Planning for Electric Taxi Charging System from the Perspective of Transport Energy Supply Chain: A Data-Driven Approach in Beijing

Yinghao Jia[1], Huimiao Chen[2], Jiaoyang Li[3], Fang He[1], Meng Li[4], Zechun Hu[2], Zuo-Jun Max Shen[1, 5, 6]

[1]Department of Industrial Engineering
Tsinghua University
Beijing, China
(jiayh15@mails.tsinghua.edu.cn)

[2]Department of Electrical Engineering
Tsinghua University
Beijing, China

[3]Department of Automation
Tsinghua University
Beijing, China

[4]Department of Civil Engineering
Tsinghua University
Beijing, China

[5]Department of Industrial Engineering and Operations Research
University of California, Berkeley
Berkeley, CA, USA

[6]Department of Civil and Environmental Engineering
University of California, Berkeley
Berkeley, CA, USA

Abstract—**Administration in big cities is strongly promoting electric taxis (ETs) by providing purchasing subsidies, accessorial public facilities and many other encouraging policies. However, how to allocate the limited resources to optimize the benefits brought by ETs remains a headache for most researchers. Applying data mining technology, this research gathers real-time vehicle trajectory data of 39,053 urban conventional taxis (CTs) and 408 suburban ETs in Beijing for 4 weeks to extract the model of customers' travel demand and ET driving patterns. Based on the transport energy supply chain derived from Global Positioning System (GPS) data, we develop a data-driven method to design ET charging infrastructure in the near future.**

Keywords—**Electric taxis; transport energy supply chain; charging station planning; data-driven approach.**

I. INTRODUCTION

Taxi, an indispensable part of public transportation, is now facing trends of electrification. Considering the environmental effects of reducing traffic emissions and petroleum dependency, government agencies implement strong policies to promote the development of electric taxi (ET) fleet, such as deploying public charging infrastructure and forcing that newly added taxis must be electric vehicles [1].

However, how to allocate limited charging stations to optimize the operation of ETs has been a headache for a long time, as ETs' charging behavior is difficult to capture. The flow-capturing models [2]-[5] and network equilibrium models [6]-[9] were proposed, while unfortunately both these studies make assumptions of electric vehicle drivers' behavior, which may be inconsistent with the real world data.

Along with the development and popularization of real-time vehicle location systems using Global Positioning System (GPS) and wireless communication features, more and more researchers focus on applying data mining technology to recognizing taxi driving pattern and allocating charging stations [10], [11]. Reference [12] proposes a data-driven optimization-based approach to allocate chargers for battery electric vehicle (BEV) taxis throughout a city with the objective of minimizing the infrastructure investment. In this research, the service time is estimated based on the dwell time extracted from conventional taxi (CT) drivers' behavior. Authors of [13] present a simulation model to improve the electrification rate of the vehicle miles traveled (VMT) by taxi fleet in based on real-time vehicle trajectory data of over 46 thousand taxis in Beijing. The assumption of keeping CT drivers' driving pattern unchanged is again adopted.

Though the above research reveals various deep insights about charging station allocation problem, more or less, the assumption of keeping taxi drivers' driving pattern unchanged after electrified is kept to maximize the utility of CTs' trajectory data. However, due to the limited battery size and "range anxiety" of BEV[1], there shows significant difference between these two kind of taxis [14], [15]. Furthermore, the fundamental goal of taxi fleet is to serve customers' demand, which is different from commercial fleet and private vehicles repeating the same daily path and making the randomness deeply rooted in individual taxi driving pattern.

In this work, we apply data-driven technology to deal with ET charging infrastructure location problem from the perspective of transport energy supply chain system: viewing charging stations as transport energy suppliers and customers' mobility needs as the transport energy demanders, thus helping absorb the layer of ET and abandon the key assumption of ETs' behavior unchanged.

[1] To make it clear, the assumption of keeping driving pattern unchanged for plug-in-electric vehicles (PHEV) is relatively reasonable because the equipped internal combustion engine generator can provide electricity to the motor once the battery power is almost exhausted.

978-1-5386-2894-2/17/$31.00 ©2017 IEEE

978-1-5386-2894-2/17/$31.00 ©2017 IEEE

The main procedures and contributions of the paper are summarized below.

- Process the raw data, and extract customers' travel demand model from real-time vehicle trajectory data of 39,053 urban CTs.
- Analyze ET driving pattern based on the 408 suburban ETs GPS data together with 32 interviews and make comparisons with the pattern of CTs.
- Kick off the assumption of keeping taxis driving pattern unchanged after electrified based on the real-world data analysis, and build a matching model which directly bridges charging stations and customers' demand from the perspective of transport energy supply chain.
- Use P-median method to mathematically model the location problem and apply Kmeans of clustering technologies to give a heuristic approach. A numerical case study is provided in Beijing.

The remainder of the paper is organized as follows. Section II describes the GPS-based taxi travel data in Beijing and the processing method. In Section III, the data-driven ETs' pattern analysis is presented. In Section IV, the transport energy supply chain framework is invited while the mathematical model of ET charging station planning is formulated. Finally, Finally, conclusions are drawn in Section V.

II. Data-Driven Spatial and Temporal analysis of Customers' Demand

Research on customers' travel demand is also a main application field of taxi GPS data, through which we can get much knowledge about passenger origin-destination (OD) matrix. For example, reference [16] applies taxi GPS data to spatial-based clustering in order to discover the passenger demand patterns over time, and authors of [17] adopt L1-Norm support vector machine (SVM) as a feature selection tool to select the most salient feature patterns determining the taxi performance. However, in most papers, models are formulated based on plenty of assumptions about passengers' demands and drivers' behaviors, which may be inconsistent with the real world data.

Herein, for the purpose of better characterization of the comprehensive travel patterns of individual taxis, we examined the real-time vehicle trajectory data of 39,053 taxis in Beijing from May 4th 2016 to May 31st 2016, within the fifth ring area of Beijing (that is latitude between 116.2 and 116.55, while longitude between 39.75 and 40.03), collected by smartphone and on-board diagnostic (OBD). The dataset includes over 4.5 billion data points, which track each taxi's location and speed every 30 seconds. Table I shows one sample of the records in the dataset.

To clean up the raw data, we preprocess the GPS points that are missing, drifting and duplicated as well as extract key information from ocean of GPS points. Specifically, three kinds of problems may occur, i.e.:

- Missing data. Due to the bad performance network and signal delay, sometimes the data collectors fail to send GPS information to the cloud, leading to missing data. In the data preprocessing phase, we apply method of interpolation to supplement missing points in the trajectory. If the situation of missing data lasts for over 5 minutes, this trip is abandoned.
- Drifting data. Normally the GPS gives a relatively precise location and limits the error range within several meters, while sheltered and interfered by other buildings the GPS location may get drifted. Here we utilize GPS mapping technology to correct drifting points with the assistance of digital map and GIS system.
- Duplicated data. First we filter out the points whether they are resting points or duplicated data, according to the criteria that resting behaviors accompanied with "load" changing from 0 to 1 (See Table I). Then the duplicated data without "load" change can be deleted as it won't influence the trip extraction process.

By using record items like ID and Load (Load=1 means there is customer on the taxi at that time, and Load=0 otherwise), we segment continuous GPS points into individual trips, from which we only extract the essential OD information like pick-up and drop-off time and locations, as well as real travelling distance. Table II shows a OD trip sample, noting that timestep is the discretization of time with the interval of 1 minute. For instance, "timestep = 7" means the happening time is between 0:06 AM to 0:07AM on May 4th 2016.

TABLE I. Record Sample

ID	Time stamp	Speed(km/h)	Longitude	Latitude	Load
84471	201411120715	32	116.8198	40.3431	1

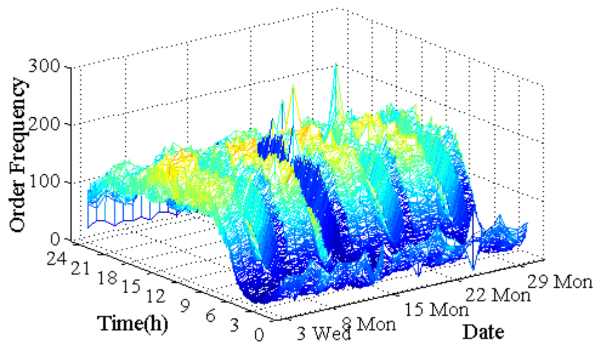

Fig. 1. Travel Demand Order Temporal Distribution.

TABLE II. OD TRIP SAMPLE

ID	Pick-up timestep	Pick-up longitude	Pick-up latitude	Drop-off timestep	Drop-off longitude	Drop-off latitude	Travel distance(km)
1	1	2	4	7	10	26	5.771420

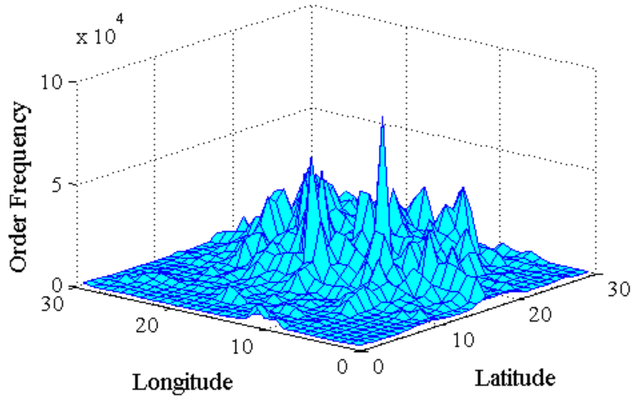

Fig. 2. Travel Demand Order Spatial Distribution.

Different from the irregular individual taxi driving pattern[2], the extracted OD trip data show the regularity in both time and space, shown in Figs. 1 and 2. Fig. 1 depicts the temporal pattern of customers' demands, where we can observe that: 1) Demand fluctuation follows distinct tendencies, such as daily or weekly cycle; 2) demand is much higher on weekdays compared with that in weekends; 3) taxi orders made in day time are significantly more than those in night time. Fig. 2 presents the travel demand order spatial distribution, where we can see that Chaoyang District, southwest area of Beijing, has obviously higher order frequency than others

To better describe the travel information of taxis, we regard some successive trips, where the previous trip's destinations and the next trip's origin are time- and space-adjacent, as a trip-chain [11].

III. ETs DRIVING BEHAVIOR ANALYSIS

Due to the limitation of battery range, charging speed and number of charging stations, there exists an obvious difference between the driving pattern of ETs and that of conventional ones, such as ETs spending more time in refueling, showing a tendency of driving around charging stations and getting charged frequently especially after finishing a long-distance trip. Reference [14] dealt with real ET GPS records to understand operational and charging patterns of ETs compared with those of internal combustion engine vehicles. Reference [15] reveals the fact that without global information, ET drivers tend to choose nearest charging stations.

To have a better understanding of ET travel patterns, our team did a survey in 2015 to investigate their driving patterns in suburbs of Beijing. During the survey, we investigated into 33 ET drivers and collected their questionnaire as well as analyzing GPS data of 408 ETs[3], in order to get a whole picture of driving behaviors. Nearly all the taxi drivers apply single-shift pattern, in contrast to double-shift pattern using by over 50% of CT drivers[4], due to the fact that ET needs a long-time (6-8 hours) charging, which is consistent with the fact that a large part of GPS data occurs around charging station, as shown in Fig. 3.

In addition to slow charging, we also noticed that quick charging is indispensable for ETs, for ET drives normally 120-160km a day, much more than the battery range, what gets worse due to the range anxiety. Thus ET drivers usually charges once or twice a day depending on how long they drive and individual range anxiety level (see Fig. 4). Range anxiety refers to the fact that an ET driver tends to get charged when the ratio of battery electricity left drops below certain anxiety level. Fig.4 shows that on average, an ET driver will feel anxious and go to charge when the battery ratio is below 0.221.

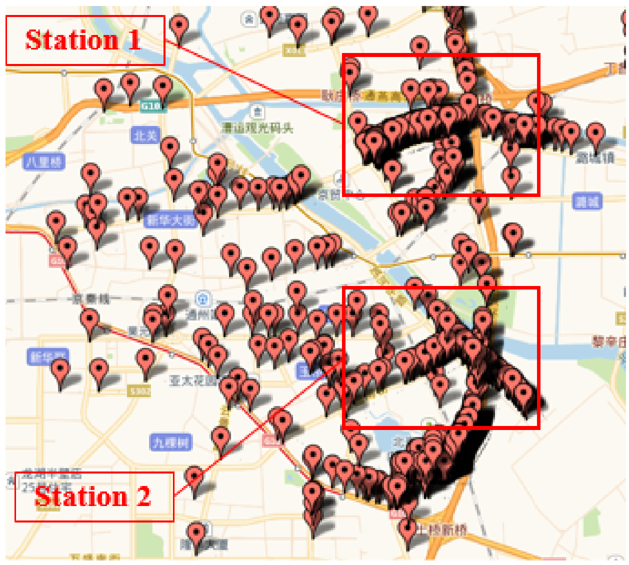

Fig. 3. An example of ET parking location distribution.

[2] In our previous work, the data-driven results show that the trajectory of a single taxi is widely distributed in roads of the city.

[3] Tongzhou, one of the biggest suburbs in Beijing, has a fleet of ETs consisting of 200 E150 put into use in November 2013 and 300 E200 in the end of 2014.

[4] Single-shift refers to the situation that each taxi is owned by one driver and only works in the daytime or nighttime in a day, while double-shift means two drivers take turns to run a taxi in a day.

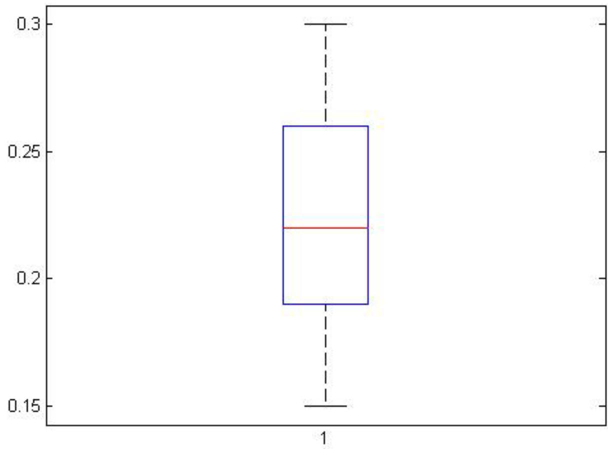

Fig. 4. Range anxiety distribution.

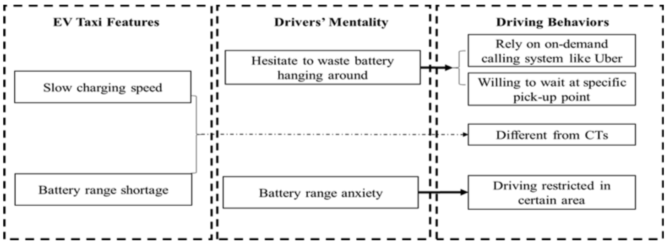

Fig. 5. ET driving pattern summary.

As for order receiving pattern, we found that the most popular way (average willing point [5] 3.94) to acquire passengers is using on-demand taxi calling system like Uber, which can significantly reduce battery consuming caused by hanging along the street. Another popular way (average willing point 1.76) is waiting at pickup points near charging stations. These findings are quite insightful in the following model, as shown in Fig. 5.

IV. Transport Energy Supply Chain Model

The basic idea of transport energy supply chain model relies on the fundamental function positioning of ET fleet, i.e. supplying transport service for customers' demand and getting charged at stations. Specifically, ETs acquire transport ability, in the form of battery energy, from charging stations. On the other hand, each customer's demand can be captured as traveling a distance from origin to destination and, during which consuming energy. From the perspective of transport energy supply chain model, as is shown in Fig. 7, it is obvious that the ET fleet acts as a retailer or porter, carrying energy from charging stations to customers' origin position. For this research we only consider part of supply chain inside red rectangle.[6]

What's more, Fig. 6 reveals that the supply chain structure of CT fleet is quite different from electric one, due to the essential difference in vehicle range and refueling frequency. The data analysis in Chapter 2&3 together with former researches shows, the limited battery range of ETs (E_{kj}) leads to the fact that they are closely bound with charging stations (C_k), normally occurring within service range (R_{ck}). On the contrary, CTs get refueled only several times a week and gas stations are already popular all over the place (the number of green in right dot is far more than blue dot in left), thus cutting the direct connections between CTs and gas stations (like what red cross does in the right picture). Inspired by the findings above, our initial model mainly captures the dominant supplier and end customers for electrified situation, choosing to omit the layer of ET fleet and bridging directly between charging stations and customers' demand to simply the location problem.

Here we set the objective function of optimizing charging infrastructure planning problem as to minimize the transport energy waste during trips between ETs and charging stations, as applying distributed charging stations location instead of centralized system is originally aimed at reducing the cost of taxis finding charging stations and leaving for orders. For more explanation, the energy to cover distance while delivering is not included in objective function, as it is determined by customers' demand and the location of charging infrastructure could not affect it.

This model is particularly suitable with the situation where ETs are restricted to operate in certain area, equipped with e-hailing service which can help customers make appointments. Adopting the assumption that each taxi only serves one trip-chain (consisting of several customers' trips of end-to-end) before getting charged at the assigned station, this model can be transformed to median-based location problem [18]. Fixed charge or p-median models[7] are to be selected corresponding to whether there exists constraint about capacity of each charging stations.

$$\min \sum_{j\in \mathbf{U}} \sum_{i\in \mathbf{U}} N_i d_{ij} Y_{ij} \tag{1}$$

subject to:

$$\sum_{j\in \mathbf{U}} Y_{ij} = 1, \forall i \in \mathbf{U} \tag{2}$$

$$Y_{ij} - X_j \le 0, \forall i, j \in \mathbf{U} \tag{3}$$

$$\sum_{j\in \mathbf{U}} X_j = p \tag{4}$$

$$X_j \in \{0,1\}, \forall j \in \mathbf{U} \tag{5}$$

$$Y_{ij} \in \{0,1\}, \forall i, j \in \mathbf{U} \tag{6}$$

In the above model, N_i is the number of OD points of trip-chains; d_{ij} is the distance between location i and location j; Y_{ij} is an assignment variable, which equals 1 if orders at location i is assigned to a charging station at location j, and 0 otherwise; X_j is a binary decision variable, which equals 1 if we locate a charging station at location j and 0 otherwise; p

[5] Willing point is a standard to know how much drivers would like to do in that way. Max is 5 standing for absolutely wanted, while minimum is 0 for "don't want to do".

[6] The influence of power plant and grid will be further studied in the proceeding researches.

[7] Here we assume that the charging station is incapacitated and ETs use fast charging technology, which makes it possible to fulfill a battery in 30 minutes.

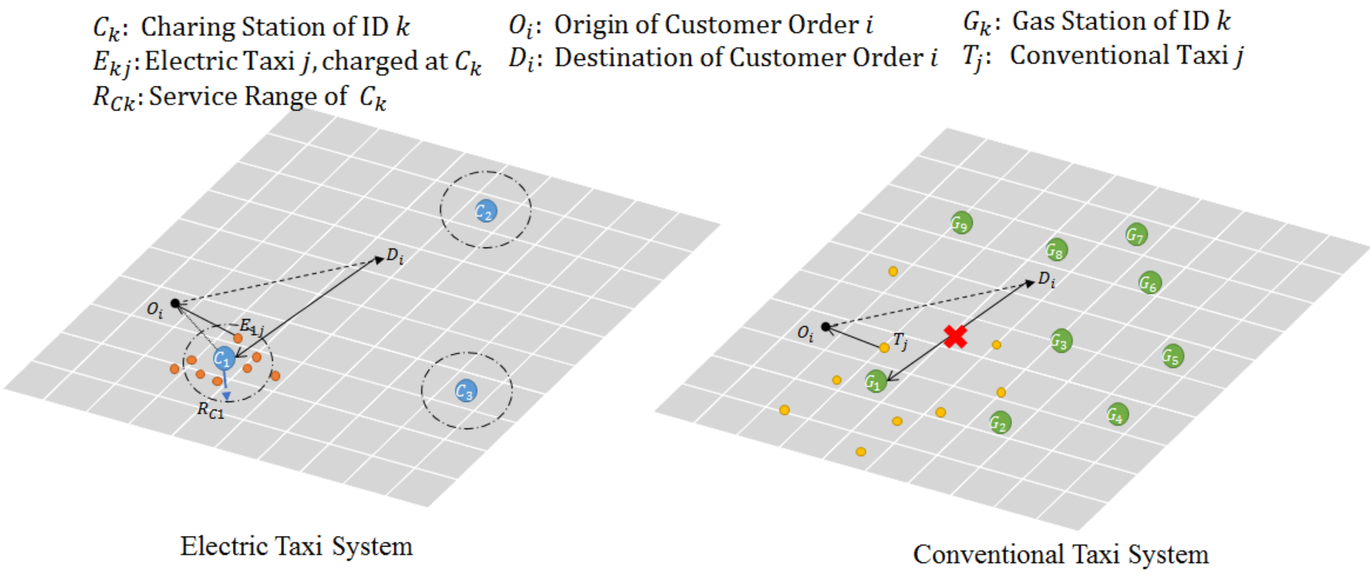

Fig. 6. Difference between ET system and CT system.

Fig. 7. Transport energy supply chain of ET system.

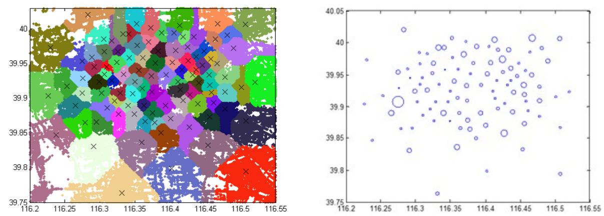

Fig. 8. Locations and capacity of charging stations.

is the number of charging station to be built; **U** is the set of locations in the concerned area of ET system. The objective function (1) minimizes the demand-weighted total distance, which is equivalent to the sum of transport energy waste during trips between ETs and charging stations. Constraints (2) stipulate that each node is assigned, while constraints (3) limit assignments of orders to a charging station. Constraint (4) states that p charging stations are to be located. Finally, constraints (5) and (6) are integrality constraints.

As the multi-source location problem is a NP-hard problem, heuristic methods are proposed, like Kmeans [13], which cluster all the trip-chain' OD (Origin and Destination) points and apply iteration algorithm to solve it. The principles behind are consistent with this transport energy supply chain system. Fig. 8 shows the result of locating 50 charging stations in Beijing urban area. On the right picture, the larger the circle is, the more OD is assigned to station. Charging capacity are allocated correspondingly. With the variable charging capacity in each charging station according to the potential demand of transport energy, the charging infrastructure can be better utilized to support the development of ET system.

V. Conclusion

In this paper, we present a planning method for ETs in urban environments from the perspective of transport energy supply chain. Inspired by real-world GPS trajectory data, this research abandons the widely-used assumption that keeping taxis driving pattern unchanged after electrified. Unlike private vehicles or commercial transport fleet, electric taxis' driving trajectory is essentially determined by customers' demand and shows irregularity and randomness for individual taxi driver. Thus we invite the concept of transport energy supply chain to review customers' order fulfillment process as the energy flow in electric taxi system. Further, we temperately make the link of ETs in this supply chain endogenous to simply the charging station location problems. Finally a p-median model is proposed and a case study in Beijing is solved by Kmeans clustering method.

VI. Acknowledgement

The authors appreciate the insightful comments and valuable suggestions offered by Sparkzone Institue, especially Wang Yang and Ziyang Guo.

The research is partially supported by grants from National Natural Science Foundation of China (71501107) and Beijing Natural Science Foundation (8173056).

REFERENCES

[1] Ministry of Environmental Protection of the People's Republic of China. (2016) [Online]. http://dqhj.mep.gov.cn/dtxx/201703/t20170323_408663.shtml, accessed Apr. 14, 2017.

[2] Berman, O., Larson, R.C., and Fouska, N., "Optimal location of discretionary service facilities," *Transp. Sci.* 26 (3), pp. 201–211, 1992.

[3] Hodgson, M.J., "A Flow-capturing location-allocation model," *Geograph. Anal*, 22 (3), 270–279, 1990.

[4] Hodgson, M. and Berman, O., "A billboard location model," *Geograph. Environ. Model*, 1, 25–45, 1997.

[5] Shukla, A., Pekny, J., and Venkatasubramanian, V., "An optimization framework for cost effective design of refueling station infrastructure for alternative fuel vehicles," *Comput. Chem. Eng.*, 35, 1431–1438, 2011.

[6] He, F., Wu, D., Yin, Y., Guan, Y., "Optimal deployment of public charging stations for plug-in hybrid electric vehicles," *Transp. Res. Part B: Methodol.*, 47 (1), 87–101, 2013.

[7] He, F., Yin, Y., and Zhou, J.. "Integrated pricing of roads and electricity enabled by wireless power transfer," *Transp. Res. Part C: Emerg. Technol.*, 34, 1–15, 2013.

[8] Dong, J., and Lin, Z., "Within-day recharge of plug-in hybrid electric vehicles: energy impact of public charging infrastructure," *Transp. Res. Part D: Transp. Environ.*, 17 (5), 405–412, 2012.

[9] Chen, Z., He, F., and Yin, Y, "Optimal deployment of charging lanes for electric vehicles in transportation networks," *Transp. Res. Part B: Methodol.*, 91, 344–365, 2016.

[10] Dong, J., Liu, C., and Lin, Z., "Charging infrastructure planning for promoting battery electric vehicles: an activity-based approach using multiday travel data.," *Transp. Res. Part C: Emerg. Technol.*, 38, 44–55, 2013.

[11] Cai, H., Jia, X., Chiu, A.S., Hu, X., and Xu, M., "Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet," *Transp. Res. Part D: Transp. Environ.*, 33, 39–46, 2014.

[12] Yang, Y., Yao, E., Yang, Z., and Zhang, R., "Modeling the charging and route choice behavior of BEV drivers.," *Transp. Res. Part C: Emerg. Technol.*, 65, 190–204, 2016.

[13] Li, M., Jia, Y., Shen, Z., and He, F., "Improving the electrification rate of the vehicle miles traveled in Beijing: A data-driven approach," *Transport. Res. A: Pol.*, 97, pp. 106-120, 2017.

[14] Tian, Zhiyong, *et al.*, "Understanding operational and charging patterns of electric vehicle taxis using GPS records," *IEEE 17th International Conference on Intelligent Transportation Systems*, 2014.

[15] Tian, Zhiyong, *et al.*, "Real-Time Charging Station Recommendation System for Electric-Vehicle Taxis," *IEEE Transactions on Intelligent Transportation Systems*, 17(11), pp. 3098-3109, 2016.

[16] L. Junghoon, S. Inhye, and G. Park, "Analysis of the passenger pick-up pattern for taxi location recommendation," *Fourth International Conference on Networked Computing and Advanced Information Management*, 2008.

[17] Li, B., Zhang, D., and Sun, L., *et al.*, "Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset," *IEEE International Conference on Pervasive Computing and Communications Workshops*, 2011.

[18] Daskin, M. S. (2008). What you should know about location modeling. *Naval Research Logistics (NRL)*, *55*(4), 283-294.