

11-2016

Battery capacity design for electric vehicles considering the diversity of daily vehicles miles traveled

Zhiheng Li
Tsinghua University

Shan Jiang
Tsinghua University

Jing Dong
Iowa State University, jingdong@iastate.edu

See next page for additional authors

Follow this and additional works at: https://lib.dr.iastate.edu/ccee_pubs

 Part of the [Civil Engineering Commons](#), and the [Transportation Engineering Commons](#)

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/ccee_pubs/201. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

Battery capacity design for electric vehicles considering the diversity of daily vehicles miles traveled

Abstract

In this paper, we study battery capacity design for battery electric vehicles (BEVs). The core of such design problems is to find a good tradeoff between minimizing the capacity to reduce financial costs of drivers and increasing the capacity to satisfy daily travel demands. The major difficulty of such design problems lies in modeling the diversity of daily travel demands. Based on massive trip records of taxi drivers in Beijing, we find that the daily vehicle miles traveled (DVMT) of a driver (e.g., a taxi driver) may change significantly in different days. This investigation triggers us to propose a mixture distribution model to describe the diversity in DVMT for various driver in different days, rather than the widely employed single distribution model. To demonstrate the merit of this new model, we consider value-at-risk and mean-variance battery capacity design problems for BEV, with respect to conventional single and new mixture distribution models of DVMT. Testing results indicate that the mixture distribution model better leads to better solutions to satisfy various drivers.

Keywords

Battery electric vehicle (BEV), Daily vehicle miles traveled (DVMT), Battery capacity design

Disciplines

Civil Engineering | Transportation Engineering

Comments

This is a manuscript of an article published as Li, Zhiheng, Shan Jiang, Jing Dong, Shoufeng Wang, Zhennan Ming, and Li Li. "Battery capacity design for electric vehicles considering the diversity of daily vehicles miles traveled." *Transportation Research Part C: Emerging Technologies* 72 (2016): 272-282. DOI: [10.1016/j.trc.2016.10.001](https://doi.org/10.1016/j.trc.2016.10.001). Posted with permission.

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Authors

Zhiheng Li, Shan Jiang, Jing Dong, Shoufeng Wang, Zhennan Ming, and Li Li

Battery Capacity Design for Electric Vehicles Considering the Diversity of Daily Vehicles Miles Traveled

Zhiheng Li^{1,2}, Shan Jiang¹, Jing Dong³, Shoufeng Wang¹, Zhennan Ming¹, Li Li^{1,4*}

1. Department of Automation, Tsinghua University, Beijing, China 100084

2. Graduate School at Shenzhen, Tsinghua University, Shenzhen, China, 518055

3. Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, Iowa 50010, USA

4. Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, 210096, P.R. China

Citation: Li, Z., Jiang, S., Dong, J., Wang, S., Ming, Z., & Li, L. (2016). Battery capacity design for electric vehicles considering the diversity of daily vehicles miles traveled. *Transportation Research Part C: Emerging Technologies*, 72, 272-282.

Abstract: In this paper, we study battery capacity design for battery electric vehicles (BEVs). The core of such design problems is to find a good tradeoff between minimizing the capacity to reduce financial costs of drivers and increasing the capacity to satisfy daily travel demands. The major difficulty of such design problems lies in modeling the diversity of daily travel demands. Based on massive trip records of taxi drivers in Beijing, we find that the daily vehicle miles traveled (DVMT) of a driver (e.g., a taxi driver) may change significantly in different days. This investigation triggers us to propose a mixture distribution model to describe the diversity in DVMT for various driver in different days, rather than the widely employed single distribution model. To demonstrate the merit of this new model, we consider value-at-risk and mean-variance battery capacity design problems for BEV, with respect to conventional single and new mixture distribution models of DVMT. Testing results indicate that the mixture distribution model better leads to better solutions to satisfy various drivers.

Keywords: battery electric vehicle (BEV), daily vehicle miles traveled (DVMT), battery capacity design

Manuscript received February 29, 2016, revised May 24, 2016, August 31, 2016, and September 27, 2016.

* Corresponding author. Email: li-li@tsinghua.edu.cn; Tel: +86 (10) 62782071

Mailing address: Room 806, Central Main Building, Tsinghua University, Beijing China 100084

1. Introduction

Electric vehicles (EV) become increasingly used in ground transportation (Tal *et al.*, 2014), because EV can reduce the air pollution in urban regions as well as fuel costs of drivers, because of the low cost of electricity relative to that of the conventional fuel. It was reported in PG&E (2009) that more than 70% of carmakers were now developing battery electric vehicles (BEVs) or plug-in hybrid electric vehicles (PHEVs).

Current BEVs still have some shortcomings. As pointed out by Pearre *et al.* (2011), one major problem of a BEV was its shorter travel distance comparing to that of a gasoline vehicle. Meanwhile, the recharging of a BEV is more time-consuming compared with the refilling of a gasoline car. This problem prevents drivers who care about travel convenience from purchasing BEVs. One possible solution to this problem is to design an appropriate battery capacity for a BEV so that this capacity can cover daily travel distances of most drivers so that drivers only need to recharge their BEVs at home or workplaces. Clearly, the feasibility of this solution lies in an appropriate demand analysis and supply design.

On the demand side, researchers collected empirical driving data to investigate the energy consumption and charging patterns of drivers. For example, GPS equipment was installed on a number of BEVs and internal combustion engine vehicles (ICEVs) to collect the necessary information (e.g., timestamp, latitude, longitude and velocity) to determine energy consumption patterns (Greaves *et al.*, 2013; Dong and Lin, 2014; Wu *et al.*, 2015a; Wu *et al.*, 2015b). Based on GPS data, vehicles' charging patterns and VMT between two consecutive charging were obtained and studied in Wu *et al.* (2015b); Dong and Lin (2014). Recently, the relationship between energy consumption and driving styles also received increasing interests (Tang *et al.*, 2015a; Tang *et al.*, 2015c; Li *et al.*, 2015; Li *et al.*, 2016; Chen *et al.*, 2016).

On the supply side, researchers analyzed the proper capacity of BEVs to satisfy daily vehicle miles traveled (DVMT) of most drivers¹. For example, Pearre *et al.* (2011); Traut *et al.* (2013); Franke and Krems (2013); Greaves *et al.* (2013) had investigated whether the mile range of a certain type BEV can satisfy the current DVMT demand of the vehicle. Dong and Lin (2014) presented a probability model to describe the BEV feasibility. Lin (2014) developed an optimization framework to design BEVs, in which the objective reflected a trade-off between the battery's price and DVMT demand.

To balance the demand and supply to satisfy most drivers, most existing studies focused on modeling the diversity in DVMT and designing BEVs with respect to such diversity. There are three kinds of models for DVMT in existing approaches (Greene, 1985), summarized as follows:

¹ We decide to keep the United States customary units, since most literatures that we cited in this field adopted the United States customary units.

The first kind of models assumes that DVMT is single-value. That is, such models assumed that DVMTs of drivers were the same and only the averaged value was used to characterize the DVMT of all drivers all the time.

The second kind of models assumes that DVMTs of drivers are different, but the DVMT of a driver is relatively constant for different days. Hence, such models use a single distribution to characterize the diversity of drivers (Liu *et al.*, 2012; Tang *et al.*, 2015b; Leng *et al.*, 2016). However, empirical observations show that the DVMT of a driver can be significantly different from day to day. So, such models cannot fully describe the uncertainty in travel demands of drivers and are thus unable to accurately measure satisfaction degrees of different drivers.

The third kind of models aims to characterize both the diversity of DVMTs for different days and the diversity of DVMTs for different drivers (Pearre *et al.*, 2011; Lin *et al.*, 2012; Kontou *et al.*, 2015; McCollum *et al.*, 2016). For example, a special distribution model was proposed in Tamor and Milačić (2015a) and Tamor *et al.* (2015b). It assumes that the DVMT of a driver can be described by the mixture of a negative exponential distribution and a normal distribution. Moreover, the key parameters of this mixed distribution for different people can be described by a log-logistic distribution. This hybrid distribution model of DVMT were tested in several regions. However, we still do not know whether this hybrid model works for other regions.

To solve this problem, we propose a new mixture distribution model to describe the diversity of DVMTs for various drivers in different days. It is slightly different from the two-dimensional distribution models that had been proposed in Bhat and Eluru (2009); Chen *et al.* (2010) to characterize the uncertainty of traffic flow measures. Rather than consider two factors (drivers and days) in an equivalent and exchangeable position, this mixture distribution model can be viewed as a hierarchical distribution model, in which we first characterize drivers by their average daily vehicle miles traveled (ADVMT) and then further describe their diversities in different days.

Different from the parametric model proposed in Tamor and Milačić (2015a) and Tamor *et al.* (2015b) which has a fixed number of parameters to calibrate, this mixture distribution DVMT model is a non-parametric model which has some merits. First, its estimation process is quite straightforward and easy. Second, it makes no assumptions about the probability distributions of the variables being assessed and thus can be used to describe any kind of drivers in all regions.

Thus, it provides us a unique and convenient tool for further analysis. To demonstrate its benefit, we consider value-at-risk and mean-variance battery capacity design problems for taxi BEVs, with respect to conventional single and new mixture distribution models of DVMT. Noticing that municipal governments of many cities had enacted a series of regulations and policies to promote the development of new energy vehicles (Zou *et al.*, 2016), we believe the design of taxi BEVs will attract increasing attentions. Tests indicate that the mixture distribution model better characterizes the diversity of taxi drivers and thus leads to better solutions to satisfy most of them.

To give a detailed presentation of our finding, the rest of this paper is organized as follows. *Section 2* presents the mixture distribution model for DVMT. *Section 3* presents optimization models to design a proper BEV capacity with respect to this new DVMT model. *Section 4* provides some numerical tests based on taxis travel data collected in Beijing. The difference between single and mixture DVMT distributions based optimization models are highlighted. Finally, *Section 5* concludes the paper. The nomenclature list of this paper can be found in *Appendix A*.

2. The Mixture Distribution DVMT Model

In this paper, we assume that the DVMT of a driver in a particular day is a random variable. It follows a mixture distribution that is characterized by the kind of driver he/she belongs to and the associated DVMT distribution function of this kind of drivers. More precisely, suppose we have finite n kinds of drivers, where the proportion of the i th kind of drivers is p_i , $i = 1, \dots, n$, satisfying

$$\sum_{i=1}^n p_i = 1 \quad (1)$$

Notice that the DVMT, denoted by x of the i th kind of drivers, may vary from day to day, we can use a one-dimensional probability density function $f_i(x)$ to describe the diversity (uncertainty) of DVMT for the i th kind of drivers. So, the probability density function of the daily vehicle miles traveled (DVMT, denoted by x) of any a driver is (Lindsay, 1995)

$$f(x) = \sum_{i=1}^n p_i f_i(x) \quad (2)$$

In practice, we will first categorize drivers into different kinds according to ADVMT and then estimate the DVMT distribution function for each kind of drivers, based on empirical observations. This trick makes a trade-off between modeling details and model complexity.

Suppose we preselect the number of driver kinds as n . Then, we compute the ADVMT for each driver and label the largest ADVMT value observed as A_{\max} . The drivers whose ADVMT values fall in $[(i-1) \cdot A_{\max} / n, i \cdot A_{\max} / n)$ will be taken as the i th kind of drivers, $i = 1, \dots, n$. The numbers of each kind of drivers will be summed up to calculate the proportions p_i of each kind of drivers. Finally, we use a discrete distribution to describe and estimate the one-dimensional probability density function $f_i(x)$ of DVMT for each kind of drivers. Usually, the raw data are discretized with a step size of 1 mile to estimate the discrete distributions. So, we do not fit any existing distribution models.

Fig.1 gives an illustration of the mixture distribution of DVMT obtained from the taxis data collected in Beijing, China, where we group the drivers into 9 kinds. As shown in Fig.1(b) and Fig.1(c), the DVMT distributions of different kinds of drivers can be significantly different. This

indicates the need of considering the diversity (uncertainty) of drivers' daily travel demands in the BEV batter capacity design.

place Fig. 1 about here.

Fig.2 provides an illustration of the relationship between the single distribution and the mixture distribution models of DVMT, using the same data. Indeed, if we neglect the diversity of DVMT for a kind of drivers in different days, we can just use the average DVMT values to distinguish different kinds of drivers. Thus, we can degenerate to the conventional single distribution model of DVMT which only assumes various drivers may have different DVMT. The major difference between the single distribution model of DVMT shown in Fig.2(b) and the existing single distribution models of DVMT (Dong and Lin, 2014) is that we consider much more kinds of driver than existing studies.

place Fig. 2 about here.

It should be pointed out that the selection of the number of driver kinds may significantly influence the shape of the mixture distribution of DVMT. For most design objectives (e.g., the two design objectives considered in *Section 3*), we can select an appropriate number of drivers to keep an appropriate balance between model feasibility and model complexity (Pearre *et al.*, 2011). In other words, using the new model, we can reach a satisfactory solution, without resorting to brute force tests. We will discuss the selection strategy in *Section 4*.

3. BEV Capacity Design Considering Mixture Distribution DVMT Model

In this paper, we focus on battery capacity design problems for BEVs, where the maximum travel distance is the decision variable. Our objective is to guarantee that the generalized battery capacity can fulfill satisfaction degrees of most BEV drivers (Lin, 2014; Zhang *et al.*, 2016).

One major difficulty lies in how to depict satisfaction degrees of various drivers with respect to uncertainty of DVMT. As pointed out in Rockafellar *et al.* (2000); Yin (2008); Markowitz (2014), there were various methods to deal with the influence of uncertainty in decision. In this paper, we choose two representative methods to model and control the uncertainty, namely Value-at-Risk analysis (Rockafellar *et al.*, 2000; Rockafellar *et al.*, 2002; Sarykalin *et al.*, 2008) and Mean-Variance analysis (Markowitz, 1952; Markowitz, 1987; Kolm *et al.*, 2014; Markowitz, 2014). We will discuss these two methods in the following subsections, respectively.

3.1 Value-at-Risk Design

Our first objective is: find a minimum BEV capacity (in terms of miles that can travel) r to satisfy the travel demands for at least β drivers; meanwhile, we say that a driver can be satisfied, if r is larger than α of his/her DVMT. Here $\alpha \in [0,1]$, $\beta \in [0,1]$ are two confidence levels.

Regarding the mixture distribution DVMT model (1)-(2), the Value-at-Risk design problem can be written as

$$\begin{aligned} \min_{r \in [0, A_{\max}]} \quad & r \\ \text{s.t.} \quad & \sum_{i=1}^n p_i \cdot \text{Heav}(r - VaR_{\alpha}^i) \geq \beta \end{aligned} \quad (3)$$

where $\text{Heav}(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$ is the Heaviside step function. VaR_{α}^i denotes the level α -quantile of the DVMT for the i th kind of drivers.

This objective belongs to the Value-at-Risk (VaR) design which formulates a measure of the risk (Rockafellar *et al.*, 2000; Rockafellar *et al.*, 2002; Sarykalin *et al.*, 2008). It estimates how much we might lose, for a given portfolio and a given probability. Clearly, when both α and β are approaching 1, we can say the designed BEV capacity r can satisfy most drivers in most days. In other words, this objective assumes that drivers will mainly require the battery capacity to meet their daily travel demands, without considering the price of battery.

In existing VaR design for BEV capacity where single distribution DVMT models are used, we only need to consider one confidence level which depicts the portion of satisfied drivers. We assume that the DVMT of a specific driver is constant on different days. However, when mixture distribution DVMT model is considered here, we need to set two confidence levels to examine the portion of drivers who had been satisfied in a given percentage of days.

place Fig. 3 about here.

Fig.3(a) illustrates the distribution of DVMT of each kind of drivers, where the red curve represents the BEV capacity and the blue curve represents the level α -quantile (denoted by VaR_{α}) of the DVMT for each kind of drivers (Rockafellar *et al.*, 2000). Fig.3(b) plots the probability density plot of the DVMT of the 6th kind of drivers whose ADVMT falls in the range of [125, 150] miles, where the area of the green part is α , the red part is the DVMT which cannot be satisfied by the BEV capacity. We judge whether the i th kind of drivers are satisfied by comparing the VaR_{α}^i of DVMT and BEV capacity. If the capacity is larger than the level α -quantile (denoted by VaR_{α}^i) of the DVMT for the i th kind of drivers, this kind of drivers are satisfied. The proportion of the satisfied drivers is required to be larger than β . Fig.3(c) gives the

corresponding box plot for mixture DVMT distribution, where the upper bound of the box is the 95% percentile of DVMT, the lower bound of the box is 0% percentile of DVMT. Particularly, for the 6th kind of drivers, the green part stands for the portion of satisfied drivers and red part stands for the portion of unsatisfied drivers.

To numerically solve this optimization problem, we can apply the following sorting and sequential searching strategy. First, we calculate the minimum capacity for each kind of drivers such that this capacity is larger than VaR_α of the DVMT of this kind of drivers. More precisely, for the i th kind of drivers, we labeled this minimum capacity as r_i , $i = 1, \dots, n$. Then, we sort r_i and find the minimum r which makes that the proportion of the satisfied drivers is larger than β by a linear search (Floudas and Pardalos, 2009). Here, the meaning of β can be easily explained as the percentage of drivers that can be satisfied by given the range. For example in Fig.3(c), the first 6 kinds of drivers are satisfied, so the percentage of drivers belonging to these 6 kinds is β .

3.2 Mean-Variance Design

The VaR design does not consider the fact that drivers hope the price of battery should be as cheap as possible. To find a good balance between battery capacity and battery price, we assume the objective function S of a particular driver can be described in terms of drivers' costs as

$$S = S^p + S^f \quad (4)$$

where S^p is the battery price and S^f denotes the cost of failing to meet the travel demand. Lin (2014) also considered the electricity cost, which we neglect here so as to focus on the influence of mixture distribution DVMT model.

On one hand, the price of battery increases with the capacity of battery. Adopting the formula proposed in Lin (2014), we define S^p in terms of the battery capacity r as

$$S^p = 2130 \cdot r \cdot (0.0004r + 0.2839) \cdot r^{-0.317} \quad (5)$$

On the other hand, let us assume the DVMT of this driver is a random number denoted as x . Similar to Lin (2014), we calculate the expectation cost of failing to meet the travel demand as

$$E\{S^f\} = \int_r^{A_{\max}} (32685.75 + 280.95x) \cdot f(x) dx \quad (6)$$

where we always choose $r < A_{\max}$, and the variance of the cost as

$$\text{Var}\{S^f\} = \int_r^{A_{\max}} (32685.75 + 280.95x)^2 \cdot f(x) dx - [E\{S^f\}]^2 \quad (7)$$

where $f(x)$ is the probability density function of the DVMT distribution.

The origins and details of the above formulas can be found in Lin (2014). It must be pointed out that the coefficients in Eqs.(5)-(7) were estimated from BEV markets in the United States by

Lin (2014). The coefficients should be re-estimated, when we applied this model for other counties in practice. However, we directly adopt the formulas given in Lin (2014) to model the need of Beijing taxi drivers in this paper, since we do not have such data available for China to calibrate the model parameters for Beijing taxis. This certainly causes a comparability problem. However, we believe US and China drivers face a similar situation when considering the price and cost of EV, so that the model proposed by Lin (2014) is still applicable. Since our main theme is to demonstrate the importance of considering the diversity of DVMT, this data comparability problem can be tolerated.

Our mean-variance type objective can then be written as

$$\begin{aligned} & \min_{r \in [0, A_{\max}]} E\{S\} + \lambda \sqrt{\text{Var}\{S\}} \\ \Leftrightarrow & \min_{r \in [0, A_{\max}]} S^p + E\{S^f\} + \lambda \sqrt{\text{Var}\{S^f\}} \end{aligned} \quad (8)$$

where λ is the preselected weighting coefficient.

Problem (8) can be used to examine either the overall objective function of all drivers by considering $f(x) = \sum_{i=1}^n p_i f_i(x)$ or the objective function of the i th kind of drivers by considering $f_i(x)$ instead of $f(x)$.

place Fig. 4 about here.

As the battery capacity r increases, the battery price S^p increases and the cost of failing to meet the travel demand S^f decreases. Fig.4(a) show the one-dimensional distribution of DMVT for the 6th kind of drivers and Fig.4(b) gives the variation of the objective function $S^p + E\{S^f\} + \lambda \sqrt{\text{Var}\{S^f\}}$ with respect to the battery capacity r . The weighting parameter λ of the mean-variance objective function is set as 0.2. We can find that there should exist an appropriate battery capacity that minimizes the objective function.

Since Problem (8) is a bounded single-variable optimization problem, we can enumerate all the possible solutions (e.g., with a step size 1 mile) to find the global optimal solution (Floudas and Pardalos, 2009).

4. Numerical Tests

4.1 Testing Data and Discussion on the Number of Driver Kinds

In this paper, we take taxi services in Beijing as an illustrative example to explain how to design battery capacity for electric taxis (Zou *et al.*, 2016). This is mainly because that the variation of DVMTs in different days are usually more significant for a taxi driver than that for a

commuting traveler. However, it must be pointed out that there are also many fleet service drivers (e.g., home movers, express mail carriers, water delivery workers) whose DVMTs vary notably from day to day. Our model is also suitable for such drivers.

Our dataset contains 3,000,000 taxi trip records of over 30,000 taxis in June, 2014, in Beijing. From this dataset, we can retrieve the traveling distance of every taxi per day. Since the raw dataset contains some abnormal data (some drivers run more than 300 miles per day), we simply filter out all the DVMT data that are larger than 250 miles. This certainly discards some normal data. Noticing the ratio of all the discarded data is less than 0.1%, this slight difference can be omitted.

As mentioned above, the selection of the number of driver kinds influences the shape of the mixture DVMT distribution. Fig.5 shows two mixture DVMT distributions with 9 and 18 kinds of drivers. We can see that the general shape of the mixture DVMT distributions roughly remains.

The definition of mixture distribution in Eq.(1)-(2) makes an implicit assumption that all drivers with similar ADVMT have similar distribution of DVMT. To verify this assumption, we use the Mann-Whitney U test.

We first assume all the drivers in one kind to follow the same distribution of DVMT and estimate this overall distribution using the samples of all drivers in all days. Then, we calculate the distribution of DVMT of each driver using the samples of this driver in all days. Finally, the distribution of DVMT for each driver is compared with the overall distribution of DVMT to check how many drivers have the same distribution with the overall distribution.

place Fig. 5 about here.

We introduce the following ratio to denote how many drivers belonging to the j th kind can be viewed to follow the same distribution of the j th kind of drivers.

$$\rho'(j) = \frac{\sum_{i=1}^{N_j} ranksum_fit(i, j)}{N_j} \quad j = 1, 2, 3 \dots K \quad (9)$$

where N_j is the number of the j th kind of drivers. The indicator function $ranksum_fit$ is defined as

$$ranksum_fit(i, j) = \begin{cases} 1 & \text{test indicates a rejection of the null hypothesis} \\ 0 & \text{test fails to reject the null hypothesis} \end{cases} \quad (10)$$

where the null hypothesis is that the DVMT distribution of i th driver is same as the j th kind of drivers.

Moreover, we define the following overall ratio to denote how many drivers with similar ADVMT have similar distribution of DVMT.

$$\rho(K) = \frac{\sum_{j=1}^K \sum_{i=1}^{N_j} \text{ranksum_fit}(i, j)}{N} \quad (11)$$

where N is the number of all drivers, satisfying $\sum_{j=1}^K N_j = N$.

Fig.6 below gives the testing result, where drivers are categorized in 300 kinds. We can see that most drivers in each kind follows a similar distribution.

place Fig. 6 about here.

4.2 Value-at-Risk Design

In this section, we compare the VaR design of BEV capacity, based on the single distribution DVMT model (Dong and Lin, 2014) and the new mixture DVMT model.

When we adopt the mixture DVMT model, we require 90% drivers be satisfied in at least 90% days. That is, we assume $\alpha = 0.9$ and $\beta = 0.9$ here. It should be pointed out that we use 90%-quantile mainly as an illustrative example here. When we increase α approaching 1, solutions found by the single distribution DVMT model and mixture distribution DVMT model would become close.

We first study the influence of the number of driver kinds in Fig.6, where we vary the number of driver kinds from 1 to 200 with a step size 1. We can see that the desired BEV capacity becomes stable at about 202 miles, when the number of driver kinds reaches at 100. So, we choose 202 miles as our solution when the mixture DVMT distribution model is adopted. Further analyzing the travel distance of every driver in all days, we can find that actually 88.72% drivers are satisfied with this battery capacity. This small discrepancy is caused by grouping approximation and can be tolerated usually.

Fig.7 illustrates the influence of the number of driver kinds. The DVMT model with the highest accuracy must take every individual driver as a special kind to construct the mixture distributions. However, this overelaborated model will consume the highest computation costs and is usually unnecessary. Testing results in Fig.7 show that the number of driver kinds can be chosen as small as 0.3% of the amount of drivers to achieve a good enough solution whose deviation to the global optimal solution is less than 1mile. This indicates the feasibility of grouping approximation to keep an appropriate balance between model feasibility and model complexity. Usually, we do not need to use such brute force test for every individual driver.

place Fig. 7 about here.

When we adopt the single distribution DVMT model, we directly get the battery capacity as the $\text{VaR}_{0.9}$ of the distribution. Particularly, the solution is 176 miles; see Fig.8(a). Fig.8(b) compares the two solutions. Further analyzing the travel demand of every driver in all days, we can find that, only 71.05% drivers can be satisfied with such a battery capacity in 90% days of their travels, if the BEV capacity is chosen as 176 miles. So, assuming a single distribution DVMT model is not accurate enough to serve the users.

place Fig. 8 about here.

4.3 Mean-Variance Design

In this section, we compare the mean-variance design solution for all the drivers and for each kind of drivers. Here, we group drivers into 100 kinds according to their ADVMT. The weighting parameter λ of mean-variance objective function is set as 0.2. The optimal battery capacity for all the drivers is 222 miles. However, as shown in Fig.9, the optimal battery capacities of the first 40 kinds of drivers are much smaller than 222 miles. Clearly, the DVMT models which cannot fully describe the uncertainty in travel demands of various drivers in different days, are unable to accurately measure the satisfaction degrees of distinct drivers.

The mixture distribution DVMT model provides a chance to solve this problem. However, it requires a dedicate paper to design a new type of objective to consider the purposes for different drivers, based on the new mixture distribution DVMT model. We will submit another paper to discuss this topic soon.

place Fig. 9 about here.

5. Conclusions

In this paper, we study the diversity of travel demands of different drivers in different days and its influence on designing a proper battery capacity for taxi BEVs. We propose a concise mixture distribution model of DVMT, in which we first characterize drivers by ADVMT and then describe the diversity in DVMT of drivers in different days. This trick makes a trade-off between modeling details and model complexity. Tests show that the number of driver kinds can be chosen as small as 0.3% of the amount of sampled drivers to achieve a good enough design solution. So, we do not need to use such a brute force test for individual drivers.

We test the new model incorporating two widely used design objectives that consider uncertainty: Value-at-Risk design and Mean-Variance design. The new model is shown to be a convenient tool to be incorporated with many design objectives. The numerical calculations are fast and straightforward. Results indicate that the introduction of the new model helps design a better battery capacity to satisfy most drivers' travel demands.

Whether the proposed model only fits for taxi drivers remains to be fathomed. We are currently collecting more data from more drivers to check the applicability of the proposed model. Besides, it must be pointed out that personal driving styles also influence energy consumption of each vehicles (Tang *et al.*, 2015a; Tang *et al.*, 2015c; Li *et al.*, 2015; Li *et al.*, 2016; Chen *et al.*, 2016). However, current DVMT models cannot be used describe this influence. The lack of detailed naturalistic driving data and the associated energy consumption data prevents further discussions here. Noticing the importance of driving styles, we plan to carry out related studies in the near future.

Besides, the above findings also indicate that possible improvement could be achieved in charging station location optimization (Dong *et al.*, 2014; He *et al.*, 2015) and other EV related designs, if we consider the heterogeneity of drivers' travel demand in different days.

Acknowledgements

This work was supported in part by the National Science Foundation of China under Grand 91520301.

Appendix A.

place Table A.1 about here.

References

- Bhat, C. R., Eluru, N. (2009) A copula-based approach to accommodate residential self-selection effects in travel behavior modeling, *Transportation Research Part B: Methodological*, **43**(7), 749-765.
- Chen, X., Li, L., Zhang, Y. (2010) A Markov model for headway/spacing distribution of road traffic, *IEEE Transactions on Intelligent Transportation Systems*, **11**(4), 773-785.
- Chen, Z., Xiong, R., Wang, C., Cao, J. (2016) An on-line predictive energy management strategy for plug-in hybrid electric vehicles to counter the uncertain prediction of the driving cycle, *Applied Energy*, <http://www.sciencedirect.com/science/article/pii/S0306261916300587>
- Dong, J., Lin, Z. (2014) Stochastic modeling of battery electric vehicle driver behavior, *Transportation Research Record*, **2454**(1), 61-67.
- Dong, J., Liu, C., Lin, Z. (2014) Charging infrastructure planning for promoting battery electric vehicles: An activity-based approach using multiday travel data, *Transportation Research Part C: Emerging Technologies*,

- Floudas, C. A., Pardalos, P. M. (2009) eds., *Encyclopedia of Optimization*, 2nd edition, Springer.
- Franke, T., Krems, J. F. (2013) What drives range preferences in electric vehicle users? *Transport Policy*, **30**(3), 56-62.
- Greene, D. L. (1985) Estimating daily vehicle usage distributions and the implications for limited-range vehicles, *Transportation Research Part B: Methodological*, **19**(4), 347-358.
- Greaves, S., Backman, H., Ellison, A. B. (2013) An empirical assessment of the feasibility of battery electric vehicles for day-to-day driving, *Transportation Research Part A: Policy and Practice*, **66**, 226-237.
- He, F., Yin, Y., Zhou, J. (2015) Deploying public charging stations for electric vehicles on urban road networks, *Transportation Research Part C: Emerging Technologies*, **60**, 227-240.
- Kolm, P. N., Tutuncu, R., Fabozzi, F. J. (2014) 60 Years of portfolio optimization: Practical challenges and current trends, *European Journal of Operational Research*, **234**(2), 356-371.
- Kontou, E., Yin, Y., Lin, Z. (2015) Socially optimal electric driving range of plug-in hybrid electric vehicles, *Transportation Research Part D: Transport and Environment*, **39**, 114-125.
- Leng, B., Wang, J., Li, L., Xiong, Z. (2016) Analysis of taxi drivers' behaviors within a battle between two taxi apps, *IEEE Transactions on Intelligent Transportation Systems*, **17**(1), 296-300.
- Li, L., Yang, C., Zhang, Y., Zhang, L., Song, J. (2015) Correctional DP-based energy management strategy of plug-in hybrid electric bus for city-bus-route, *IEEE Transactions on Vehicular Technology*, **64**(7), 2792-2803.
- Li, L., Yan, B., Yang, C., Zhang, Y., Chen, Z., Jiang, G. (2016) Application oriented stochastic energy management for plug-in hybrid electric bus with AMT, *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4459-4470, 2016.
- Lin, Z., Dong, J., Liu, C., Greene, D. L. (2012) Estimation of energy use by Plug-In Hybrid Electric Vehicles: Validating Gamma distribution for representing random daily driving distance, *Transportation Research Record*, **2287**, 37-43.
- Lin, Z. (2014) Optimizing and diversifying electric vehicle driving range for U.S. drivers, *Transportation Science*, **48** (4), 635-650.
- Lindsay, B. G. (1995) *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 5, Institute of Mathematical Statistics, Hayward, CA, USA.
- Liu, Y., Kang, C., Gao, S., Xiao, Y., Tian, Y. (2012) Understanding intra-urban trip patterns from taxi trajectory data, *Journal of Geographical Systems*, **14**(4), 463-483.
- Markowitz, H. (1952) Portfolio selection, *Journal of Finance*, **7**(1), 77-91.
- Markowitz, H. (1987) *Meanvariance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, Cambridge, MA.
- Markowitz, H. (2014) Mean-Variance approximations to expected utility, *European Journal of Operational Research*, **234**(2), 346-355.
- McCollum, D. L., Wilson, C., Pettifor, H., Ramea, K., Krey, V., Riahi, K., Bertram, C., Lin, Z., Edelenbosch, O. Y., Fujisawaa, S. (2016) Improving the behavioral realism of global integrated assessment models: An application to consumers' vehicle choices, *Transportation Research Part D: Transport and Environment*, <http://www.sciencedirect.com/science/article/pii/S1361920915300900>
- PG&E, (2009). The perfect storm for electric vehicle market growth in California. Smart Grid Workshop. California Public Utilities Commission.
- Pearre, N. S., Kempton, W., Guensler, R. L., Elango, V. V. (2011) Electric vehicles: how much range is required for a day's driving? *Transportation Research Part C: Emerging Technologies*, **19**(19), 1171-1184.
- Rockafellar, R. T., Uryasev, S. P. (2000) Optimization of conditional value-at-risk, *Journal of Risk*, **2**, 21-42.
- Rockafellar, R. T., Uryasev, S. (2002) Conditional value-at-risk for general loss distributions, *Journal of Banking and Finance*, **26**(7), 1443-1471.

- Sarykalin, S., Serraino, G., Uryasev, S. (2008) Value-at-risk vs. conditional value-at-risk in risk management and optimization, *Tutorials in Operations Research*.
- Tamor, M. A., Milačić, M. (2015a) Electric vehicles in multi-vehicle households, *Transportation Research Part C: Emerging Technologies*, **56**, 52-60.
- Tamor, M. A., Moraal, P. E., Reprogle, B., Milačić, M. (2015b) Rapid estimation of electric vehicle acceptance using a general description of driving patterns, *Transportation Research Part C: Emerging Technologies*, **51**, 136-148.
- Tang, T.-Q., Chen, L., Yang, S.-C., Shang, H.-Y. (2015a) An extended car-following model with consideration of the electric vehicle's driving range, *Physica A*, **430**, 148-155.
- Tang, J., Liu, F., Wang, Y., Wang, H. (2015b) Uncovering urban human mobility from large scale taxi GPS data, *Physica A*, **438**, 140-153.
- Tang, T.-Q., Yu, Q., Yang, S.-C., Ding, C. (2015c) Impacts of the vehicle's fuel consumption and exhaust emissions on the trip cost allowing late arrival under car-following model, *Physica A*, **431**, 52-62.
- Traut, E., Hendrickson, C., Klampfl, E., Liu, Y., Michalek, J. J. (2012) Optimal design and allocation of electrified vehicles and dedicated charging infrastructure for minimum life cycle greenhouse gas emissions and cost, *Energy Policy*, **51**(4), 524-534.
- Tal, G., Nicholas, M. A., Daview, J., Woodjack, J. (2014) Charging behavior impacts on electric Vehicle Miles Traveled: Who is not plugging in? *Transportation Research Record*, **2454**, 53-60.
- Wu, X., Aviquzzaman, Md., Lin, Z. (2015a) Analysis of plug-in hybrid electric vehicles' utility factors using GPS-based longitudinal travel data, *Transportation Research Part C: Emerging Technologies*, **57**, 1-12.
- Wu, X., Freese, D., Cabrera, A., Kitch, W. A. (2015b) Electric vehicles' energy consumption measurement and estimation, *Transportation Research Part D: Transport and Environment*, **34**(34), 52-67.
- Yin, Y. (2008) Robust optimal traffic signal timing, *Transportation Research Part B: Methodological*, **42**(10), 911-924.
- Zhang, Y., Qian, Z., Sprei, F., Li, B. (2016) The impact of car specifications, prices and incentives for battery electric vehicles in Norway: Choices of heterogeneous consumers, *Transportation Research Part C: Emerging Technologies*, **69**, 386-401.
- Zou, Y., Wei, S., Sun, F., Hu, X., Shiao, Y. (2016) Large-scale deployment of electric taxis in Beijing: A real-world analysis, *Energy*, **100**, 25-39.

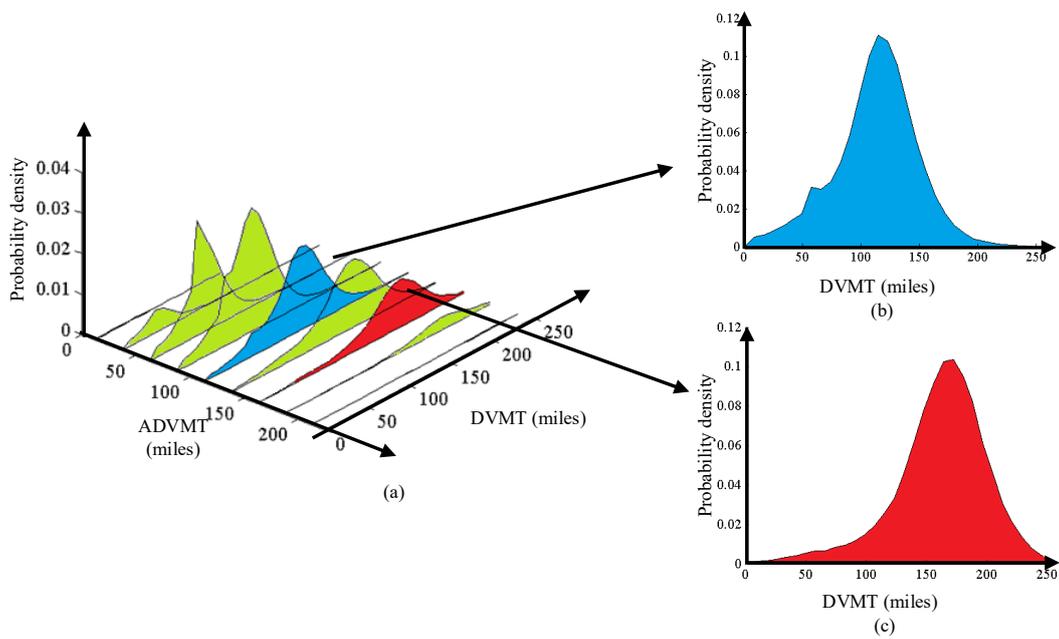


Fig. 1. An illustration of the mixture distribution of DVMT using the taxis travel data collected in Beijing, China. (a) The one-dimensional DVMT distribution for different kinds of drivers. (b) The probability density plot of the one-dimensional DVMT distribution for the 5th kind of drivers whose ADVMT falls in the range of [100, 125] miles. (c) The probability density plot of the one-dimensional DVMT distribution for the 7th kind of drivers whose ADVMT falls in the range of [150, 175] miles.

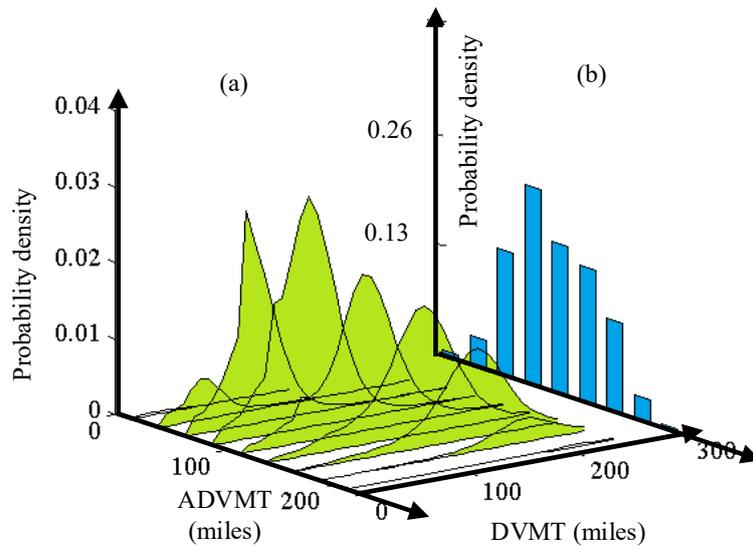


Fig. 2. An illustration of the relationship between the single distribution and the mixture distribution models of DVMT, using the taxis data collected in Beijing, China. (a) The one-dimensional DVMT distributions for different kinds of drivers. (b) The mapped single distribution model of ADVMT (In conventional single distribution DVMT models, we assume drivers are different, but the DVMTs of a driver is relatively constant in different days. Hence, we usually directly call it the single distribution model of DVMT).

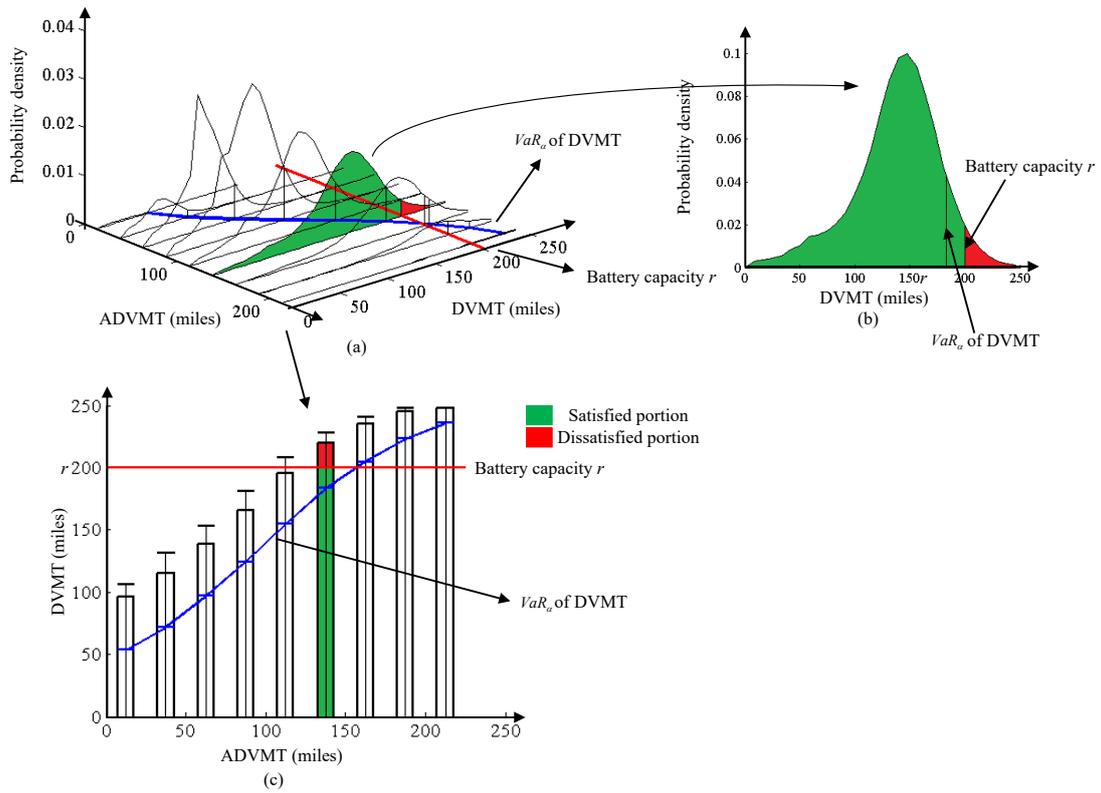


Fig. 3. The detail illustration of the solving objective using the taxis data collected in Beijing, China. (a) The distribution of DVMT of each kind of drivers, where we group the drivers into 9 kinds. The red curve represents the battery capacity which is set as 200 miles just for an example. The blue curve represents the VaR_α of the one-dimensional DVMT distributions for each kind of drivers. Here, we choose $\alpha = 0.9$. (b) The one-dimensional DVMT distribution of the 6th kind of drivers whose ADVMT falls in the range of [125, 150] miles. (c) The box plot for mixture DVMT distribution.

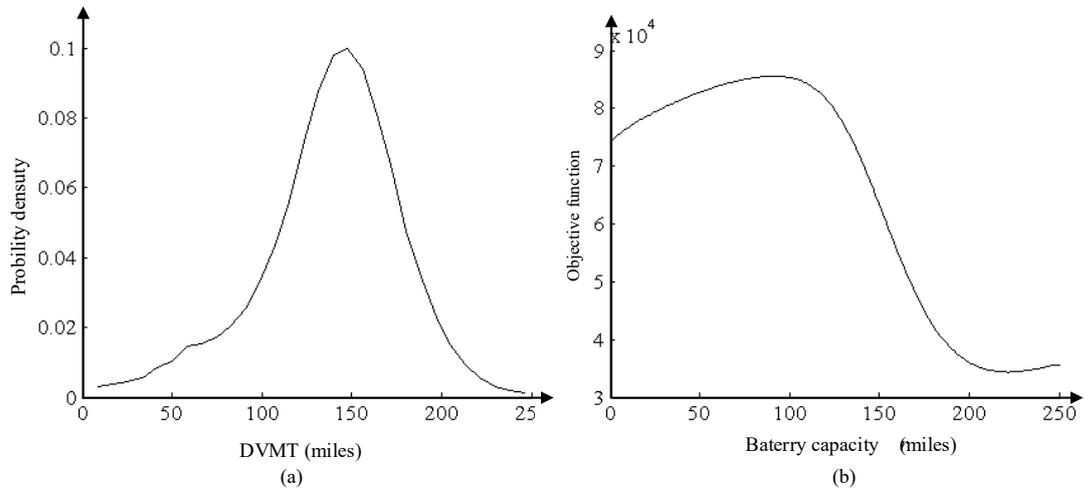


Fig. 4. (a) The one-dimensional DVMT distribution of the 6th kind of drivers. (b) The variation of the objective function $S^p + E\{S^f\} + \lambda\sqrt{\text{Var}\{S^f\}}$ with respect the battery capacity r .

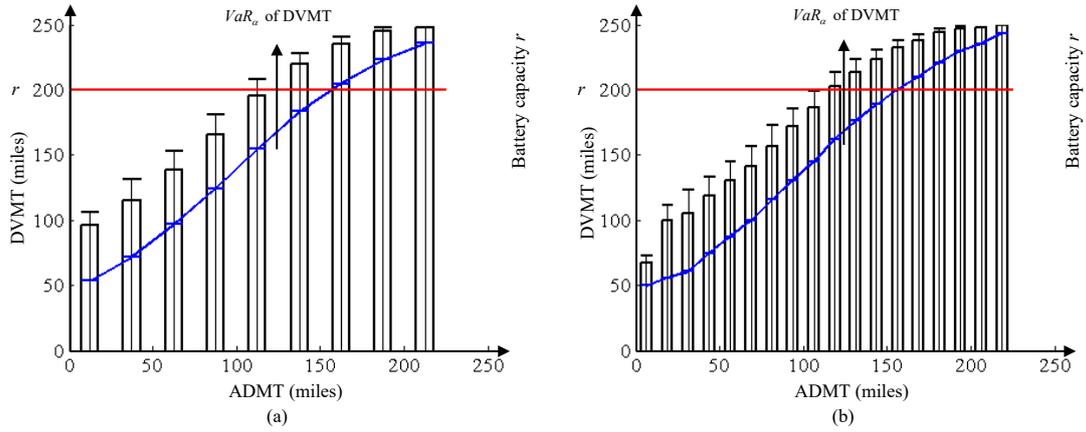


Fig. 5. The box-plot of the mixture DVMT distributions, when we select (a) 9 kinds of drivers; (b) 18 kinds of drivers.

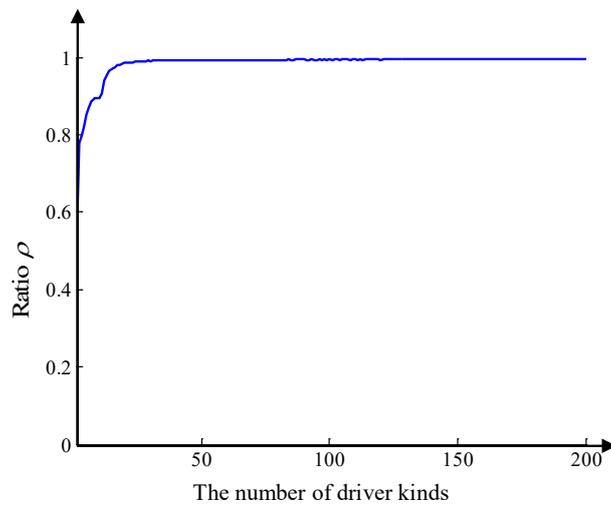


Fig. 6. The relationship between the overall ratio about how many drivers with similar ADVMT have similar distribution of DVMT distributions and the number of kind drivers.

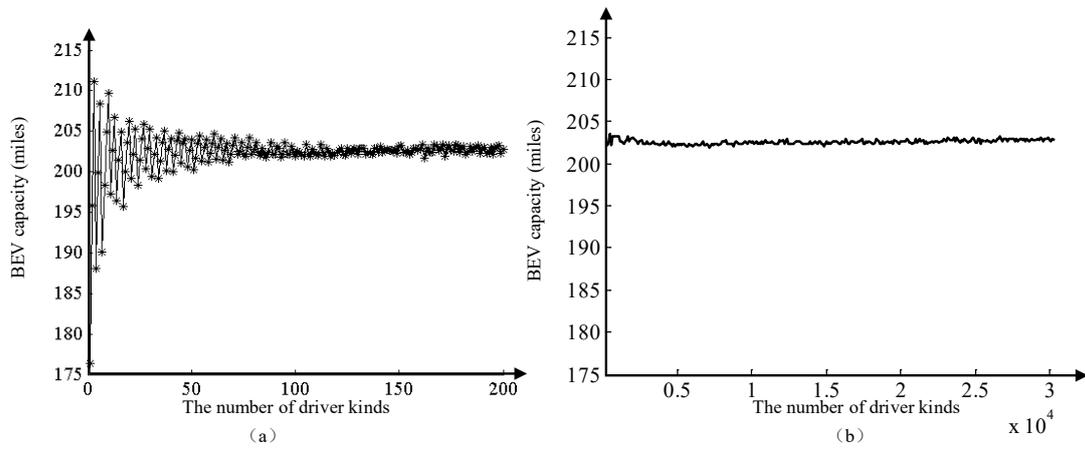


Fig. 7. The desired BEV capacity vs. the number of driver kinds, where we set the confidence level $\alpha = 0.9$ and $\beta = 0.9$. In subfigure (a), the number of driver kinds increase from 1 to 200 with an incremental step 1; while in subfigure (b), the number of driver kinds increase from 201 to 30300 (there are totally 30347 samples) with an incremental step 100.

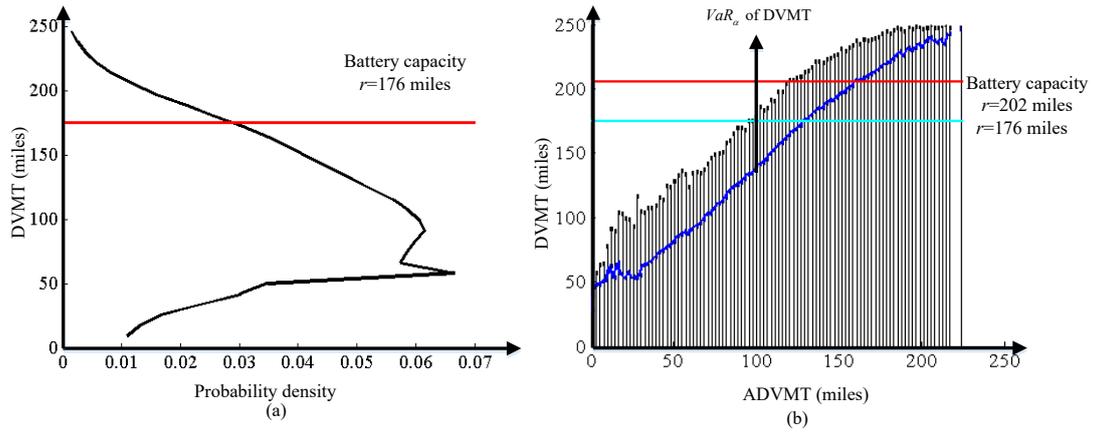


Fig. 8. The illustration of VaR design for BEV capacity, (a) when the single DVMT model is considered; (b) when the mixture DVMT model is considered (here we choose 100 kinds of drivers).

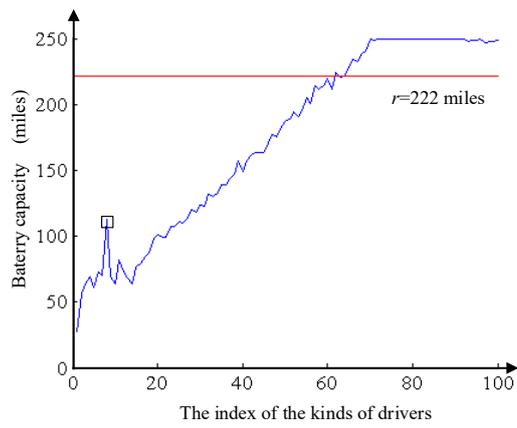


Fig. 9. A comparison of the mean-variance design solution for all the drivers and for each kind of drivers. Here, we group drivers into 100 kinds.

Table. A1 The nomenclature list

Notation	Meaning of notation	Remark
n	The number of driver kinds	
p_i	The proportion of the i th kind of drivers is	$i = 1, \dots, n$
$f_i(x)$	A one-dimensional probability density function to describe the diversity of DVMT for the i th kind of drivers	
$f(x)$	A probability density function to describe the diversity of DVMT for all drivers	
A_{\max}	The largest ADVMT value observed	
r	Batter capacity of BEVs in terms of miles can travel	
α	The confidence level that characterizes the portion of days in which capacity r is larger than the DVMT of a driver	$\alpha \in [0,1]$
β	The confidence level that characterizes the portion of drivers whose DVMT can be satisfied in a α -level	$\beta \in [0,1]$
S	The objective function	
S^p	The battery price	
S^f	The cost of failing to meet the travel demand	
λ	The weighting parameter in mean-variance analysis	

List of Figure captions

Fig. 2. An illustration of the mixture distribution of DVMT using the taxis travel data collected in Beijing, China. (a) The one-dimensional DVMT distribution for different kinds of drivers. (b) The probability density plot of the one-dimensional DVMT distribution for the 5th kind of drivers whose ADVMT falls in the range of [100, 125] miles. (c) The probability density plot of the one-dimensional DVMT distribution for the 7th kind of drivers whose ADVMT falls in the range of [150, 175] miles.

Fig. 2. An illustration of the relationship between the single distribution and the mixture distribution models of DVMT, using the taxis data collected in Beijing, China. (a) The one-dimensional DVMT distributions for different kinds of drivers. (b) The mapped single distribution model of ADVMT (In conventional single distribution DVMT models, we assume drivers are different, but the DVMTs of a driver is relatively constant in different days. Hence, we usually directly call it the single distribution model of DVMT).

Fig. 3. The detail illustration of the solving objective using the taxis data collected in Beijing, China. (a) The distribution of DVMT of each kind of drivers, where we group the drivers into 9 kinds. The red curve represents the battery capacity which is set as 200 miles just for an example. The blue curve represents the $VarR_\alpha$ of the one-dimensional DVMT distributions for each kind of drivers. Here, we choose $\alpha = 0.9$. (b) The one-dimensional DVMT distribution of the 6th kind of drivers whose ADVMT falls in the range of [125, 150] miles. (c) The box plot for mixture DVMT distribution.

Fig. 4. (a) The one-dimensional DVMT distribution of the 6th kind of drivers. (b) The variation of the objective function $S^p + E\{S^f\} + \lambda\sqrt{\text{Var}\{S^f\}}$ with respect the battery capacity r .

Fig. 5. The box-plot of the mixture DVMT distributions, when we select (a) 9 kinds of drivers; (b) 18 kinds of drivers.

Fig. 6. The relationship between the overall ratio about how many drivers with similar ADVMT have similar distribution of DVMT distributions and the number of kind drivers.

Fig. 7. The desired BEV capacity vs. the number of driver kinds, where we set the confidence level $\alpha = 0.9$ and $\beta = 0.9$. In subfigure (a), the number of driver kinds increase from 1 to 200 with an incremental step 1; while in subfigure (b), the number of driver kinds increase from 201 to 30300 (there are totally 30347 samples) with an incremental step 100.

Fig. 8. The illustration of VaR design for BEV capacity, (a) when the single DVMT model is considered; (b) when the mixture DVMT model is considered (here we choose 100 kinds of drivers).

Fig. 9. A comparison of the mean-variance design solution for all the drivers and for each kind of drivers. Here, we group drivers into 100 kinds.

List of Table captions

Table. A1 The nomenclature list