Title
Leveraging Big Data and Coordinated Charging for Effective Taxi Fleet Electrification: The 100% EV Conversion of Shenzhen, China

Journal
IEEE Transactions on Intelligent Transportation Systems, 23(8)

ISSN
1524-9050

Authors
Bauer, Gordon S
Zheng, Cheng
Shaheen, Susan
et al.

Publication Date
2022-08-01

DOI
10.1109/tits.2021.3092276

Peer reviewed

Leveraging big data and coordinated charging for effective taxi fleet electrification: the 100% EV conversion of Shenzhen, China

Gordon S. Bauer, Cheng Zheng, Susan Shaheen, and Daniel M. Kammen, *Member, IEEE*
IEEE Paper: eCF Paper Id: T-ITS-19-10-1166

***Abstract*— Taxis provide an important market for electric vehicles (EVs), but long charging durations and limited charger availability have prevented rapid adoption. Leveraging over two weeks of high-resolution GPS and battery data from almost 20,000 EVs in the all-electric Shenzhen taxi fleet, we analyze the potential to improve fleet-wide operations by optimizing both the location and timing of vehicle charging. We construct machine learning models to predict travel time, queuing time at charging stations, and charge consumption by time of day. Contrary to the emphasis on charging station siting in the literature, we find that optimizing charging locations would have a relatively limited impact. Instead, providing drivers with better real-time information about queuing times at charging stations, and enabling flexibility in battery charge during shift changes could reduce down-time per vehicle by over 30 minutes per day, while increasing the number of economically viable charging stations by over 50%. Moreover, taking full advantage of break periods and nighttime to charge could reduce downtime per vehicle by over one hour per day, reducing revenue losses due to charging by roughly 90%. These results are verified with evidence from real-time charging station data and driver shift-change data. Policy recommendations from this study include establishing citywide open data platforms to integrate real-time data on vehicle trajectory, battery charge, and charger availability, and providing drivers and companies with training on best charging practices. As a number of cities worldwide move toward fully electrified taxi fleets, this analysis has large-scale implications for decarbonized, cleaner urban areas.**

***Index Terms*—electric vehicles, charging optimization, big data, fleet simulation, taxi electrification, China[1]**

I. Introduction

Meeting the Paris Climate Agreement's initial 2°C (now 1.5°C) target will require a wholesale shift to electrified transportation [1]. Transportation represents the fastest-growing source of the world's greenhouse gas (GHG) emissions, with passenger cars accounting for close to a sixth of carbon dioxide emissions and car sales set to more than double by 2050 [2]. Battery electric vehicles (BEVs) could reduce transportation-related carbon emissions and urban air pollution, [3], [4] but adoption has been slow due to several barriers, including higher upfront cost, limited driving range, and slower charging speed relative to conventional vehicles [5], [6].

In this study, we test the hypothesis that if electrified and operated in a coordinated way, taxi fleets hold the potential to overcome these barriers and drive dramatic increases in transportation electrification. Because taxis accumulate mileage more quickly than vehicles used only for personal use, BEVs used as taxis capture more savings from lower operating costs than personal BEVs [7], [8], and they provide better returns on public electrification investments in terms of reduced carbon emissions and air pollution per vehicle [9]. Because taxis are typically driven in urban cores, BEVs used in this context could increase public health benefits while also exposing many consumers to the technology, which might increase private BEV sales [10].

Previous studies have shown that many challenges lie in the way of complete taxi electrification. In a London pilot project, Uber found that over 80% of BEV drivers lacked access to home charging, and insufficient public infrastructure prevented drivers from serving as many rides as they could with internal combustion engine vehicles (ICEVs) [10]. Furthermore, transportation network company (TNC) drivers with BEVs have reported declining rides because their vehicles lacked sufficient charge as well as revenue losses owing to time spent charging and looking for charging stations [11], [12]. In a study in South Korea, BEV taxis provided a lower benefit-to-cost ratio compared with natural-gas-powered taxis because of limited charging infrastructure and battery range [13]. In

[1]G. Bauer is supported by a graduate research fellowship from the National Science Foundation. D. Kammen thanks the Karsten Family Foundation and the Zaffaroni Family for support. This work was supported by the United States National Science Foundation through the CyberSEES Program under Award 1539585 (to Callaway, Brewer & Kammen).

G. Bauer is with the Energy and Resources Group, University of California, Berkeley, CA 94720 USA (e-mail: gbauer@berkeley.edu).

C. Zheng is with Aspiring Citizens Cleantech and also with the Lee Kuan Yew Centre for Innovative Cities at Singapore University of Technology and Design (e-mail: andy.zheng@aspiringcitizens.com).

S. Shaheen is a professor in Civil and Environmental Engineering and Co-Director of the Transportation Sustainability Research Center at the University of California, Berkeley in Berkeley, CA 94720 USA (e-mail: sshaheen@berkeley.edu).

D. M. Kammen is with the Energy and Resources Group, University of California, Berkeley, CA 94720 USA and also with the Goldman School of Public Policy, University of California, Berkeley, CA 94720 USA (e-mail: kammen@berkeley.edu).

Beijing, researchers found that EV taxis with 130 to 160 km of battery range travel only 118 km per day, compared to 250 km for conventional vehicles, suggesting that they do not earn nearly as much revenue [14].

China is rapidly pushing toward the adoption of BEVs, and an increasing number of cities have set ambitious targets for taxi fleet electrification including: Taiyuan [15], Shenzhen [16], Guangzhou [17], Chengdu [18], and Beijing [19]. Shenzhen, China represents an especially interesting case study because it is the first top-tier Chinese city to have fully replaced its taxi fleet of over 20,000 vehicles with BEVs, using the BYD e6 model with 400 km battery range (80 kWh capacity with an average consumption of 0.2 kWh/km) [20].

To build a sustainable roadmap for taxi electrification that can be adapted and adopted elsewhere, the Shenzhen electric fleet must provide the same level of service as a conventional fleet at low cost and with significant carbon benefits. Local interviews conducted with drivers and reports in the media suggest, however, that time spent charging – in some cases over three hours per day – results in lost revenue, compounded by problems with queuing at popular charging stations [21].

Many studies have developed methods to optimize the siting of charging stations [22]–[25], but despite the unplanned nature of charging infrastructure siting in Shenzhen, sub-optimal station locations may not be the fundamental cause of operational inefficiencies. Drivers report that charging activity is concentrated during the afternoon because dayshift drivers feel obligated to fully charge their vehicle before delivering it to the nightshift driver – most vehicles are driven by two drivers for 12 hours per day each, and most dayshift drivers end their shift between 5 and 7 PM [26]. Some drivers use apps to find available stations, but others go to the same preferred station each day near the shift change location, leading to problems with queuing at popular stations. In turn, this uncertainty in waiting time means drivers go to charge two to three hours before a shift change, long before they typically need to: most drivers go to charge around 50% state of charge, and most afternoon charging events take less than one hour (this charging pattern is similar between weekdays and weekends). If left with extra time between charging and the end of their shift, drivers are often unwilling to accept trips that travel too far from the shift change location, reducing revenue opportunities during this period.

In this study, we use over two weeks of GPS and battery state of charge (SOC) data from about 20,000 electric taxis in Shenzhen to evaluate the potential of different interventions to mitigate the problems described above. The data come from January, May, and June 2019, and they consist of snapshots taken every five minutes from each vehicle. Using these data, we conduct simulations of four proposed interventions that could reduce the charging burden: 1) optimizing the location of charging stations to minimize travel time to charging stations, 2) optimizing the dispatch of vehicles to charging stations to minimize both travel and queuing times, 3) shifting more daytime charging to early morning hours when demand for taxi trips is low, and 4) shifting charging to times when vehicles are idle.

Several previous simulation studies have found that optimization strategies could improve the efficiency of electric taxi fleets. Lu et al. (2012) showed that having a dispatching strategy for electric taxis in Taipei, Taiwan successfully reduces charging wait times [27]. Bauer et al. (2018) conducted simulations on electrification scenarios for automated taxis in Manhattan and found that optimization of charging enabled the electric fleet to operate at lower cost than a fleet with conventional vehicles [9]. Bauer et. al (2019) expanded this work to TNC fleets in San Francisco and New York City and found similar results [28]. Tian et al. (2016) proposed a framework to recommend charging station locations to e-taxi drivers in Shenzhen [29]. However, the study was limited as the SOC data were crudely inferred from estimated charging locations. Tian et al. (2017) considered the scenario of a major shutdown of an EV charging station in Shenzhen and developed a strategy to re-allocate the charging demand to reduce queuing time and increase the usage rate of charging stations [30]. Finally, Dong et al. (2018) proposed a real-time framework to recommend location and charging time to electric taxi drivers in Shenzhen, and conducted simulations that showed significant potential improvements in charging station use and queuing times at charging stations [31].

No previous study, however, has analyzed impact on revenues or compared the effectiveness of multiple interventions. This deeper analysis is facilitated by the rich granularity of the data, including both vehicle state-of-charge and the timing and locations of shift changes. More importantly, no previous study has explored the underlying causes of the apparent inefficiencies in fleet operations. In particular, we find that inefficient charging behavior may be caused in large part by taxi drivers' preference for changing shifts at full charge in the afternoon, and that changing shifts at partial charge could significantly improve fleet operations.

In this study, we develop several machine learning models to predict operational characteristics of the taxi fleet and present a framework for how this modeling platform can be implemented in practice. We conduct simulations to estimate and compare the potential impact of these various interventions on fleet performance, driver revenue, and charging infrastructure use. Finally, using one day of driver shift-change data, we compare the performance of groups of drivers with different charging patterns to verify the simulation results.

II. Methodology

Our study methodology consists of the following steps employing GPS and charging data. Vehicle data for approximately 19,224 electric taxis in Shenzhen were prepared by Aspiring Citizens Cleantech (ACC) between May 27 and June 9 and between January 17 to 19, 2019, for a total of 17 days of data. Data include snapshots of location, state of charge (% of battery capacity), and operation status (hired or available) every five minutes for each taxi while the vehicle is turned on. To verify simulation results, ACC also provided vehicle data and driver shift change data for July 25, 2019, including timestamps for when each driver logged in or out of each vehicle. Finally, ACC also provided summaries of interviews conducted with approximately 30 drivers in June and July 2019. These interviews were conducted with drivers who were approached at charging stations by ACC or while ACC staff were taking taxi trips. Our analysis with these data consists of four key methodologies: 1) charging inference; 2) queuing inference; 3) developing predictive models; and 4) simulation-based optimization analysis focused on four intervention strategies: a) charging station location, b) dispatch to charging stations, and c) shifting more charging to nighttime and d) shifting more charging to break periods. Each of these interventions involves an optimization problem that is solved through simulation, maximizing the objective function for each vehicle or charging station while obeying a series of constraints. While this method does not guarantee global optimality, it shows what might be feasibly achieved in the real world. With more advanced optimization methods, it is possible that even greater savings than those described in this study could be achieved.

A. Charging inference

The development and ownership of charging infrastructure in Shenzhen has been highly decentralized, so there is no database of locations and numbers of chargers at each station. While regulators are assembling this data, it is incomplete. New infrastructure is still being added frequently, so any data that exists will soon be out of date. Thus, as with other issues analyzed in this paper, the best source of data for inferring charging station locations and charger availability is vehicle GPS location and state of charge.

To begin our data analysis, we first inferred charging events by finding locations where the vehicle SOC increased. We discarded potential charging events where SOC changes by less than 10% to filter out events caused by erroneous SOC readings; these discarded events account for almost 5% of charging events but less than 0.5% of total charging (see supplemental information for cumulative distributions). We then performed hierarchical clustering on all of these locations, using 200 m as the maximum distance between any two vehicles in the same cluster. This cluster diameter was considered appropriate because it is roughly the size of a city block in downtown Shenzhen and also roughly the distance between the two most distance charging plugs at the largest inferred charging station in Shenzhen (see supplemental information for satellite image). To infer charging locations, we then took the mean values of latitude and longitude of all points in each cluster, and then removed outlier charging events greater than 200 m from the closest charging location.

To estimate the total number of chargers at each station, we found the maximum number of vehicles charging simultaneously at the location across the total period in the dataset. Note that this estimate of the number of chargers at each station does not account for usage aside from taxis. To validate these inferred estimates, we collected data on actual charger availability for 18 charging stations, whose locations match the charging station locations we inferred from the taxi data. We collected these data by taking a sequence of screenshots of two different charging apps each hour over the course of four days between July 19 to 22, 2019, then using Google Vision API [32] to extract the number of available chargers from each screenshot. The estimated availability for these 18 charging stations inferred from the taxi data closely tracks the availability displayed by the apps, especially during peak charging periods in the early morning and late afternoon (see supplemental information for details). As such, we assume our estimates of the number of chargers at each charging station is sufficiently accurate for the simulations conducted in this work. Future work can improve accuracy by expanding data collection from the charging apps and integrating it with taxi data more thoroughly.

B. Queuing inference

For every timestamp, we identified which vehicles were relocating to each charging station (i.e., idle vehicles whose next activity was charging). We inferred these vehicles to be queuing, if they satisfied the following criteria: 1) the vehicle moved less than 1 km over the previous five minutes and was either: a) the closest vehicle to the charging station and within 500 m or b) within 100 m of another vehicle queuing at the same station. In other words, after finding the closest queuing vehicle to each station, we inferred lines of queuing vehicles by sequentially adding the next-nearest queuing vehicle. We define queuing time as the time elapsed between the start of a vehicle joining a charging queue and the beginning of the charging event.

C. Predictive models

We developed all predictive models with the h2o package using R 3.5.1. We tuned hyper-parameters using five-fold cross-validation, and we validated model performance with 10% of the data left out of model training for testing purposes. We found the difference in performance between the gradient boosting models and other model types to be negative or insignificant, so we only report results using gradient boosting machines for all models (see supplemental information for more details).

First, we constructed a model to predict the average queuing time before charging at each charging location for 30 minutes into the future. Several studies have proposed methods for optimal scheduling of charging to maximize charging operator profit while minimizing wait time and power grid impacts [33], [34]. However, these studies do not consider how drivers acting independently without the ability to reserve a timeslot for charging might use a forecast of queue times to identify the best station for their needs. Our approach based on machine learning has several advantages over classical queuing theory: we can forecast queue times even though we do not have exact data on which vehicles are queuing (queuing is inferred as described in the previous section), or data on how many vehicles other than taxis are queuing; and we can analyze the impact of data that could be easily displayed to drivers, i.e. forecasts for average queue times 30 minutes in the future, rather than theoretical estimates for current queue times. The predictive model includes the following features: the number of vehicles queuing at the charging station, the number of vehicles charging, the flows in and out of the station and queue, the size and location of the station, fixed effects for each station, and a variety of transformations of each of these variables. As reliable prediction requires sufficient charging events at each charging station, we only applied the predictive model to charging stations with at least 20 inferred chargers, resulting in 93 charging stations accounting for over 75% of charging events, for a total of 374,976 observations. The model prediction tracks the actual queuing time during peak hours quite well, with a root mean squared error of 4.3 minutes (see supplemental information for details).

Second, we also constructed a model to estimate travel time between any two points in the city of Shenzhen, using 4.7 million trip durations as training outcomes. We clustered trip origin and destination locations into 500-meter grid cells and downloaded travel-time estimates between each pair of grid cells from OpenStreetMaps. These travel-time estimates do not account for traffic congestion and served as inputs in our model to predict trip durations, along with straight-line distance, time of day, day of week, and origin and destination locations. Over 85% of predictions fall within five minutes of the actual trip duration, with a root mean squared error (RMSE) of 4.7 minutes (see supplemental information for details). By forecasting travel times based on real-world taxi trips, this model incorporates average road congestion by time of day and day of the week. We assume that taxi re-routing does not significantly affect congestion because taxis represent less than 1% of all vehicles on the road in Shenzhen [35].

We also developed a model to predict surplus SOC at the end of the dayshift for each vehicle, as will be described in the next section. These three models are summarized in Table I below.

TABLE I
SUMMARY OF PREDICTIVE MODELS

	Outcome variable		
	Queue time	Travel time	Surplus SOC
Model type	Gradient boosting machine		
Parameters	Five-fold cross-validation Learning rate = 0.01 Number of trees = 10,000 Stopping condition: Measure RMSE every 100 trees, terminate if improvement < 0.001		
Sample size	374,976	4,660,593	210,397
Features	vehicles queuing (count and flow)	OpenStreetMaps estimate (min)	average surplus SOC by vehicle
	vehicles charging (count and flow)	straight-line distance (km)	average time of charging by vehicle
	total chargers	time of day	average SOC after charging by vehicle
	station location (longitude, latitude)	day of week	time of day
	station fixed effects	origin location (longitude, latitude)	day of week
	time-lagged variables of vehicle counts and flows	destination location (longitude, latitude)	SOC at start of charging event
RMSE	4.3 minutes	4.7 minutes	13.5% SOC

D. Optimization analysis

1) Relocate charging stations:

Using the travel time model described above, we identified the detour time required to visit a charging station in each grid cell C, defined as the additional travel time from the last drop-off cell A to the next pick-up cell B via the charging station C ($t_{AC} + t_{CB}$), compared with the travel time directly from cell A to cell B (t_{AB}, see Fig. 1 below).

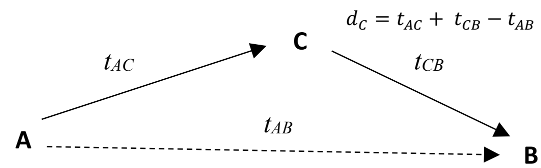


Fig. 1. Schematic of the calculation of detour time (d_C) caused by driving to a charging station.

We then used the following heuristic approach to analyze the impact of optimally locating charging stations. We relocated each existing charging station to the grid cell C that minimized total detour time of all charging events taking place there and then reassigned each charging event to the charging station that minimized detour time, following the objective function shown in (1) below. We repeated these two steps until the simulated detour times converged, as defined by changing by less than 5%, which occurred after three iterations.

$$\min \sum_i \; t_{A_iC} + t_{CB_i} - t_{A_iB_i} \qquad (1)$$

2) Dispatch to charging stations:

Using the queuing and travel time models described above, we predicted the average queuing time at all stations for the next 30 minutes and assigned vehicles to the station with available chargers that minimized the total delay (the sum of detour time plus queuing time). Given potential inaccuracies in our estimate of the number of chargers available, as a conservative approach we only relocated vehicles to a new charging station if we estimated less than 90% of the chargers to be occupied. After reassignment, we updated the estimated queuing time and the number of available chargers at each station. We repeated these two steps until the simulated queuing times changed by less than 5%. This process is described by the objective function shown in (2) and constraints listed in (3-4) below, where q_C is the estimated queuing time, d_C is the travel detour time, and p_C is the number of charging ports at the charging station.

$$\min(d_C + q_C); \qquad (2)$$
$$d_C = t_{A_iC} + t_{CB_i} - t_{A_iB_i}, \qquad (3)$$
$$n_{charging} < 0.9 * p_C \qquad (4)$$

3) Shift more charging to nighttime:

Based on our analysis and results from the driver interviews, we assumed that charging has a negligible opportunity cost between 1 AM and 7 AM on weekdays and between 2 AM and 8 AM on weekends due to low trip demand. As such, we argue that the SOC that the nightshift driver had at the start of this "zero-cost" period could be interpreted as surplus SOC that the dayshift driver did not need to charge before the shift change.

We defined these before-shift-change charging events as the last charging events for each vehicle before 8 PM or the first charging event after 8 PM in cases where there was no charging event between 10 AM and 8 PM. Based on the average amount of surplus SOC that each vehicle had on other days, the time of day, and the starting SOC, we trained a model to predict how much surplus SOC a vehicle would have at the start of the "zero-cost" period, if the dayshift driver charged to 100% immediately before changing shifts. For each prediction, we then calculated the SOC at the shift change that would give the vehicle a 95% and 99% chance of reaching the "zero-cost" time without falling below 10% SOC. Based on the average power of the charging station selected for the before-shift-change charging event, we then calculated the amount of time that could be saved by following our recommendation for the SOC at shift change. This process is described in (5) below, where $t_{n,d}$ is the time savings of taxi n on day d, t is time of day, and P is the power of the charging station.

$$t_{n,d} = f(t, SOC) * P \qquad (5)$$

4) Shift more charging to break times:

First, we identified break periods for each vehicle, defined as periods when the vehicle was neither serving a trip nor charging and the vehicle's odometer reading did not increase for more than 30 minutes. Using the queuing and travel time models described above, we estimated the amount of time available for charging during each break period, and the maximum amount of SOC that could be charged. We then conducted a simulation to determine the potential to satisfy charging needs during these break periods by removing all charging events, i.e. assuming the vehicles' SOC did not change during existing charging events. We then added charging during break periods incrementally to keep SOC between 0% and 100% and calculated the amount of additional charging time needed to maintain SOC above 10% between each break (see supplemental information for example profile). This process is described by the objective function in (6) and constraints listed in (7-12) below.

$$min(SOC_{emerg})\,; \qquad (6)$$

$$SOC_{end} > 10, \qquad (7)$$

$$\Delta SOC_{max} = (t_{idle} - \min d_C + q_C) * P_C, \qquad (8)$$

$$\Delta SOC_{break} = \min(\Delta SOC_{max}\,, 100 - SOC_{start}), \qquad (9)$$

$$SOC_{end} = SOC_{start} + \Delta SOC_{break}, \qquad (10)$$

$$\Delta SOC_{emergency} = \max(0, -1 * SOC_{end}), \qquad (11)$$

$$SOC_{start} = SOC_{end} + \Delta SOC_{emergency} \qquad (12)$$

E. Study limitations

There are several limitations to this analysis that can be addressed with expanded data access. First, without complete real-time data from existing charging stations, there may be inaccuracy in our estimates of charger availability. In the next phase of our research, we will expand charging station data collection efforts to conduct more accurate and detailed charger availability analysis. Similarly, we also plan to obtain and integrate data on driver shift change locations and times to inform our estimates of detour time and surplus SOC during the afternoon shift change. Given the limited timespan of the vehicle data available to us, it is possible the interventions we analyze would have different impact at other times of year. However, as shown in Table II, the standard deviation of results between days is less than 5%, and we also found no significant difference between results from January and May-June, suggesting that our results are generalizable.

Finally, this analysis does not consider a variety of behavioral factors that influence decisions on where and when to charge including: the availability of rest places and food, opportunities to meet friends and other drivers, and desire to maintain a large SOC buffer at all times (for example, it appears some drivers never let their batteries fall below 50% SOC). To analyze these factors in depth and to test potential strategies in the real world, we plan to conduct a pilot project via a smartphone app that will provide drivers with real-time information on queuing times at charging stations as well as an accounting tool to keep track of SOC charged across shifts, to facilitate changing shifts at partial charge.

III. Results

As shown in Fig. 2 below, analysis of the Shenzhen taxi data shows there are two major peaks in charging events in the early morning and late afternoon (likely preceding shift changes), along with a smaller peak during the lunchtime period. Each peak is accompanied by a significant number of taxis queuing at charging stations, especially in the afternoon. Notably, this afternoon peak is also accompanied by a decreased number of hired vehicles serving trips and an increased total number of active vehicles, suggesting that these charging events result in significant lost revenue.

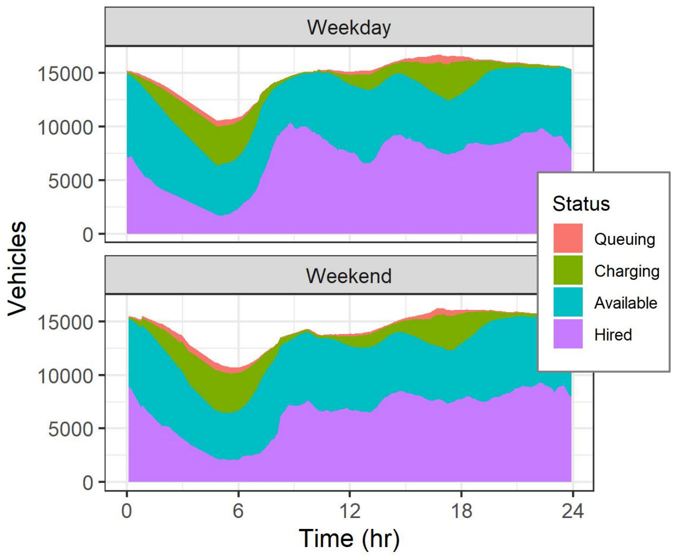


Fig. 2. Number of taxis by operating status over time of day, for both weekends and weekdays.

A comparison with TNC trip volumes in Shenzhen across three months in 2019 shows correlation between decreased taxi trip volumes and increased TNC trip volumes during the afternoon period, suggesting that taxi trip volumes are constrained by vehicle supply (TNC data are not shown here due to issues with confidentiality).

Drawing on our literature review, our analysis, and taxi driver interviews, we identify three key problems in taxi fleet operations: 1) queuing time due to inefficient routing of vehicles to charging stations, 2) spatial mismatch between charging station locations and charging demand, and 3) temporal mismatch between charging events and the time periods of lowest opportunity cost of charging. In the following sections, we analyze each of these problems and the impact of proposed interventions. Table II below summarizes the intervention results.

TABLE II
SUMMARY OF IMPACTS OF ALL PROPOSED INTERVENTIONS

Strategies/ interventions	Time savings (min/vehicle/day)		Revenue generated (USD/ vehicle/day)	Electricity savings (USD/ vehicle/day)
	dayshift	total		
Optimal charging station locations	5 ± 0.2	8 ± 0.4	0.72	0
Optimal dispatch to charging stations	10 ± 0.5	14 ± 0.7	1.45	0
Flexible SOC during shift change	25 ± 3	0	3.77	0.43
Optimal charging during break periods	72 ± 7	123 ± 12	11.16	0.58

Impact on revenue is estimated by multiplying the daytime savings by the average revenue generated per vehicle during that time. Results for flexible shift change SOC are derived from the scenario providing 99% reliability of maintaining a 10% buffer in SOC between the last daytime charging event and early morning hours. Exchange rate between USD and Chinese Yuan (RMB) (0.14:1) was recorded on August 1, 2019. Error values show standard deviation of results between days of data.

A. *Optimal charging locations*

The charging market in Shenzhen is highly fragmented with over 100 different charging station operators [36], and planning has not been integrated with charging demand data, potentially leading to a spatial mismatch. Meanwhile, drivers report that they prefer larger charging stations to smaller ones due to greater reliability in expected queuing time. This is reflected in charger usage data; concentrating chargers into fewer large stations at the best locations may increase usage rates. Many studies have suggested siting taxi charging stations close to areas with high trip density can improve fleet performance [22]–[25], but given that drivers in Shenzhen typically charge near their shift-change location, we assume that the optimal infrastructure siting strategy would minimize the detour time incurred by visiting a charging station on the way to drivers' next destination. As described in the methodology section D.1, we analyzed the impact of a scenario in which charging stations are relocated to the locations that minimize detour time. As shown in Fig. 3 below, this scenario results in a much higher density of charging stations in densely populated areas of the city close to major corridors. We find that on average, optimizing charging station locations could save each vehicle eight minutes per day.

B. *Optimal charging dispatch*

While existing charging network apps show real-time availability at charging stations, these data are fragmented between platforms and may not reflect actual queuing times due to many vehicles arriving during a short time window. Accurate queuing time predictions may help drivers locate charging stations with less queuing time and improve the certainty of the total time needed to charge, allowing drivers to start charging closer to the end of their shift.

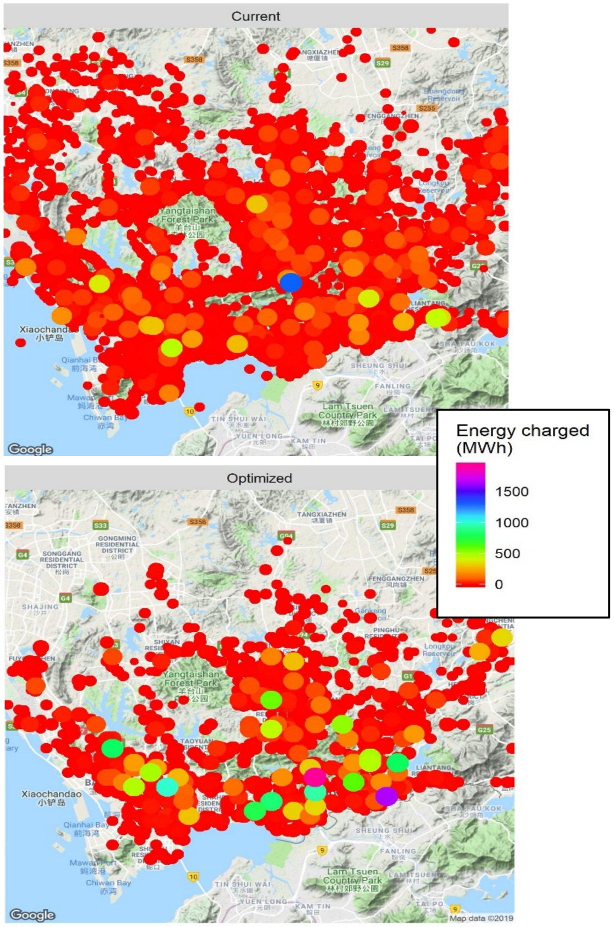


Fig. 3. Heat maps comparing the distribution of energy charged at current charging stations with the distribution after relocating stations to the locations that minimize the detour time.

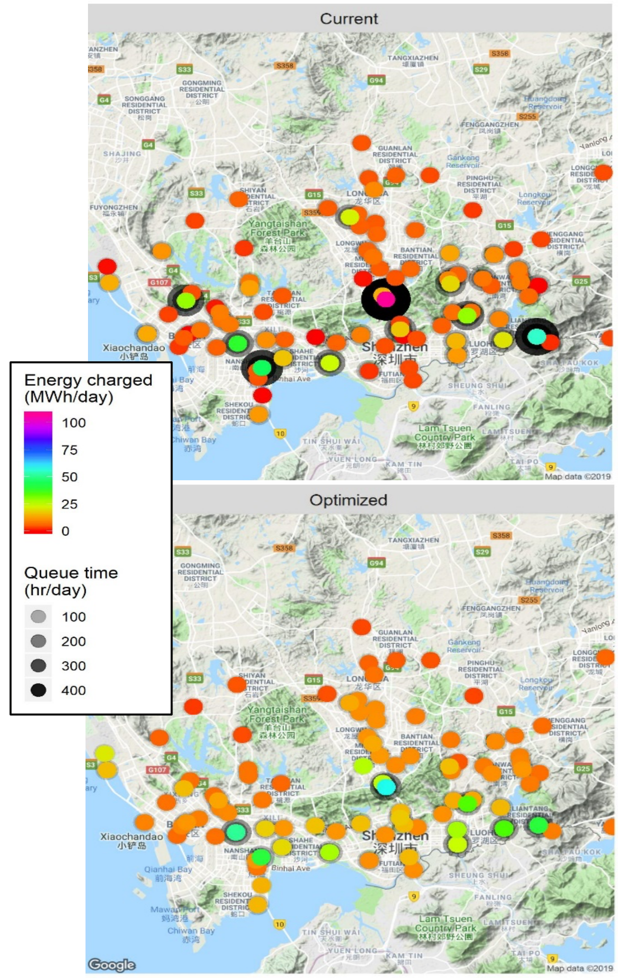


Fig. 4. Heat maps depicting charger use and total queuing time (black rings) before (top) and after (bottom) dispatching optimization.

Using the queuing model described earlier, we estimated the amount of time vehicles could save if they charged at the station causing the least delay, which is defined as the sum of the detour and queuing time. We find that through optimization it is possible to reduce the time spent queuing per charging session by almost 50%, from over 10 minutes to about five, with a total time savings of 14 minutes per vehicle per day. While over 26% of vehicles currently spend over 30 minutes queuing per day, in the simulation we find that this proportion could be reduced to less than 10% (see supplemental information for details). We find that optimizing the dispatch of vehicles to charging stations could save almost 4,500 hours of aggregate downtime per day, potentially resulting in over $10 million per year in additional revenue.

Optimizing the dispatch of vehicles to charging stations could also greatly increase the economic sustainability of the charging network, which may be threatened by the pending removal of charger installation subsidies [37]. As shown in Fig. 4 above, currently most charging is concentrated in a few large charging stations with over 100 inferred chargers each, resulting in large total queuing times at each of these stations (shown by the black rings). Drivers report that they prefer larger stations because the queuing time is more certain due to higher turn-over rates. Not surprisingly, by providing accurate estimates of the expected queuing times at each station, we find that charging events can be dispersed among more charging stations.

Due to the non-linear relationship between usage and amortized cost per kilowatt-hour, we find that this dispersion can have positive impacts on charging economics. Using the cost parameters reported by [36], including costs for charger construction, maintenance, and land, we calculated the cost of charging amortized per kilowatt-hour, assuming a 10% discount rate and a 10-year charging station lifetime. As shown in Fig. 5 below, the increased usage rate at many stations results in much lower amortized cost. Under the current pricing regulation, charging stations are not allowed to charge customers more than $0.11 per kWh on top of the electricity price [36]. Without subsidies on charger installation, less than 10% of inferred charging stations would be economically sustainable, including less than 50% of the top 93 inferred charging stations. In contrast, with optimal dispatching of vehicles to charging stations, 75% of inferred charging stations would be profitable without subsidies.

Interestingly, once vehicles have been dispatched to optimal charging stations, repeating the heuristic location optimization described in the first section has little effect. We find that only stations on the city's periphery are significantly affected, resulting in time savings of less than one minute per vehicle per day (see supplemental information for details).

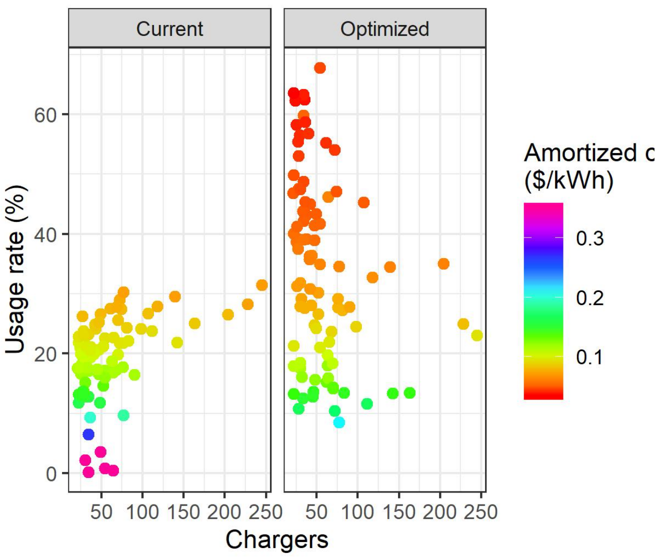


Fig. 5. Usage rate of major charging stations (average percent of time each charger is occupied) by number of chargers at each station, before (left) and after (right) dispatching optimization.

C. Shift more charging to early morning hours

We also studied the potential for enhancing efficiency by shifting some charging from the dayshift to early-morning hours, when charging carries a lower opportunity cost due to lower demand for taxi trips.

As seen in Fig. 6, during early-morning hours both the proportion of the fleet that is hired and the hourly revenue per vehicle drop substantially, suggesting there are more idle vehicles than are needed to serve demand. Based on this finding, we estimated the number of "surplus vehicles" that could go charge without affecting the fleet's capacity to serve trip demand. This value is defined as the number of idle vehicles that could be removed from duty while maintaining the ratio of hired vehicles to idle vehicles at or below 1:1.25, the minimum hired ratio observed during the daytime. We find that the number of vehicles charging during early-morning hours (defined as 1 AM - 7 AM on weekdays and 2 AM – 8 AM on weekends) could almost double without affecting the fleet's capacity to serve trip demand.

Despite this apparent incentive to charge more during the nightshift, most drivers interviewed reported that they feel an obligation to charge to full SOC before the afternoon shift change to ensure that dayshift and nightshift drivers both pay for a fair share of the electricity and have an equal opportunity to serve long trips.

There is no established policy by regulators or companies that requires drivers to change shift at full SOC. If there were a data-driven accounting tool to facilitate payment between drivers to compensate for time spent charging, the barrier to optimizing charging times among drivers could be overcome. Based on the prediction model described in the methodology, we find that starting with a 72% SOC during the afternoon shift change gives the nightshift driver a 95% probability (on average) of reaching the early-morning hours without falling below 10% SOC (see supplemental information for details). Based on the derived charging power of each charging station (33 kW on average), adding this SOC flexibility to the afternoon shift change would save each dayshift driver 40 minutes per day and increase their fare revenue by over US $5 per day. Given that electricity prices are also lower between 11 PM and 7 AM, this intervention would also reduce charging costs by over US $1 per vehicle per day. Even if nightshift drivers require 99% probability that they will maintain at least 10% SOC, a flexible SOC policy during shift changes would save each dayshift driver over 25 minutes per day. In aggregate, this intervention could save day shift drivers 8,000 to 12,000 hours per day, potentially yielding over $25 million per year in additional revenue. This analysis does not include the potential for reducing queuing and detour times by increasing flexibility in when and where drivers charge, and so these estimates are likely conservative.

D. Shift charging to break periods

Early-morning hours are not the only time of day when charging has a negligible opportunity cost. We estimate that each vehicle takes over two hours of breaks per day on average, even when only including periods when the vehicle spends at least 30 minutes idle in the same location (not including charging). Over 60% of vehicles spend at least as much time on these breaks as they spend charging during the dayshift. When asked why they do not currently use break periods to charge, several drivers reported that they see no need to do so, because they must charge to full SOC before changing shift regardless. Without the constraint of changing shift at full SOC, drivers could save time by charging during breaks. Such an intervention could also reduce fatigue driving and improve safety by encouraging drivers to take longer breaks every few hours to ensure that they remain alert while driving.

As described in the methodology, we developed a simulation model to estimate the minimum amount of additional charging time required given full use of both break periods and early morning hours, finding that on average vehicles could satisfy all their charging needs with only 10 additional minutes, saving over 70 minutes per vehicle per day.

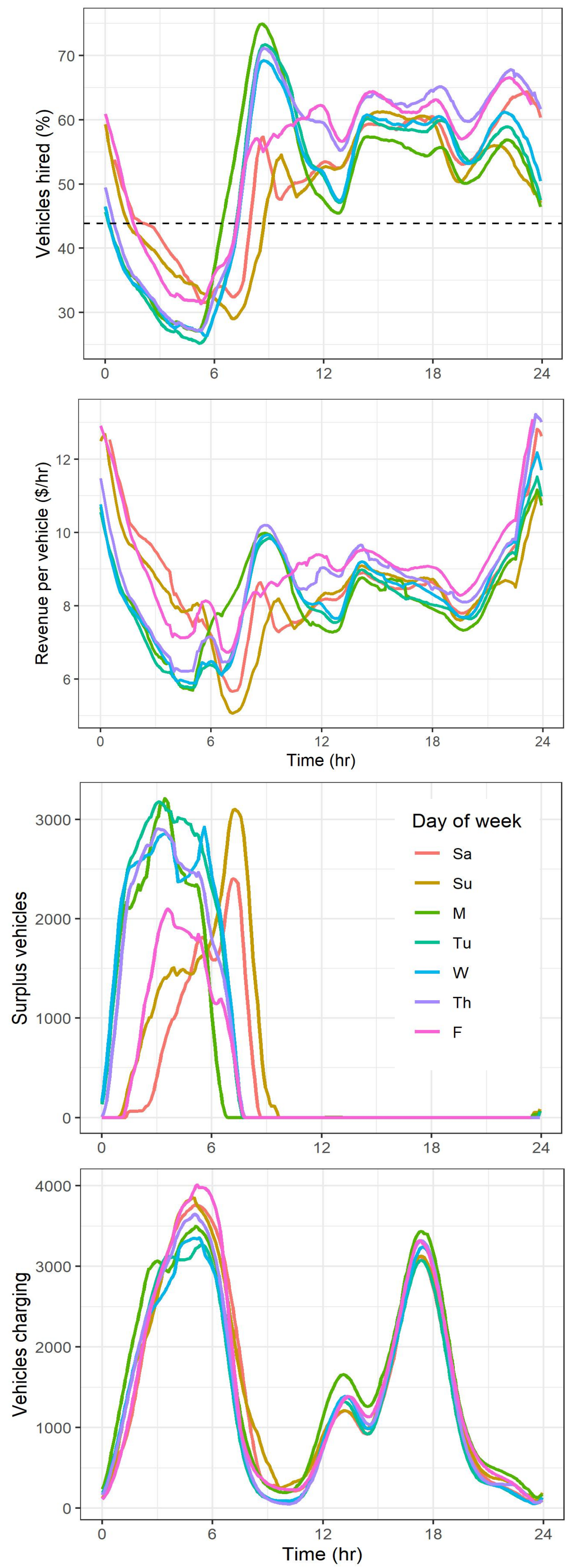


Fig. 6. From top (all hourly moving averages disaggregated by both time of day and day of week): a) proportion of vehicles hired (not including vehicles that are disengaged or charging); b) average net revenue per vehicle per hour (fare revenue minus electricity cost); c) number of "suplus vehicles," defined as the number of vehicles in the fleet that could be charging while keeping the fraction of vehicles hired below the minimum value observed during the daytime (44%); d) number of vehicle charging.

This optimization strategy, shown as the "optimized" vehicle charging profile in Fig. 7, results in a larger early-morning peak, a broader lunchtime peak, and a substantially smaller afternoon peak. This reduced demand for charging during the afternoon increases the fleet's capacity to serve trip demand during afternoon peak hours, yielding over $75 million per year in additional revenue.

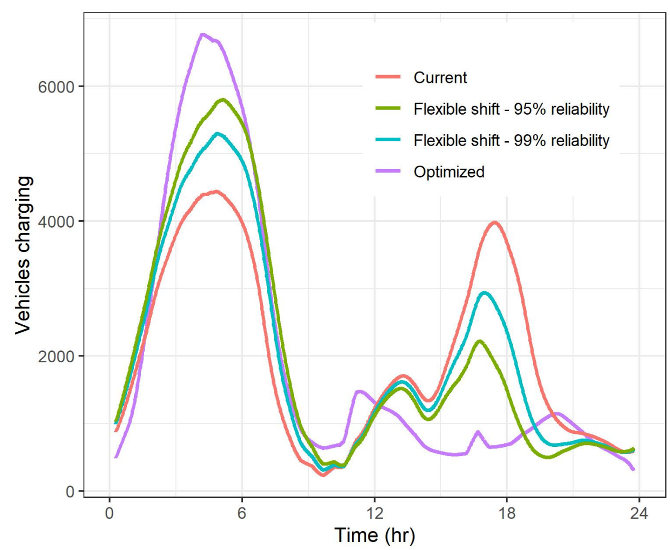


Fig. 7. Number of vehicles charging by time of day in four different scenarios. The "flexible shift" scenarios allow drivers to change shift at a sufficient SOC to provide either 99% (green) or 95% (blue) confidence for the nightshift driver to operate without charging until the early morning hours (defined as 1am on weekdays and 2am on weekends). The optimized scenario assumes full usage of both break periods and early morning hours for charging. The large nighttime peak introduces the possibility of scarcity of available chargers. However, given that we inferred over 12,000 chargers in the taxi data, we expect that the dispatching optimization, described above, would be able to mitigate this potential issue.

E. Evidence from current driver behavior

Using the fleet's driver shift-change data, we find that some drivers have already adopted charging patterns aligned with some of the strategies described above. For example, for 1,251 vehicles or roughly 7% of the fleet, dayshift drivers ended their shift with a 60 to 85% SOC. Compared with "full SOC" dayshift drivers in Fig. 8 below, "flexible SOC" dayshift drivers tend to charge earlier in the afternoon, likely coinciding with their lunch break, and "flexible SOC" nightshift drivers tend to charge slightly earlier in the early-morning hours.

As shown in Table III below, compared with drivers that change shifts at full charge, on average these "flexible SOC" drivers earn more revenue ($5 per vehicle per day), operate for slightly more time, and continue accepting trips with less time before their shift change (1.42 hours versus 2.25 hours). They also charge further from the shift-change location (6.1 km versus 2.9 km), meaning they have more choice of where to charge. This increased choice likely results in less queuing time and greater certainty of charging time, both of which reduce the overall opportunity cost of charging.

Additionally, on average "flexible SOC" dayshift drivers stop charging at 92%, saving 10 minutes per charging event by avoiding slow charging speeds at close to full SOC. These findings are consistent with our simulation results that suggest that flexibility in the SOC required at shift changes can increase driver earnings by enabling drivers to charge during break times and at charging stations with shorter queuing times.

TABLE III
VEHICLE STATISTICS BY CHARGING BEHAVIOR

Attribute	SOC at afternoon shift change (%)		p-value
	60 to 85 (flexible SOC)	>85 (full SOC)	
Number of vehicles	1251	10616	n/a
Total revenue (USD/vehicle/day)	211.62	206.15	***
Dayshift revenue (USD/vehicle/day)	107.09	102.31	***
Distance between charging station and shift change location (km)	6.1	2.9	***
Time between charging and shift change (hr)	3.63	2.30	***
Time between shift change and last trip dropoff (hr)	1.42	2.25	***
Operating time (hr/day)	19.67	19.48	*
Revenue per operating hour (USD/vehicle/hr)	10.79	10.57	**
Starting SOC of charging events (%)	53	52	x
Ending SOC of charging events (%)	92	96	***
Charging time (hr)	1.09	1.26	***

p-values describe results of two-tailed t-tests, with the following significance levels: x >= 0.05, * < 0.05, ** < 0.01, *** < 0.001

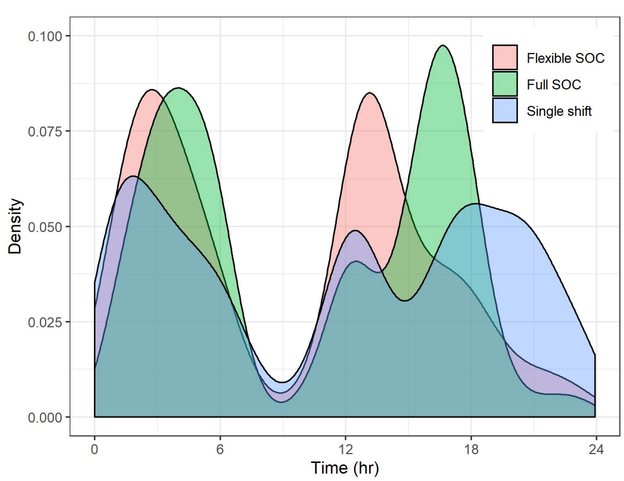


Fig. 8. Density profile of charging events by start time on July 25, 2019, grouped by shift-change type. "Flexible SOC" represents vehicles changing shift in the afternoon at 60 to 85% SOC, "full SOC" represents vehicles changing shift in the afternoon at 85 to 100% SOC, and "single shift" represents vehicles that did not report a shift change on this day.

IV. Conclusion

With the Shenzhen taxi fleet data on vehicle trajectory and battery state of charge, we illustrate the capability of big data to reveal system-level inefficiencies and inform simple optimization strategies to facilitate effective total electrification. Enabling flexible SOC during the afternoon shift change could reduce aggregate vehicle downtime by over 10,000 hours per day, while fully using break periods and early morning hours to charge could save over 20,000 hours per day. We also find that optimizing vehicle dispatching to charging stations could improve the economic sustainability of charging infrastructure by increasing the percentage of viable chargers without subsidies from less than half to 75%.

These findings naturally lead to recommendations for policies that encourage driver behavioral change and coordinated charging. By testing and implementing such strategies, cities and fleet operators could greatly alleviate the operating burden arising from electrification. For more effective fleet electrification, cities should consider developing data platforms that integrate demand-side data on the charging needs of various transportation modes with supply-side data on charger availability and construction. This combination would enable critical feedback control so cities can predict and balance the demand and supply for electrified transportation. Aside from the real-time fleet optimization discussed in this study, data integration can also be used to minimize the stress vehicle charging places on the power grid and maximize usage of intermittent renewable energy resources. Meanwhile, switching from conventional vehicles to BEVs effectively will require significant behavioral changes, and both taxi drivers and fleet operators could benefit from targeted training to better understand how to best use this new technology.

Shenzhen provides a strong leadership model for municipal governments to consider as they seek full electrification of both fleets and private transportation. Given China's goal of phasing out conventional vehicles over the next several decades [38], the findings of this study have large-scale implications for other Chinese megacities aiming to shift to low-carbon transportation.

Acknowledgment

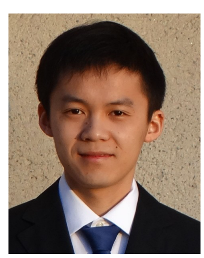
Shenzhen Transport Authority and Aspiring Citizens Cleantech's PAIR team provided access to data used for this study. Jennifer Conde, Kelly Jiang, and Soomin Woo at University of California, Berkeley helped develop early versions of predictive models.

References

[1] M. Schneider, "The Road Ahead for Electric Vehicles," *ICCG Reflect.*, vol. 54, pp. 1–8, 2017.
[2] H. Hao, Y. Geng, and J. Sarkis, "Carbon footprint of global passenger cars: Scenarios through 2050," *Energy*, vol. 101, pp. 121–131, 2016, doi: 10.1016/j.energy.2016.01.089.
[3] T. R. Hawkins, B. Singh, G. Majeau-Bettez, and A. H. Strømman, "Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles," *J. Ind. Ecol.*, vol. 17, no. 1, pp. 53–64, 2013, doi: 10.1111/j.1530-9290.2012.00532.x.
[4] H. Cai and M. Xu, "Greenhouse gas implications of fleet electrification based on big data-informed individual travel patterns,"

Environ. Sci. Technol., vol. 47, no. 16, pp. 9035–9043, 2013, doi: 10.1021/es401008f.
[5] E. H. Green, S. J. Skerlos, and J. J. Winebrake, "Increasing electric vehicle policy efficiency and effectiveness by reducing mainstream market bias," *Energy Policy*, vol. 65, pp. 562–566, 2014, doi: 10.1016/j.enpol.2013.10.024.
[6] C. King, W. Griggs, F. Wirth, K. Quinn, and R. Shorten, "Alleviating a form of electric vehicle range anxiety through on-demand vehicle access," *Int. J. Control*, vol. 88, no. 4, pp. 717–728, 2015, doi: 10.1080/00207179.2014.971521.
[7] J. B. Greenblatt and S. Saxena, "Autonomous taxis could greatly reduce greenhouse-gas emissions of US light-duty vehicles," *Nat. Clim. Change*, vol. 5, no. September, pp. 860–865, 2015, doi: 10.1038/nclimate2685.
[8] R. Li and G. Fitzgerald, "Ride-Hailing Drivers Are Ideal Candidates for Electric Vehicles," *Rocky Mountain Institute*, 2018. https://rmi.org/ride-hailing-drivers-ideal-candidates-electric-vehicles/ (accessed Sep. 18, 2018).
[9] G. S. Bauer, J. B. Greenblatt, and B. F. Gerke, "Cost, Energy, and Environmental Impact of Automated Electric Taxi Fleets in Manhattan," *Environ. Sci. Technol.*, vol. 52, no. 8, 2018, doi: 10.1021/acs.est.7b04732.
[10] S. R. George and M. Zafar, "Electrifying the Ride-Sourcing Sector in California," San Francisco, 2018.
[11] J. Hagman and J. H. M. Langbroek, "Conditions for electric vehicle taxi: A case study in the Greater Stockholm region," *Int. J. Sustain. Transp.*, vol. 0, no. 0, pp. 1–10, 2018, doi: 10.1080/15568318.2018.1481547.
[12] C. Said, "Uber's new plan to woo drivers: It's electric," *San Francisco Chronicle*, 2018. https://www.sfchronicle.com/business/article/Uber-s-new-plan-to-woo-drivers-It-s-electric-13005719.php (accessed Sep. 18, 2018).
[13] S. Baek, H. Kim, and H. J. Chang, "A feasibility test on adopting electric vehicles to serve as taxis in daejeon metropolitan City of South Korea," *Sustain. Switz.*, vol. 8, no. 9, 2016, doi: 10.3390/su8090964.
[14] Y. Zou, S. Wei, F. Sun, X. Hu, and Y. Shiao, "Large-scale deployment of electric taxis in Beijing: A real-world analysis," *Energy*, vol. 100, pp. 25–39, 2016, doi: 10.1016/j.energy.2016.01.062.
[15] Global Opportunity Explorer, "Taiyuan: World's Fastest Electric Taxi Fleet Overhaul," *Global Opportunity Explorer*, 2018. .
[16] Shenzhen Local Treasure, "Since August, Shenzhen has added a electric car to use a pure electric vehicle," *Bendibao.com*, 2018. .
[17] Xinhua News Agency, "Guangzhou City promotes pure electric taxis," *Gov.cn*, 2019. .
[18] Xinhua, "Electric taxis hit road in Sichuan," *China.org.cn*, 2019. .
[19] D. Galeon, "70,000 Beijing taxis are being converted to electric power," *World Economic Forum*, 2017. https://www.weforum.org/agenda/2017/03/beijing-is-converting-its-fleet-of-70-000-taxis-to-electric-power.
[20] Z. Tian, Y. Wang, C. Tian, F. Zhang, L. Tu, and C. Xu, "Understanding operational and charging patterns of Electric Vehicle taxis using GPS records," *2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014*, pp. 2472–2479, 2014, doi: 10.1109/ITSC.2014.6958086.
[21] Y. Juanjuan, "How to solve the problem of charging thousands of electric taxis in Shenzhen?," *Gaogong.com*, 2017. .
[22] L. Hu, J. Dong, Z. Lin, and J. Yang, "Analyzing battery electric vehicle feasibility from taxi travel patterns: The case study of New York City, USA," *Transp. Res. Part C Emerg. Technol.*, vol. 87, no. December 2017, pp. 91–104, 2018, doi: 10.1016/j.trc.2017.12.017.
[23] M. Li, Y. Jia, Z. Shen, and F. He, "Improving the electrification rate of the vehicle miles traveled in Beijing: A data-driven approach," *Transp. Res. Part Policy Pract.*, vol. 97, pp. 106–120, 2017, doi: 10.1016/j.tra.2017.01.005.
[24] S. Y. He, Y. H. Kuo, and D. Wu, "Incorporating institutional and spatial factors in the selection of the optimal locations of public electric vehicle charging facilities: A case study of Beijing, China," *Transp. Res. Part C Emerg. Technol.*, vol. 67, no. June, pp. 131–148, 2016, doi: 10.1016/j.trc.2016.02.003.
[25] C. J. R. Sheppard, A. Harris, and A. R. Gopal, "Cost-Effective Siting of Electric Vehicle Charging Infrastructure With Agent-Based Modeling," vol. 2, no. 2, pp. 174–189, 2016.
[26] Southern Metropolis Daily, "Shenzhen: 80% of electric taxi drivers surveyed believe that charging affects daily operations," *Sohu.com*, 2018. .
[27] J. L. Lu, M. Y. Yeh, Y. C. Hsu, S. N. Yang, C. H. Gan, and M. S. Chen, "Operating electric taxi fleets: A new dispatching strategy with charging plans," *2012 IEEE Int. Electr. Veh. Conf. IEVC 2012*, pp. 1–8, 2012, doi: 10.1109/IEVC.2012.6183233.
[28] G. S. Bauer, A. Phadke, J. B. Greenblatt, and D. Rajagopal, "Electrifying Urban Ridesourcing Fleets at No Added Cost through Efficient Use of Charging Infrastructure," Berkeley, CA, 002, 2018.
[29] Z. Tian *et al.*, "Real-Time Charging Station Recommendation System for Electric-Vehicle Taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3098–3109, 2016, doi: 10.1109/TITS.2016.2539201.
[30] Z. Tian, L. Tu, Y. Wang, F. Zhang, and C. Tian, "Impact of core charging station's cease operation in the entire charging station system: A case study in Shenzhen," *Proc. - 3rd IEEE Int. Conf. Big Data Comput. Serv. Appl. BigDataService 2017*, pp. 90–95, 2017, doi: 10.1109/BigDataService.2017.12.
[31] Z. Dong, C. Liu, Y. Li, J. Bao, Y. Gu, and T. He, "REC: Predictable Charging Scheduling for Electric Taxi Fleets," *Proc. - Real-Time Syst. Symp.*, vol. 2018-Janua, pp. 287–296, 2018, doi: 10.1109/RTSS.2017.00034.
[32] Google, "Google Vision AI," 2019. .
[33] Y. Kim, J. Kwak, and S. Chong, "Dynamic pricing, scheduling, and energy management for profit maximization in PHEV charging stations," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1011–1026, 2016.
[34] A. Rabiee, A. Ghiasian, and M. A. Chermahini, "Long term profit maximization strategy for charging scheduling of electric vehicle charging station," *IET Gener. Transm. Distrib.*, vol. 12, no. 18, pp. 4134–4141, 2018.
[35] "Guangdong and Shenzhen ease restrictions on car purchases to boost economy." https://news.cgtn.com/news/3d3d774d78456a4d35457a6333566d54/index.html (accessed Oct. 19, 2020).
[36] A. Crow, D. Mullaney, Y. Liu, and Z. Wang, "A New EV Horizon: Insights From Shenzhen's Path to Global Leadership in Electric Logistics Vehicles," Boulder, Colorado, 2019.
[37] People's Network, "From construction to operation, who is benefiting from the adjustment of the charging pile subsidy.," *CNR.cn*, 2019. http://auto.cnr.cn/gdzx/20190819/t20190819_524736953.shtml (accessed Sep. 30, 2019).
[38] The Economist, "China moves towards banning the internal combustion engine," 2017. .

Gordon S. Bauer is a Ph.D. student in the Energy and Resources Group at University of California, Berkeley, where he conducts research at the intersection between sustainable transportation, data science, and human behavior. He received a B.A. in Chemistry with Highest Honors from Williams College in 2014. He is a recipient of the National Science Foundation Graduate Research Fellowship and the Link Energy Fellowship.



Cheng Zheng is the founder and chief scientist of Aspiring Citizens Cleantech, and adjunct fellow with Lee Kuan Yew Centre for Innovative Cities at Singapore University of Technology and Design. He received a B.Eng. in Mechanical Engineering from the National University of Singapore in 2008, and a Ph.D. in Mechanical Engineering from University of California, Berkeley in 2014.

Susan Shaheen is a pioneer and thought leader in emerging mobility strategies. She is a Professor in the Department of Civil Engineering and Co-Director of the Transportation Sustainability Research Center (TSRC) at the University of California, Berkeley. She has authored 64 journal articles, over 120 reports and proceedings articles, 11 book chapters, and co-edited two books. In May 2019, she received the most influential paper award from World Conference on Transportation Research. She was named "Faculty of the Year" by the Institute of Transportation Studies, Berkeley in December 2018 and received the 2017 Roy W. Crum award from the Transportation Research Board (TRB) for her distinguished achievements in transportation research. In May 2016, she was named one of the top 10 academic thought leaders in transportation by the Eno Transportation Foundation. She is a member of the TRB Executive Committee and chair of the subcommittee for Shared-Use Vehicle Public Transport Systems of TRB. She also is a member of the Mobile Source Technical Review Subcommittee to the U.S. Environmental Protection Agency's Clean Air Act Advisory Committee.

Daniel M. Kammen is an IEEE member, and Professor with the Energy and Resources Group where he is the Chair, the Goldman School of Public Policy where he directs the Center for Environmental Policy, and the Department of Nuclear Engineering. He is the Founding Director of the Renewable and Appropriate Energy Laboratory. He has authored over 400 publications and more than 40 Congressional Testimonies. He was appointed by Secretary of State Hillary Clinton in 2010 as the First Energy Fellow of the Environment and Climate Partnership for the Americas initiative. He served as a Science Envoy for U.S. Secretary of State John Kerry from 2016 to 2017 Kammen was Chief Technical Specialist for Renewable Energy for the World Bank 2010 - 2011. He has served as a contributing or coordinating lead author on various reports of the Intergovernmental Panel on Climate Change (IPCC) since 1999. The IPCC shared the 2007 Nobel Peace Prize.