	392
User unitary cost (euros/hour)	6,45	6,25	7,5
System cost (euros/hour)	385	325	865

From the results (table 2), it can be observed that the dispatching mode has the lowest number of vehicles requested and therefore the highest income for the drivers. However, it presents larger waiting time for the users than stand mode, which leads to the lowest system cost.

- Estimation of the optimum number of vehicles that minimizes the generalized system cost function in Barcelona (using 9 years of real data containing customer trip origins and destinations).

The social optimum number of taxis per hour and km2 is 33-34 (figure 14 down), while the maximum number of vehicles that will not generate losses to taxi drivers is 30. Therefore, social optimum solution in not feasible in this case unless local authorities subsidize taxi sector. If subsidization is not feasible, there will be 30 taxis/hour*km2 instead of 33-34, increasing the waiting time of the customers by a few minutes.

3. Conclusion

The entry of third parties to the taxi sector together with the deregulation of their markets has resulted in the redesign of operational rules and policy frameworks of the taxi sector worldwide. New models should be developed for supporting decision makers in how to regulate and how to operate the taxi market from a public and private perspective respectively. During the next years, studies on the impact of innovative mobility schemes, some of them involving taxis, such as mobility as a service, will be supported by new modeling and simulating capabilities, allowing to replicate and to analyze such complicate environments. Aggregated, activity-based and agent-based models can fulfill these needs thank to the larger availability and granularity of the data.

Still before the acceptance of MaaS schemes the taxi sector has seen the introduction of third party service providers aiming to facilitate the matching of supply and demand, not always related to the taxi sector or even against them in some cases. Two kinds of applications have been developed in this direction: the first group of applications focused on sharing available places in vehicles aiming at sharing the transport costs among the different users (e.g. blablacar), without a willingness to make profits but to reduce costs; the second group of applications aimed at providing a platform for putting in contact potential customers and drivers, just as traditional dispatching centers were doing during decades of years, but with a more innovative and attractive way for the customers.

With regards to data availability, the large amount of available data allows for developing data intensive models, not only for transport planning purposes but also for operation issues. There is a need for defining the right modeling frameworks allowing for correctly handling the new datasets and delivering useful and reliable transport models. Nowadays agent-based modeling is a powerful tool able to reproduce complex environments with reliability and provide detailed and useful outputs without large computation needs, while aggregated models are easier to build and able to provide faster results towards the main variables on average terms. In order to develop such models, data science skills should complement the domain expertise.

Finally, the introduction of autonomous vehicles merged with new sharing schemes will change significantly the way taxis work and therefore will request for enhanced models with fleet management capabilities.

References

[1] United Nations 2007, World Urbanization Prospects: The 2007 Revision.

[2] Douglas G. (1972). Price Regulation and optimal service standards. The taxicab Industry.

[3] De Vany A. (1975). Capacity Utilization under Alternative Regulatory Restraints: An Analysis of Taxi Markets. Chicago Journals, The journal of Political Economy, 83(1): 83 – 94. https://doi.org/10.1086/260307

[4] Schroeter J. R. (1983). "A model of taxi service under fare structure and fleet size regulation". The Rand Company, The Bell Journal of Economics, Vol 14, No. 1: 81 – 96. https://doi.org/10.2307/3003539

[5] Manski, C. F. and Wright, J. D. (1976). Nature of equilibrium in the market for taxi services. Transportation Research Record 619: 296 – 306.

[6] Cairns R. D. and Liston-Heyes C. (1996). Competition and regulation in the taxi industry. Journal of Public Economics 59: 1 – 15. https://doi.org/10.1016/0047-2727(94)01495-7

[7] Arnott R. (1996). Taxi Travel Should Be Subsidized. Journal of Urban Economics 40: 316 – 333. https://doi.org/10.1006/juec.1996.0035

[8] Yang H. and Wong S. C. (1998). A network model of urban taxi services. Transport Research B, 32(4): 235 – 246. https://doi.org/10.1016/S0191-2615(97)00042-8

[9] Yang H., Yan Wing Lau, Sze Chun Wong and Hong Kam Lo (2000). A macroscopic taxi model for passenger demand, taxi utilization and level of services. Transportation 27: 317-340. https://doi.org/10.1023/A:1005289504549

[10] Yang h., Ye M., Tang W. H. And Wong S. C. (2005). Regulating taxi services in the presence of congestion externality. Transport Research A 39: 17 – 40. https://doi.org/10.1016/j.tra.2004.05.004

[11] Yang H., Fung C. S., Wong K. I. and Wong S. C. (2010). Nonlinear pricing of taxi services. Transport Research A 44: 337 – 348. https://doi.org/10.1016/j.tra.2010.03.004

[12] Bai Z. and Wang J. (2012). Equilibrium model of urban taxi services network based on the integrated service modes. Advanced Engineering Forum, 5: pp 82-87. https://doi.org/10.4028/www.scientific.net/AEF.5.82

[13] Cetin, T and K. Y. Eryigit (2011). "Estimating the Effects of Entry Regulation in the Istanbul Taxicab Market", Transportation Research Part A: Policy and Practice, 45(6), 2011, 476-484. https://doi.org/10.1016/j.tra.2011.03.002

[14] Cetin, T and K. Y. Eryigit (2013). The Economic Effects of Government Regulation: Evidence from the New York Taxicab Market", Transport Policy, 25, 2013, 169-177. https://doi.org/10.1016/j.tranpol.2012.11.011

[15] 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. https://doi.org/10.1016/j.sbspro.2011.08.020

[16] Holland, J.H.. Miller, J.H. (1991). Artificial Adaptive Agents in Economic Theory. American Economic Review 81 (2), 365-71.

[17] Kikuchi S., Rhee J. and Teodorovic D. (2002) Applicability of an agent-based modeling concept to modeling of transportation phenomena. Yugoslav Journal of Operations Research 12 (2), 141-156. https://doi.org/10.2298/YJOR0202141K

[18] Teodorovic D. (2003) Transport modeling by multi-agent systems: a swarm intelligence approach. Transportation planning and Technology 26 (4), 289-312. https://doi.org/10.1080/0308106032000154593

[19] Chen B. and Cheng H. H. (2010) A Review of the Applications of Agent Technology in Traffic and Transportation Systems. IEEE transactions on intelligent transportation systems vol. II (2) pp. 485 - 497. https://doi.org/10.1109/TITS.2010.2048313

[20] Bailey, W. A. and Clark, T. D. (1987) A simulation analysis of demand and fleet size effects on taxicab service rates. Proc. of the 19th conference on Winter simulation, 838-844. https://doi.org/10.1145/318371.318705

[21] Bailey, W. A. and Clark, T. D. (1992) Taxi management and route control: a systems study and simulation experiment. Proc. of the 24th conference on Winter simulation, 1217-1222. https://doi.org/10.1145/167293.167897

[22] Kim, H., Oh. J. D., and Jayakrishnan, R. (2005). Effect of taxi information system on efficiency and quality of taxi services. Transportation Research Record, Journal of the Transportation Research Board 1903, 96-104. https://doi.org/10.1177/0361198105190300111

[23] Song, Z. Q. and Tong, C. O. (2006) A simulation based dynamic model of taxi service. Proc. of Dynamic Traffic Assignment: First International Symposium on Dynamic Traffic Assignment.

[24] Song, Z. Q. (2006). A Simulation Based Dynamic Taxi Model. Master thesis at the University of Hong Kong.

[25] Chen Y. M. and Wang B. Y. (2009) Towards Participaroty Design of Multi-agent Approach to Transport Demands. DCSI International Journal of Computer Science Issues 4 (1) pp. 10 - 15.

[26] Lioris J. E., Cohen G., and La Fortelle A. (2010) Evaluation of Collecvite Taxi Systems by Discrete-Event Simulation. Second International Conference on Advances in System Simulation pp. 34-39. https://doi.org/10.1109/SIMUL.2010.18

[27] Shi W. (2010) Dynamic simulation and quantitative analysis of urban taxi services. Thesis for the degree of Master of Philosophy at the University of Hong Kong.

[28] Kim H., Yang I. and Choi K. (2011) An Agent-based Simulation Modelfor Analyzing the Impact of Asymetric Passenger Demand on Taxi Service. KSCE Journal of Civil Engineering 15 (1), 187-195. https://doi.org/10.1007/s12205-011-1037-8

[29] Cheng S. F. and Nguyen T. D. (2012) Taxisim: A Multiagent Simulation Platform for Evaluating Taxi Fleet Operations. AAMAS Workshops, LNAI 7068, 359-360. https://doi.org/10.1007/978-3-642-27216-5_26

[30] Salanova J.M., Estrada M. A., Mitsakis E. and Stamos I. (2013). "Agent Based Modeling for Simulation of Taxi Services". Journal of Traffic and Logistics Engineering (JTLE) (ISSN: 2301-3680), 1 (2)" 159 – 163. https://doi.org/10.12720/jtle.1.2.159-163

[31] Salanova Grau J., Estrada M. (2015). Agent based modelling for simulating taxi services. Procedia – Computer Sciences, Vol 52, pp 902-907. https://doi.org/10.1016/j.procs.2015.05.162

[32] Maciejewski M., Salanova J. M., Bischoff J., Estrada M. (2016). Large-scale microscopic simulation of taxi services. Berlin and Barcelona case studies. Journal of Ambient Intelligence and Humanized Computing 7(3), pp 1-9 https://doi.org/10.1007/s12652-016-0366-3

[33] Salanova Grau J.M., Moreira-Matias L., Saadallah Amal, Tzenos P., Aifandopoulou G., Chaniotakis E.m Estrada M. (2018) informed versus non-informed taxi drivers: agent-based simulation framework for assessing their performance, Transportation Research Board 96th Annual Meeting, Washington, USA. https://doi.org/10.1016/j.procs.2018.04.042

[34] Salanova J. M. (2013). Modeling of taxi cab fleets in urban environment. PhD thesis, Polytechnic University of Catalonia, BarcelonaTECH.

[35] Salanova J. M., Estrada M., Amat C. (2014). Aggregated modeling of urban taxi services. Procedia - Social and Behavioral Sciences (ISSN: 18770428), Vol 160, pp 352-361. [DOI:10.1016/j.sbspro.2014.12.147] https://doi.org/10.1016/j.sbspro.2014.12.147

[36] CENIT (2013). Observatori del Taxi 2012. Document Final. Institut Metropolità del Taxi.

[37] Salanova Grau, J.M., Estrada, M. (2017), Social optimal shifts and fares for the Barcelona taxi sector, Transport Policy, https://doi.org/10.1016/j.tranpol.2017.12.007

[38] Salanova J. M., Maciejewski M., Bischoff J., Estrada M., Tzenos P., Stamos I. (2017). Use of probe data generated by taxis. Big Data for regional science. Routledge Advances in Regional Economics, Science and Policy. Taylor & Francis Group. https://doi.org/10.4324/9781315270838-3