Received April 6, 2021, accepted April 17, 2021, date of publication April 20, 2021, date of current version April 30, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3074568*

# Optimal Decision-Making Method for a Plug-In Electric Taxi in Uncertain Environment

**YANG YOU, JISONG ZHU, YICHUAN HUANG, AND ZHAOXIA JING, (Member, IEEE)**
School of Electric Power Engineering, South China University of Technology, Guangzhou 510640, China
Corresponding author: Zhaoxia Jing (zxjing@scut.edu.cn)

This work was supported by the Science and Technology Planning Project of Guangdong Province, China, under Grant 2019B090911001.

**ABSTRACT** This paper studies the optimal decision-making problem for a plug-in electric taxi (PET) in a time-varying complex environment, i.e., a passenger environment, charging station environment, traffic environment, and taxi company management system, in order to maximize PET profit in a short-term operating cycle. First, this problem is formulated as a sequential decision-making problem composed of multiple decision slots. Then, to make the model more practical, the model is divided into two parts: an external environment and an electric taxi model for refinement. The uncertainty and time-varying characteristics of four environmental aspects, including passengers, charging stations, traffic, and taxi company management systems, are analysed and modelled. The transitions between adjacent processes and the environmental feedback of each process are modelled by further subdividing both the serving process and the charging process of the PET into multiple subprocesses, including cruising, carrying passengers, driving to the charging station, queueing before charging, and connecting to the power grid for charging. There are several uncertain factors in the sequential decision-making process for the PET, which leads to difficulty in solving the problem. To address this difficulty, the model-free algorithm SARSA is chosen. Finally, the effectiveness of the proposed method is verified by simulation results.

**INDEX TERMS** Plug-in electric taxi, decision making, uncertainty, SARSA algorithm, load modeling.

## NOMENCLATURE

| | |
|---|---|
| $E_n$ | Environment state at decision-making slot $n$. |
| $S_n$ | PET state at decision-making slot $n$. |
| $A_n$ | Action selected at decision-making slot $n$. |
| $R_n$ | Reward at decision-making slot $n$. |
| $G_n$ | Return following decision-making slot $n$. |
| $t_n$ | Discrete time steps at decision-making slot $n$. |
| $d_{fp}(t)$ | PET cruising distance at time $t$. |
| $w_q(t)$ | PET queueing time at time $t$. |
| $m_e(t)$ | PET unit charging price at time $t$. |
| $m_s(t)$ | Unit kilometre price of PET serving at time $t$. |
| $v(t)$ | PET speed at time $t$. |
| $f_d$ | PET energy consumption per kilometre at time $t$. |
| $d_{cp}$ | PET passenger carrying distance (a random variable). |
| $d_{fc}$ | Distance for PET driving to charging station (a random variable). |
| $w_{fp}$ | PET cruising time |
| $w_{cp}$ | PET passenger carrying time. |
| $w_{fc}$ | PET driving time to charging station. |
| $w_g$ | Time to connect to the grid for charging. |
| $\sigma_{fp}^2$ | Variance of $d_{fp}(t)$ probability distribution. |
| $\sigma_q^2$ | Variance of $w_q(t)$ probability distribution. |
| $T_{fp}^p$ | Peak passenger travel time. |
| $T_{fp}^g$ | Off-peak passenger travel time. |
| $T_q^p$ | Peak queueing time. |
| $T_q^g$ | Off-peak queueing time. |
| $T_e^p$ | Peak charging price time. |
| $T_e^g$ | Off-peak charging price time. |
| $T_v^p$ | Peak traffic time. |
| $T_v^g$ | Off-peak traffic time. |
| $T_s^{dt}$ | PET operating time during the day. |
| $T_s^{nt}$ | PET operating time at night. |
| $D_{fp}^p$ | Mean value of $d_{fp}(t)$ probability distribution during $T_{fp}^p$. |
| $D_{fp}^g$ | Mean value of $d_{fp}(t)$ probability distribution during $T_{fp}^g$. |
| $W_q^p$ | Mean value of $w_q(t)$ probability distribution during $T_q^p$. |

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

| | |
|---|---|
| $W_q^g$ | Mean value of $w_q(t)$ probability distribution during $T_q^p$. |
| $M_e^p$ | PET unit charging price during $T_e^p$. |
| $M_e^g$ | PET unit charging price during $T_e^g$. |
| $V^p$ | PET speed value during $T_v^p$. |
| $V^g$ | PET speed value during $T_v^g$. |
| $M_s^{dt}$ | PET unit kilometre price for service during $T_s^{dt}$. |
| $M_s^{nt}$ | PET unit kilometre price for service during $T_s^{nt}$. |
| $e_{\max}$ | PET battery capacity. |
| $p_c$ | PET charging power. |
| $m_{sf}$ | PET flag-down fare. |
| $\gamma$ | Discount-rate parameter in SARSA algorithm. |
| $\alpha$ | Learning-rate parameter in SARSA algorithm. |
| $\varepsilon$ | Probability of taking a random action in an $\varepsilon$-greedy policy. |
| $r_{pv}$ | Penalty when PET does not have enough energy to complete task. |

## I. INTRODUCTION

There is a consensus in all countries worldwide to vigorously develop new-energy vehicles, as represented by electric vehicles, to replace traditional fuel vehicles [1]. The large-scale development of electric vehicles will present great changes to the transportation system and have a profound impact on electric power and energy systems [2]–[4].

Since uncoordinated large-scale electric vehicle charging will threaten the safety and stability of the power grid, there has been a wealth of research on electric vehicle dispatching strategies under the background of electric power systems. Ref. [5] studied the optimal operation scheduling of electric vehicles in a smart distribution network, aiming at improving voltage profiles and reducing network losses. Ref. [6] proposed the application of electric vehicles to participate in the primary frequency regulation of a power grid. The goal of optimal control is to suppress the frequency fluctuations in a power grid. In [7] and [8], plug-in electric vehicle storage was exploited for load flattening and voltage regulation. Ref. [9] proposed a novel business model to optimize the charging energy of PEVs to maximize aggregator profits.

Obviously, all the above studies have focused on private electric vehicles, and little attention has been paid to electric taxis. Electric taxis are different from private electric vehicles in terms of power characteristics, running time, and power consumption [14]. Specifically, compared with private electric vehicles and electric buses, electric taxi owners are completely profit oriented. However, most studies have put forward scheduling strategies from the perspective of power systems [5]–[9], ignoring the interests of electric vehicle owners or electric vehicle fleets. Considering the profit-seeking of taxi owners, the effects of these methods will be greatly reduced if they are directly applied to electric taxis. Second, electric taxis will be in use most of the time, and owner actions will be affected by many external factors (such as passengers, traffic conditions, etc.), which leads to great uncertainty. However, most of the current studies make deterministic assumptions [10], [11] or deterministic predictions [12], [13] about electric vehicle loads. From a practical point of view, these methods apply only to private electric vehicles (which are parked most of the time) and electric buses (which have relatively fixed travel times and routes) but not to electric taxis. In summation, most of the existing studies cannot be directly applied to electric taxis, so electric taxis need more attention and research.

Since 2014, electric taxis have gradually attracted the attention of researchers. Ref. [15] and [16] proposed a multi-objective optimizing model for electric taxi charging station deployment based on taxi trajectory data. Ref. [17] and [18] studied the optimal charging problem for a plug-in electric taxi by considering factors such as time-varying service incomes and charging costs. Ref. [19] provided a charging station recommendation system for electric taxis to achieve spatial optimization of the electric taxi charging problem. Ref. [20] launched an online method to calculate proper real-time prices from the viewpoint of a utility company, such that the collective charging load of a fleet could track the desired value as the response to prices. Ref. [21] proposed a multiagent framework for PET operation and developed the multistep Q(λ) learning approach for PETs to make decisions under various situations. However, the main shortcomings of these existing studies are as follows:

- The influences of important factors other than the power grid (such as those related to the transportation system) on the operating decisions of PETs have been ignored.
- The models rely too much on historical operation data or the deterministic assumption of the PET behaviour chain, so they cannot reflect the autonomous response of electric taxi owners to changes in the external environment and the uncertainty of their behaviours.
- Since most studies rely heavily on real-time environment information (such as electricity price and traffic data), these methods place specific requirements on communication technologies, business models, and supporting facilities in real-world applications, which makes the applications more difficult.

In this paper, the operating processes of PETs are mathematically expressed, and then the operating behaviours of PETs are modelled from the source (the key factor affecting the operating decision of PETs). The uncertainty of the PET operating environment leads to uncertainty in the behaviour of PET owners, which greatly increases the difficulty of solving this optimization problem. To overcome this difficulty, the SARSA algorithm is selected to solve the problem accordingly. The main contributions of this paper are summarized as follows:

- Because electric taxi owners are completely profit oriented, this paper studies the optimal decision-making problem of PETs from the perspective of actual electric taxi owners to maximize the daily income of a single electric taxi.
- PET driver behaviours are influenced by many environmental factors, creating great uncertainty. This paper focuses on the influence of environmental factors

(passengers, charging stations, traffic, taxi company management systems) on the behaviours of PET drivers and puts forward a PET operating behaviour model based on environmental uncertainty. The model can reflect the autonomous responses of PET drivers to the environment and the uncertainty of this behaviour.

- The optimal decision-making method for PETs proposed in this paper is mainly based on operating environment characteristics, which reduces the dependence of the model on historical PET data and real-time environmental data. Therefore, the model has greater expandability and wider application scenarios.

The rest of this paper is organized as follows. In Section II, the operation optimization problem of a single PET in a short-term operating cycle is simplified, and the difficulty of solving the problem is analysed. The model is further refined in Section III, mainly by including the operating environment of PETs and environment feedback signals to PETs, the states in the operating processes of PETs, and the transitions between states. Section IV introduces the specific solution methods and steps of the problem. The simulation results are shown in Section V, and Section VI offers conclusions from this paper.

## II. PROBLEM FORMULATION

In simple terms, this paper addresses how to maximize the benefits of a PET in its short-term operating cycle by properly arranging the service time and charging behaviours. A short operating cycle of a PET is usually composed of multiple service processes and charging processes. A PET needs to decide between the service process and charging process, that is, to decide whether to serve or charge in the next process. A Markov chain model is used to describe this problem. It is assumed that a short-term PET operating cycle contains $N+1$ decision-making slots. At the decision-making slot $n = 0, 1, 2, \ldots, N$, based on the environment $E_n \in \mathrm{E}$ where the current decision-making slot is located, the PET should select the next action $A_n \in \mathrm{A}$, i.e., $A_n = 1$ $A_n = 0$, standing for service action and charging action, respectively. Then, the PET will complete the selected behaviour. After the end of a service or charging behaviour, it will enter into the next decision-making slot. At the slot, the PET will select an action $A_n \in \mathrm{A}$ based on the new environment $E_{n+1} \in \mathrm{E}$, and as above, the cycle will continue until the end of a short-term operation. It can be said that a short operating cycle of an electric taxi contains $N$ discrete decision-making slots, or it can be said that it contains $N$ processes of interaction between a PET and its operating environment, which is represented by the following sequence:

$$E_0, A_0, E_1, A_1, E_2, \ldots, E_N, A_N$$
$$s.t.\ E \in \mathrm{E}, \quad A \in \mathrm{A} \tag{1}$$

Sequence (1) contains two transitions. The first is how to select action $a_n$ based on the current environment $e_n$, which is the transition from the environment state $e_n$ to the action $a_n$. The second is what happens to the environment after action $a_n$ has been executed, which is the transition from the action $a_n$ to the environment state $e_{n+1}$. The PET policy $\pi(a|e)$ drives the first transition. The policy of the PET is specifically described as follows: the PET evaluates each optional action (serving or charging) at each decision-making slot. The higher the PET's evaluation of an action, the greater the probability of selecting the action at the decision-making slot, which can be formulated as

$$\begin{aligned} a_n^* &= \arg\max_{a_{n,i}} p(a_{n,i}|\{e_0, a_0, \ldots, e_n\}) \\ &= \arg\max_{a_{n,i}} p(a_{n,i}|e_n) \end{aligned} \tag{2}$$

where the conditional probability originally contains the sequence information before the state $e_n$. According to Markov properties, the action at the next moment is only related to the current state and is independent of the state before the current moment, so (2) is simplified as above.

The second transition in (1) is determined by the operating environment where the PET is located. After the PET completes an action, the environment will respond to the action and present new situations to the PET so that the PET transitions to $e_{n+1}$. Here, also according to the Markov characteristic, the environmental state transition $e_n$ to $e_{n+1}$ can be expressed in the form of probability:

$$p(e_{n+1}|e_n, a_n) = \Pr\{E_{n+1} = e_{n+1}|E_n = e_n, A_n = a_n\} \tag{3}$$

Based on the above analysis, the key problem is how to optimize the policy from state to action at each decision slot so that the PET can obtain an optimal action sequence $\{A_0, A_1, A_2, \ldots, A_N\}$ according to the policy to achieve the goal of maximized revenue in the operating cycle:

$$\max_{A_n \in \mathrm{A}} \sum_{n=0}^{N} R_{n+1} \tag{4}$$

where $R_{n+1}$ is the numerical reward (such as the profit obtained from a single service or the cost of a single charge, defined in Section III (24)-(25).) that an operating environment feeds back to the PET after completing the behaviour selected at the decision-making slot $n$. It is worth noting that $R_{n+1}$ is not acquired immediately at slot $n$ but rather at slot $n+1$ after the serving or charging process has been completed, as shown in Fig. 1.

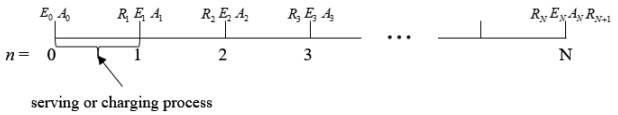



**FIGURE 1.** A short-term operating cycle for a PET.

Since the decision-making problem of a PET within an operating cycle is a sequential decision-making problem that contains multiple decision-making slots, we cannot directly optimize the income of the entire operating cycle according

to (4), so we need to convert (4) into the objective of each decision-making slot:

$$\max_{A_n \in \mathrm{A}} G_n, \quad n = 0, 1, 2, \ldots, N$$
$$s.t.\ G_n = R_{n+1} + \gamma R_{n+2} + \gamma^2 R_{n+3} + \cdots + \gamma^{N-n} R_{N+1} \tag{5}$$

Ideally, the goal of each partial decision-making slot for the PET is to maximize the reward of the entire operating cycle. However, the current and future action decisions will not affect the past rewards, as shown in (5), so the objective of a specific decision-making slot $n$ does not include the rewards $R_1, R_2, \ldots, R_n$ before the present decision-making slot $n$. In (5), $\gamma$ is a parameter where $0 \leq \gamma \leq 1$, which is called a discount rate. The closer it is to 1, the more consideration is given to future rewards.

However, it is still difficult to directly optimize (5) because when the PET is at the decision-making slot, the future state action sequence is unknown, let alone the future income. This can be further expressed in the form of expectations:

$$\begin{aligned} v_\pi(e) &\doteq \mathrm{E}_\pi[G_n | E_n = e] \\ &= \mathrm{E}_\pi[\sum_k^N \gamma^k R_{n+k+1} | E_n = e] \end{aligned} \tag{6}$$

$$\begin{aligned} q_\pi(e, a) &\doteq \mathrm{E}_\pi[G_n | E_n = e, A_n = a] \\ &= \mathrm{E}_\pi[\sum_k^N \gamma^k R_{n+k+1} | E_n = e, A_n = a] \end{aligned} \tag{7}$$

where $\mathrm{E}_\pi[\cdot]$ represents the expected value of a random variable at a given policy $\pi$. Eq. (6) expresses the probabilistic expected value of the rewards obtained by the PET in accordance with the policy $\pi$, starting from state $e$. $v_\pi(e)$ measures the value that state $e$ provides to the realization of goal (4) under the policy $\pi$. Eq. (7) expresses the expected rewards of all possible decision sequences according to the policy $\pi$, starting from the state $e$ and the execution of the action $a$. $q_\pi(e, a)$ measures the value provided by taking action $a$ in state $e$ to achieve goal (4) under the policy $\pi$.

After a series of analyses and derivations, we obtain a more effective expression of the operating decision-making optimization problem for a PET, that is, to solve the optimal policy $\pi(a|e)$ of the PET in a short-term operating process to maximize the value of each state in the operating process:

$$\pi^* = \arg\max_\pi \ v_\pi(e). \tag{8}$$

In other words, the problem is to decide the next action at each decision-making slot to maximize the value of the state of the current decision-making slot:

$$a^* = \arg\max_a \ q_{\pi^*}(e, a) \tag{9}$$

*Remark 1:* It should be noted that the above models and derivations assume Markov characteristics of the PET operation. That is, the probability of the occurrence of each possible value of $E_n$ and $R_n$ only depends on the previous state $E_{n-1}$ and the previous action $A_{n-1}$ and is completely independent of the earlier state and action. The key to solving this problem is to determine the probability of state $E_{n-1}$, action $A_{n-1}$, to the next state $E_n$ and reward $R_n$. In practical application, this probability value is actually an accumulation of experience (data can come from reality or a simulation, and this paper works with the latter).

At this point, we have clearly described the short-term operating decision-making problem for a PET. Next, the model is further refined in Section III and Section IV. Section III will further describe the operating environment of the PET. In addition, Section IV will address the state transitions between decision slots and the calculation of each decision slot's rewards.

## III. MODEL REFINEMENT

### A. ANALYSIS OF PET OPERATING ENVIRONMENT

Obviously, the actual operating environment of a PET is complex and changeable, and there are many factors that can affect the operating decision-making process of the PET, such as weather, road conditions, and so on. Therefore, when analysing the operating environment of PETs, to ensure that the research results are realistic, a comprehensive understanding is required; in addition, to ensure the feasibility and efficiency of solving the problem, environmental information that has less obvious influences on the PET decision-making process should be selectively ignored. Regarding the impacts of operating environments on electric taxis, researchers have compiled statistics and analyses on actual data or simulated data. According to existing studies [21]–[23], this paper selects the environmental factors that have a direct and greater impact on the operating decision and revenue of a PET: passengers, charging stations, transportation, and taxi company management systems, which constitute the operating environment of the PET in this paper. First and foremost, in this paper, the operating space of the PET operating problem is limited to a small city or a specific district of a large city (such as the central business district, residential quarters, residential and commercial mixed area, etc.). When the PET operating area is small, it is assumed that the impact of the PET's spatial position on its operational decision-making process is far less than its temporal position. Therefore, in this paper, only the time-varying characteristics of the environment in the PET operating area are considered. In different spatial locations, the environment has different time-varying characteristics, which has different influences on PET operation. The influence of the time-varying characteristics of these four aspects (passengers, charging stations, transportation, and taxi company management systems) of the environment on PET operation is analysed below.

#### 1) PASSENGER ENVIRONMENT

In different time periods, passenger travel will show different characteristics (such as the total number of passengers, the average travel distance, etc.). The impact of passenger travel time on the PET operating occurs mainly in the process

of PET cruising. In general, the PET cruising distance in peak travel hours shorter than the PET cruising distance in ordinary travel hours.

$$d_{fp}(t) \sim U(\bar{d}_{fp}(t), \sigma_{fp}^2) \tag{10}$$

$$\bar{d}_{fp}(t) = \begin{cases} D_{fp}^p & t \in T_{fp}^p \\ D_{fp}^g & t \in T_{fp}^g \end{cases} \tag{11}$$

The cruising distance at time $t$, $d_{fp}(t)$, is uncertain, so it is assumed that $d_{fp}(t)$ is a random variable subject to a uniform distribution, in which the mean value $\bar{d}_{fp}(t)$ related to the time period, and $D_{fp}^p < D_{fp}^g$.

#### 2) CHARGING STATION ENVIRONMENT

The PET charging process is completed at a charging station, so the environment of the charging station is an important factors affecting the operating decision of a PET. Different charging stations may have completely different load curves and different charging peaks during the day. Whether charging stations are available during the peak electric vehicle charging times will affect PET queueing time $w_q(t)$ before charging; in addition, this can affect the PET charging price $m_e(t)$ (PET charging price generally consists of two parts, energy price and charging station service charge, and the charging station can adjust the service charge according to load conditions within the station).

$$w_q(t) \sim U(\bar{w}_q(t), \sigma_q^2) \tag{12}$$

$$\bar{w}_q(t) = \begin{cases} W_q^p & t \in T_q^p \\ W_q^g & t \in T_q^g \end{cases} \tag{13}$$

$$m_e(t) = \begin{cases} M_e^p & t \in T_e^p \\ M_e^g & t \in T_e^g \end{cases} \tag{14}$$

The queueing time $w_q(t)$ is uncertain, so it is assumed to be a random variable complying with uniform distribution, in which the mean value $\bar{w}_q(t)$ is related to the time period, and $W_q^p > W_q^g$.

#### 3) TRAFFIC ENVIRONMENT

During PET service calls and while driving to the charging station, the PET speed will be affected by the traffic environment (traffic congestion) at all times. Different PET driving speeds will affect the time for a PET to complete a serving process or charging process, thus affecting the revenue per unit time; in addition, this will affect the energy consumption of the PET $f_d$ (previous studies have shown that the energy consumption of a PET is correlated with its average speed, which can be fitted by (16)).

$$v(t) = \begin{cases} V^p & t \in T_v^p \\ V^g & t \in T_v^g \end{cases} \tag{15}$$

$$f_d = k1 + k2 \cdot v(t) + k3/v(t) \tag{16}$$

where Eq. (16) is a basic form commonly used to fit the relationship between energy consumption and average PET speed. $k1$, $k2$ and $k3$ are fitting coefficients, whose values can be obtained by multiple linear regression based on a large amount of actual data.

#### 4) TAXI COMPANY MANAGEMENT SYSTEM

Different taxi companies will have different management and operating systems. Electric taxis need to operate according to the regulations of their companies. In a taxi company system, the shift change system (such as the shift change time, the need for surplus electricity during the shift) and the passenger fees rules (taxi starting price, unit kilometre price, night operation subsidy, etc.) will have a direct impact on the decision and income of the PET.

$$m_s(t) = \begin{cases} M_s^{dt} & t \in T_s^{dt} \\ M_s^{nt} & t \in T_s^{nt} \end{cases} \tag{17}$$

It is notable that the unit kilometre price $m_s(t)$ here is only a part of the passenger fees that will change over time, while the complete passenger fees (that is, PET serving revenue) will be given in Section III.

### B. PET MODEL

The following is a further refinement of the PET model. Before refining the state model of the PET, we first need to distinguish the objective environmental state and the observed environmental state of the PET. An objective operating environment for the PET demands that the environment will not be changed by subjective observation or cognition of the PET. An observed PET environment, as the name implies, refers to the subjective environment of the PET. The environment in Section II and the four parts of the PET operating environment discussed in Section III belong to an objective environment. In this section, to further refine the PET decision-making model, it is necessary to discuss the environment observed by the PET (hereinafter referred to as PET state). The PET state is an abstract expression of environmental information that the PET can and must obtain. The obtainable information for the PET depends on the actual background, such as whether the environmental information is disclosed and whether the technology and equipment conditions supporting the information transmission are available. However, the information required by the PET depends on which environmental information will have a greater impact on the PET operating decisions. This study solely focuses on how selecting the state in the actual application should be considered and adjusted in combination with the background of the actual problem. However, this study is a relatively universal model, focusing on the decision-making method itself. Therefore, this paper does not seriously consider the real background but rather works from the perspective of algorithm implementation and effectiveness. Finally, in this paper, time and SOC of the PET are selected to constitute the PET states. In other words, the two parameters of time and SOC are used to identify the PET state at each decision slot:

$$S \sim [S_{time}, S_{energy}] \tag{18}$$

where, $S_{time}$ is the time state and is set as a state every 30 minutes. When taking a day as an operating cycle, there are 48 time states in an operating cycle. $S_{energy}$ is the SOC of the PET. The total electric quantity of the PET is divided into 25 states. That is, there are $48 \times 25 = 1200$ PET states.

At each decision-making slot, the PET has to choose the action of the next process based on the current state $S$. The PET has two optional behaviours: serving or charging.

$$\mathrm{A} = \{a_s, a_c\} \tag{19}$$

where, $a_s$ indicates that the PET decides to offer service as the next process at the current decision-making slot, and $a_s = 1$; $a_c$ indicates that the PET decides to seek a charge as the next process at the current decision slot, and $a_c = 1$.

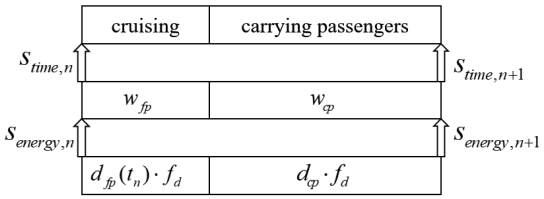



**FIGURE 2.** State transition when selecting serving.

### 1) STATE TRANSITION

The following is a further analysis of how the state of the adjacent decision slot is transformed, that is, the expression of (3): how to obtain $s_{n+1}$ according to $s_n$ and $a_n$. As briefly analysed above, there is a continuous process of serving or charging between two adjacent decision slots (depending on whether the PET chooses serving or charging behaviour at the previous decision slot). Therefore, analysing the state transition between the decision-making slots is essentially analysing which changes ($S_{time}$ state and $S_{energy}$ state transitions) occur to the state of the PET $S$ during the process of serving or charging. If the PET chooses service, the serving process (the serving process is composed of two processes of cruising and carrying passengers, as shown in Fig. 2) will take place. In addition, the $S_{time}$ state and $S_{energy}$ state transition of the PET are as follows:

$$s_{time,n+1}(s_{time,n}, a_s) = floor(\frac{t_n + w_{fp} + w_{cp}}{30}) \tag{20}$$

$$s.t. \int_{t_n}^{t_n+w_{fp}} v(t)dt = d_{fp}(t_n), \int_{t_n+w_{fp}}^{t_n+w_{fp}+w_{cp}} v(t)dt = d_{cp}$$

$$s_{energy,n+1}(s_{energy,n}, a_s) = floor(\frac{e_n - d_{fp}(t_n) \cdot f_d - d_{cp} \cdot f_d}{e_{\max}/25}) \tag{21}$$

If the PET chooses charging, the charging process (the charging process consists of three processes: driving to a charging station, queueing before charging, and connecting to the power grid for charging, as shown in Fig. 3) will take place. In addition, the $S_{time}$ state and $S_{energy}$ state transition of

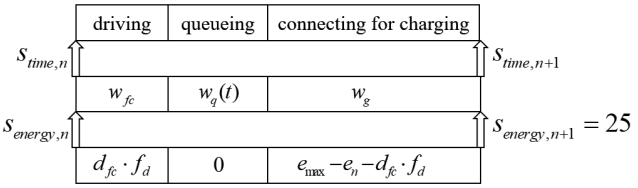



**FIGURE 3.** State transition when selecting charging.

the PET are as follows:

$$s_{time,n+1}(s_{time,n}, a_c) = floor(\frac{t_n + w_{fc} + w_q(t) + w_g}{30}) \tag{22}$$

$$s.t. \int_{t_n}^{t_n+w_{fc}} v(t)dt = d_{fc}, w_g = \frac{e_{\max} - e_n}{p_c}$$

$$s_{energy,n+1}(s_{energy,n}, a_c) = 25 \tag{23}$$

In Eqs. (20) and (22), $floor(\cdot)$ is the rounding down function. Eqs. (20) and (22) refer to the PET $S_{time}$ state after the PET completes a serving action or a charging action, respectively. It should be noted that the time s $w_{fp}$, $w_{cp}$, and $w_{fc}$ in Eqs. (20) and (22) are calculated based on the distance and speed of the corresponding action. The time $w_q(t)$ is obtained according to Eqs. (12) and (13), while the charging time $w_g$ is calculated by assuming that the PET is fully charged in every charging action. Eqs. (21) and (23) refer to the PET $S_{energy}$ state after the PET completes a serving action or a charging action, respectively. Since it is assumed that the PET is fully charged after each charge, the PET is at maximum $S_{energy}$ state after each charging action, i.e., $S_{energy} = 25$.

### 2) REWARD SIGNAL

Next, the reward signal from the environment after the PET completes a serving process or a charging process is defined:

$$r_{n+1}(s_n, a_s, s_{n+1}) = m_{sf} + \int_{t_n+w_{fp}}^{t_n+w_{fp}+w_{cp}} m_s(t)v(t)dt \tag{24}$$

$$r_{n+1}(s_n, a_c, s_{n+1}) = -\int_{t_n+w_{fc}+w_q(t)}^{t_n+w_{fc}+w_q(t)+w_g} m_e(t)p_c dt \tag{25}$$

where Eq. (24) actually refers to the carrying income obtained in a serving process, while Eq. (25) refers to the total charging cost.

## IV. SOLUTION ALGORITHM

Thus far, we have introduced the model in detail, but this is still far from the solution to the final problem. According to the analysis concept of Eqs. (2)-(9) in Section II, we need to list all possible state transitions and specify each state-transition probability before calculating Eqs. (6) and (7) to obtain the optimal strategy. However, after further refinement of the model in Section II, we find that it is nearly unreachable to list all possible state transitions (1200 PET states have been set above, which means that $1200^2 = 1.44 \times 106$ state-transition probabilities need to be set or calculated.

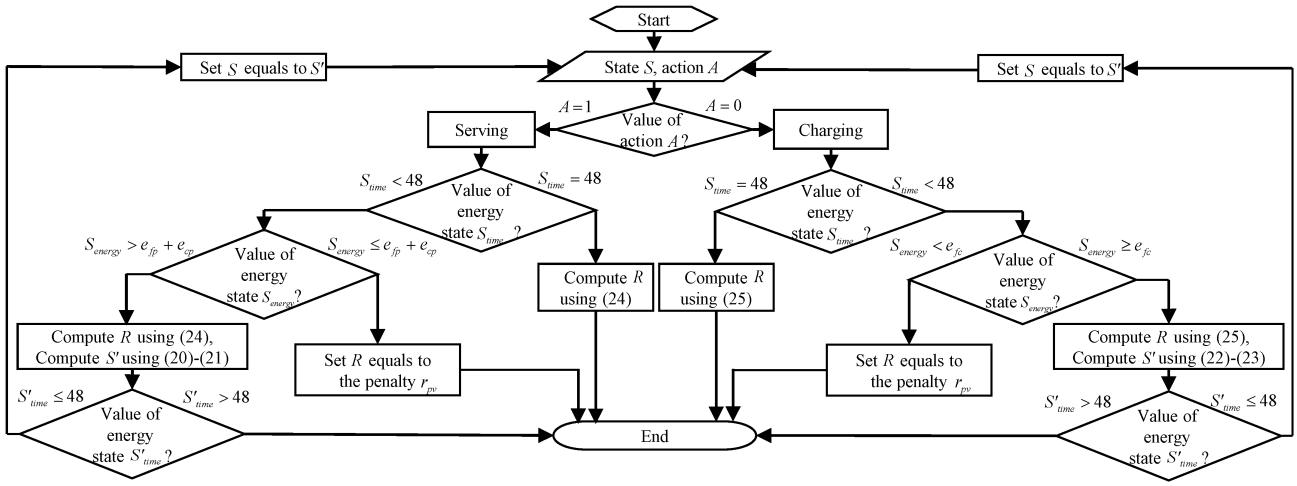



**FIGURE 4.** Sample generation process.

In addition, the state transition process contains many random variables, such as $d_{fp}(t_n)$, $d_{cp}$, $d_{fc}$ and so on). Therefore, we cannot apply ''model-based'' algorithms (such as dynamic programming) to solve this problem. We can only choose a ''model-free'' algorithm (which approximates the expected values in (6) and (7) by simulating enough sample sequences). In this paper, the commonly used ''model-free'' algorithm SARSA is selected to solve this problem. SARSA is a temporal difference (TD) learning algorithm, which combines Monte Carlo simulation and dynamic programming [24]. From the point of view of the algorithm structure, it is similar to the Monte Carlo method, which is also solved by simulating the interactive sequence from the core of the algorithm; this approach also uses the classic Bellman equation in reinforcement learning to realize self-iteration. As its name implies, the five key iteration factors of the algorithm are $S$ (the current state), $A$ (the current action), $R$ (the reward obtained by the simulation), $S'$ (the next state entered by the simulation) and $A'$ (the next action taken in the simulation). The core iteration equation for SARSA is as follows:

$$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma Q(S',A') - Q(S,A)] \quad (26)$$

The pseudocode of the algorithm is as follows.

The key to the algorithm implementation lies in Step 3, that is, how to simulate an episode of a sequence sample. There are two key points to implement Step 3: one is the policy used to simulate the sample, and the other is the definition of the final state of an episode. As shown in the pseudocode above, this paper uses the most basic and commonly used random policy in reinforcement learning algorithms, that is, an $\varepsilon$-greedy policy. This kind of policy may well balance exploitation and exploration.

$$\pi(a|s) = \begin{cases} 1 - \dfrac{\varepsilon}{2} & a = \arg\max_a \ Q(s,a) \\ \dfrac{\varepsilon}{2} & a \neq \arg\max_a \ Q(s,a) \end{cases} \quad (27)$$

**Algorithm 1** Solution to (4)-(9)

1: **Step 1:** Initialization. $Q(s,a), \forall s \in S, a \in A(s)$.
2: **Step 2:** Setting. Set values of $\alpha$ and $\gamma$.
3: **Step 3:** An episode. Set a starting state $S$.
   1) Choose action $A$ from state $S$ using policy derived from $\varepsilon$-greedy.
   2) Take action $A$, observe $R$, $S'$.
   3) Choose action $A'$ from state $S'$ using policy derived from $\varepsilon$-greedy.
   4) Update
   $$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \gamma Q(S',A') - Q(S,A)].$$
   5) Update $S \leftarrow S', A \leftarrow A'$.
   6) Go to 3) and repeat this process until $S$ is terminal.
4: **Step 4:** Repeating. Go to **Step 3**, repeat this process until all $Q(s,a)$ converge.
5: **Step 5:** Output the final policy $\pi(s) = \arg\max_a Q(s,a)$.

Eq. (27) indicates that the probability of selecting the action with the maximum action value is $1 - \frac{\varepsilon}{2}$, while the probability of another action is $\frac{\varepsilon}{2}$. This kind of policy may well balance exploitation (select the action with maximum action value) and exploration (select action other than the action with maximum value function).

Step 3 obtains an episode of data through cycles 1) to 6) until the PET reaches the final state. However, there are several different final states. Fig. 4 shows a more detailed loop logic of Step 3 and the scenarios where an episode ends. As shown in Fig. 4, the first kind of final state occurs when $S_{time} \geq 48$, that is, the time state $S_{time}$ reaches the end of the operating day. The second and third kinds occur when $S_{energy} \leq e_{fp} + e_{cp}$ and $S_{energy} < e_{fc}$ respectively, that is, the PET has chosen a serving action but without enough energy to complete cruising and carrying passengers, or the PET has chosen a charging action but without enough energy

to reach the charging station. In this paper, the second and third conditions are collectively referred to PET operating failure. Obviously, the essential reason for an operating failure is that the PET did not choose the charging action in time and in advance. A penalty signal $r_{pv}$ should be set here to avoid such an unreasonable policy.

## V. SIMULATION AND ANALYSIS

In this section, we use MATLAB to simulate and verify the effectiveness and superiority of the method proposed in this paper. First, the basic parameters of the simulation were explained. Then, according to the convergence and optimization of the algorithm, several key parameters of the SARSA algorithm were determined. Finally, the simulation results were compared with a conventional electric taxi operating scheme (scheme without SARSA algorithm).

### A. PARAMETER SETTING

We examine a PET operating during a day (24 h) starting at 5:00 AM. It should be noted that the parameters in this paper are determined by reasonable assumptions based on real data studies [21]–[23]. First, we set the basic parameters for the PET. Vehicle battery capacity $e_{\max}$ is set at 25 kWh, and charging power $p_c$ is set at 12 kW. According to (16), the corresponding energy consumption per kilometre at different driving speeds can be calculated. Assuming that $V^p = 12km/h$ in traffic congestion, the corresponding energy consumption per kilometre $f_d = 0.3kWh/km$; assuming that the general driving speed $V^g = 36km/h$, then $f_d = 0.2kWh/km$.

**TABLE 1. Peak hour settings.**

| $T_{fp}^p$ | $T_q^p$ | $T_e^g$ | $T_v^p$ | $T_s^{nt}$ |
|---|---|---|---|---|
| 8:00-9:00, 17:00-19:00 | 3:00-5:00, 16:00-18:00 | 0:00-7:00 | 8:00-9:00, 17:00-19:00 | 23:00-5:00 |

Operating environment parameters (such as peak passenger travel time, traffic congestion time, charging peak time, etc.) depend on the characteristics of the PET operating location (central business district, residential quarters, residential and commercial mixed area, etc.). In this example, we set the PET operating area as a central business district. According to the conclusions of existing studies, the time periods for peak passenger travel, traffic congestion, and charging in this region are set as shown in Table 1. In these peak hours in Table 1 and ordinary times beyond the peak hours, the characteristics of the PET operating environment are different, and their corresponding cruising distance, queuing time, charging unit price, and so on are different. When $t \in T_{fp}^p$, the cruising distance $d_{fp}(t)$ (in kilometres) follows uniform distribution $U(2, 1.33)$; otherwise, $d_{fp}(t) \sim U(6, 1.33)$. When $t \in T_q^p$, the queuing time $w_q(t)$ (in minutes) follows uniform distribution $U(20, 33.33)$; otherwise, $w_q(t) = 0$. During off-peak charging price hours $T_e^g$, PET's unit charging price $M_e^g$ is set to be 0.9 RMB/kWh, and during peak hours, $M_e^p$ is set to be RMB/kWh. During night hours $T_s^{nt}$, PET's unit kilometres price $M_s^{nt}$ is set to be 3 RMB/km, and during other hours, $M_s^{dt}$ is set to be 2.5 RMB/km.

### B. ALGORITHM CONVERGENCE

In this section, we discuss the values of SARSA algorithm related parameters, including discount rate $\gamma$, learning rate $\alpha$, penalty value $r_{pv}$, and probability $\varepsilon$, in an $\varepsilon$-greedy policy and select the parameter value that makes the result more optimal. Through the preliminary experimental comparison, we find that the value of $\gamma$ and $\alpha$ will have an impact on the optimization. The smaller the value of $\alpha$ is, the more convergent the $Q(s, a)$ will be, but the slower the convergence rate will be. The penalty value mainly affects the failure rate of an episode simulation. The smaller the penalty value is, the smaller the failure rate is. However, $\alpha$ and $r_{pv}$ hardly affect the optimization.

After comparing the results of different parameter combinations, the parameter values are finally selected as follows: $\gamma = 1$, $\alpha = 0.05$, $\varepsilon = 0.05$, $r_{pv} = -10^3$. Based on these parameters, we respectively took 25 SOC states as the initial state to conduct 3000 rounds of PET daily operating simulations. The convergence of the algorithm is judged according to whether the $Q(s, a)$ of state-action pair is stable. After 3000 rounds of simulations, in addition to $S_{energy} = 1, 2, 3$, the $Q(s, a)$ of other state-action pairs tends to be stable, from which the $Q(s, a)$ of $S_{energy} = 15$ and $S_{energy} = 25$ are selected to be displayed in Fig. 5.

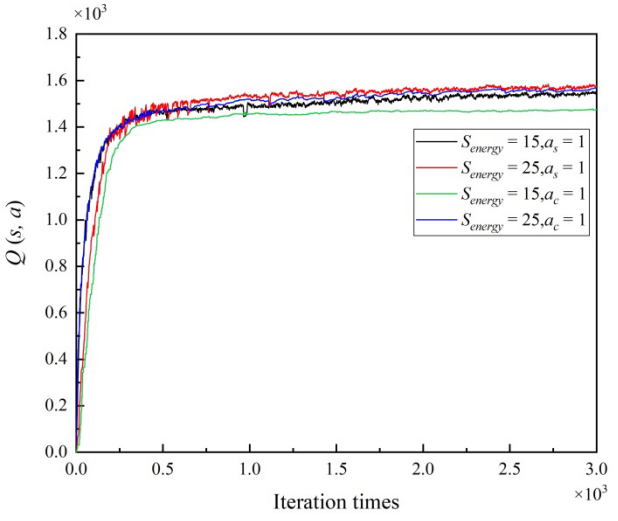



**FIGURE 5. Convergence of SARSA algorithm.**

### C. STATISTICAL PERFORMANCE

In this section, we compare the results of the optimal operating strategy proposed in this paper with the common operating strategy. To obtain a statistical result, each of the following figures is obtained via a Monte Carlo simulation that consists of 100 independent trials. In the common

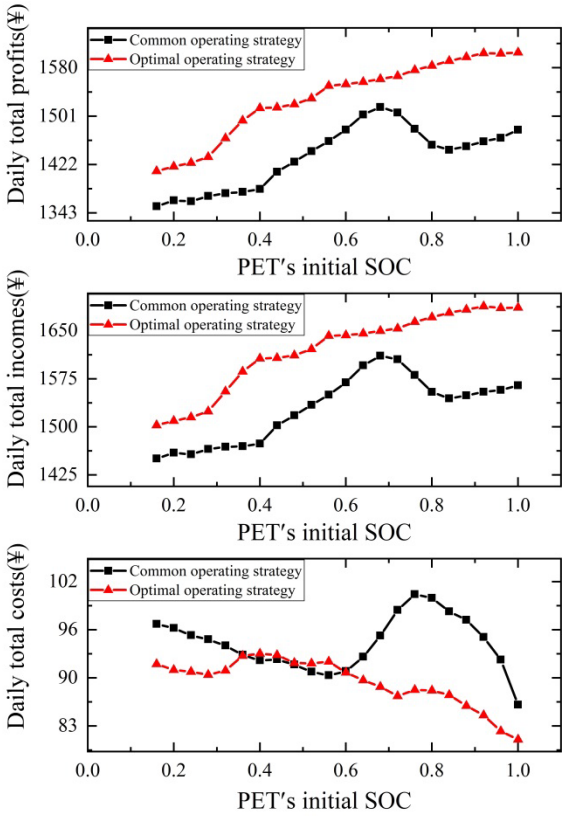



**FIGURE 6.** The daily profits, incomes, and costs with different initial SOCs by applying common operating strategy and optimal operating strategy.

operating strategy specifically mentioned here, when the remaining battery capacity of the PET is less than a certain fixed value (set it to 0.3 here), the owner will choose to charge, and in other cases, it will choose to serve, completely ignoring the environment changes (This strategy is often used to make assumptions about EV charging behaviour.). Fig. 6 shows the daily profits, incomes, and costs with different initial SOCs for the PETs. It can be observed that no matter how much the initial SOC of the PET is, the optimal operating strategy proposed in this paper can improve the daily profits of the PET. It can also be found that the initial SOC has a great impact on the daily profits, incomes, and costs, and the increase in profits varies with the initial SOC. The maximum increase is approximately 10% when the initial SOC is 0.84, and the minimum increase is approximately 3% when the initial SOC is 0.68.

Fig. 7 shows the optimal policy, which states when the PET chooses to charge and when it chooses to serve. Charging selection is selected when the charging action value $Q(s, a_c)$ is greater than the serving action values $Q(s, a_s)$ and vice versa. It should be noted that both the charging selection and serving selection are divided into strong and weak selection. Weak selections essentially mean that there is no significant difference between $Q(s, a_c)$ and $Q(s, a_s)$ in these states (the single selection in these states has little impact on the profit of the entire operating process). In the case of strong selections, $Q(s, a_c)$ and $Q(s, a_s)$ are much different, and a single selection

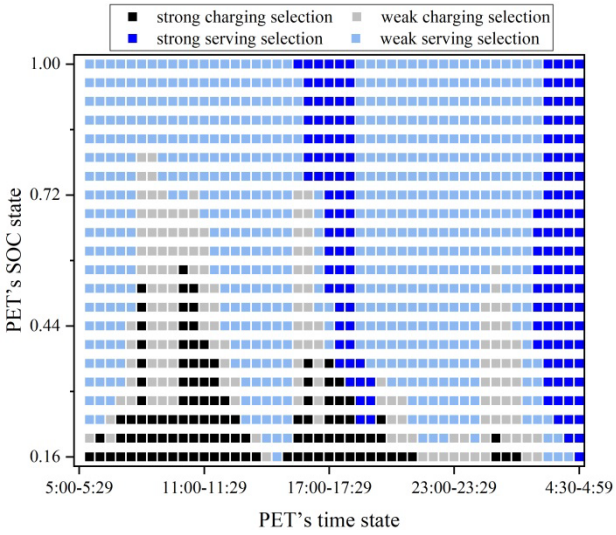



**FIGURE 7.** The optimal operating policy.

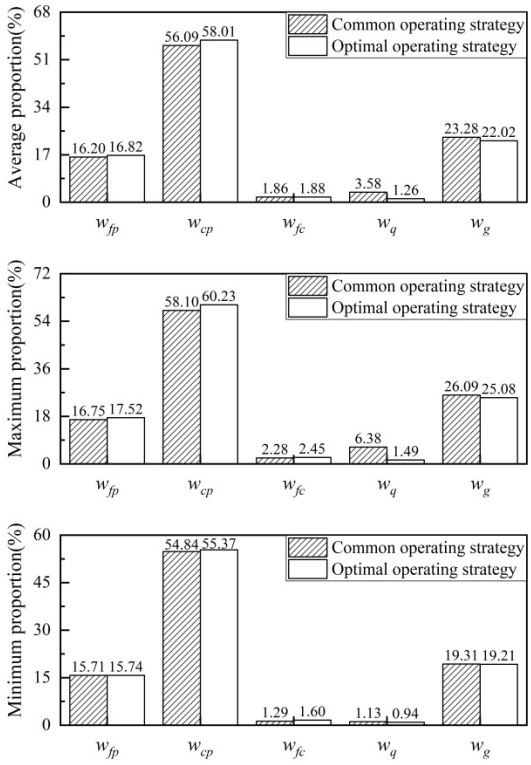



**FIGURE 8.** Average, maximum, and minimum time proportions of each operating subprocess under the common strategy and the optimal strategy.

in these states can have a significant impact on the profit of the entire operating process. As shown in Fig. 7, there are three black and grey areas in the figure. When the PET is in these states, the charging action needs to be selected.

Fig. 8 and Fig. 9 respectively compare the time proportion and energy consumption per unit kilometre of each operating subprocess under the optimal strategy and the common strategy. As seen in Fig. 8, the optimal strategy mainly increases the proportion of passenger carrying time by

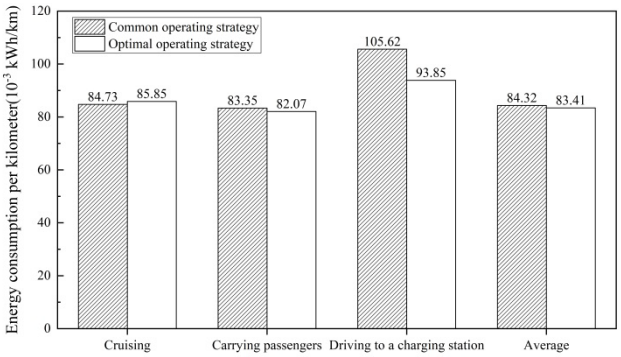



**FIGURE 9.** The energy consumption per kilometre of each operating subprocess under the common strategy and the optimal strategy.

reducing the queuing time in an operating cycle. It can be seen in Fig. 9 that the optimal strategy not only increases the daily profit of the PET owner through a reasonable time arrangement of charging and serving but also reduces the energy consumption per unit kilometre for the PET and increases energy savings.

## VI. CONCLUSION AND DISCUSSION

This paper investigates the optimal decision-making problem of serving actions and charging actions for an individual PET subject to a time-varying uncertain external environment to maximize its average profit in a short-term operating cycle. To optimize the applicable strategy even when the external environment of the PET is unknown, the external environment is analysed and modelled in terms of four aspects: passengers, charging stations, traffic, and taxi company management systems. To make the model more objective and realistic, the serving and charging processes in the PET operating cycle are further refined into multiple processes of cruising, carrying passengers, driving to the charging station, queueing before charging, and connecting to the power grid for charging. Then, the transitions between adjacent processes and the reward signals from the environment are modelled for model refinement. For this sequential decision-making problem, which contains many uncertain factors, the SARSA algorithm is selected to solve it.

A series of experiments showed the following results. First, the initial SOC has a great impact on the daily profits, incomes, and costs, and the increase in profits varies with the initial SOC. Second, the proposed strategy can improve the short-term operating profit compared with the ordinary strategy under any initial PET SOC condition. Third, the proposed strategy mainly increases the proportion of passenger carrying time by reducing the queuing time in an operating cycle. Fourth, the proposed strategy not only increases the daily profit of the PET owner through reasonable time arrangement of charging and serving but also reduces the energy consumption per unit kilometre of PET and increases energy savings.

However, this paper only considers the behavioural decisions of a single PET and the temporal characteristics of the external environment, which is only a preliminary study of PET operating problems. The model proposed in this paper is extendable and can be based on the proposed model to study the decision-making strategy of a PET and the dispatching strategy of a PET fleet under an environment with complex temporal-spatial characteristics. In addition, from the perspective of power grid operators, the model in this paper could be used as a power load model that reflects autonomous PET decision-making processes. Further, based on this load model, the flexibility of a PET participating in demand responses and the design of charging price mechanisms considering PET behavioural uncertainty can be studied.

## REFERENCES

[1] International Energy Agency (IEA), Paris, France, 2013. *Global EV Outlook-Understanding the Electric Vehicle Landscape to 2020*. [Online]. Available: http://www.iea.org/publications/globalevoutlook_2013.pdf

[2] Y. P. Yang, Y. S. Chen, X. J. Wang, and R. S. Yin, "The major problems needed to be resolved in the development of electric vehicles at the present stage in China," *Adv. Mater. Res.*, vols. 779–780, pp. 991–995, 2013.

[3] N. O. Kapustin and D. A. Grushevenko, "Long-term electric vehicles outlook and their potential impact on electric grid," *Energy Policy*, vol. 137, Feb. 2020, Art. no. 111103.

[4] J. Liu and C. Zhong, "An economic evaluation of the coordination between electric vehicle storage and distributed renewable energy," *Energy*, vol. 186, Nov. 2019, Art. no. 115821.

[5] M. Ahmadi, S. H. Hosseini, and M. Farsadi, "Optimal allocation of electric vehicles parking lots and optimal charging and discharging scheduling using hybrid Metaheuristic algorithms," *J. Electr. Eng. Technol.*, vol. 16, no. 2, pp. 759–770, Mar. 2021.

[6] M. F. M. Arani and Y. A.-R.-I. Mohamed, "Cooperative control of wind power generator and electric vehicles for microgrid primary frequency regulation," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5677–5686, Nov. 2018.

[7] K. R. Reddy and S. Meikandasivam, "Load flattening and voltage regulation using plug-in electric vehicle's storage capacity with vehicle prioritization using ANFIS," *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 260–270, Jan. 2020.

[8] X. Zhou, S. Zou, P. Wang, and Z. Ma, "Voltage regulation in constrained distribution networks by coordinating electric vehicle charging based on hierarchical ADMM," *IET Gener., Transmiss. Distrib.*, vol. 14, no. 17, pp. 3444–3457, Sep. 2020.

[9] D. Chen, Z. Jing, and H. Tan, "Optimal bidding/offering strategy for EV aggregators under a novel business model," *Energies*, vol. 12, no. 7, p. 1384, Apr. 2019.

[10] J. Su, T. T. Lie, and R. Zamora, "Modelling of large-scale electric vehicles charging demand: A New Zealand case study," *Electr. Power Syst. Res.*, vol. 167, pp. 171–182, Feb. 2019.

[11] I. Morro-Mello, A. Padilha-Feltrin, and J. D. Melo, "Spatial-temporal model to estimate the load curves of charging stations for electric vehicles," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf.-Latin Amer. (ISGT Latin Amer.)*, Sep. 2017, pp. 1–6.

[12] M. B. Arias and S. Bae, "Electric vehicle charging demand forecasting model based on big data technologies," *Appl. Energy*, vol. 183, pp. 327–339, Dec. 2016.

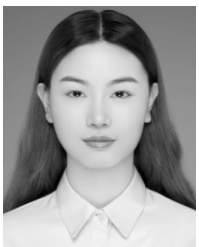
[13] B. Zhou, X. Yang, D. Yang, Z. Yang, T. Littler, and H. Li, "Probabilistic load flow algorithm of distribution networks with distributed generators and electric vehicles integration," *Energies*, vol. 12, no. 22, p. 4234, Nov. 2019.

[14] J.-M. Clairand, P. Guerra-Terán, X. Serrano-Guerrero, M. González-Rodríguez, and G. Escrivá-Escrivá, "Electric vehicles for public transportation in power systems: A review of methodologies," *Energies*, vol. 12, no. 16, p. 3114, Aug. 2019.

[15] A. Pan, T. Zhao, H. Yu, and Y. Zhang, "Deploying public charging stations for electric taxis: A charging demand simulation embedded approach," *IEEE Access*, vol. 7, pp. 17412–17424, 2019.

[16] C. Jiang, Z. Jing, T. Ji, and Q. Wu, "Optimal location of PEVCSs using MAS and ER approach," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 20, pp. 4377–4387, Nov. 2018.

[17] Z. Yang, L. Sun, J. Chen, Q. Yang, X. Chen, and K. Xing, "Profit maximization for plug-in electric taxi with uncertain future electricity prices," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 3058–3068, Nov. 2014.
[18] Z. Yang, L. Sun, M. Ke, Z. Shi, and J. Chen, "Optimal charging strategy for plug-in electric taxi with time-varying profits," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2787–2797, Nov. 2014.
[19] Z. Tian, T. Jung, Y. Wang, F. Zhang, L. Tu, C. Xu, C. Tian, and X.-Y. Li, "Real-time charging station recommendation system for electric-vehicle taxis," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 11, pp. 3098–3109, Nov. 2016.
[20] J. Yang, Y. Xu, and Z. Yang, "Regulating the collective charging load of electric taxi fleet via real-time pricing," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3694–3703, Sep. 2017.
[21] C. X. Jiang, Z. X. Jing, X. R. Cui, T. Y. Ji, and Q. H. Wu, "Multiple agents and reinforcement learning for modelling charging loads of electric taxis," *Appl. Energy*, vol. 222, pp. 158–168, Jul. 2018.
[22] 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, Apr. 2016.
[23] Z. He, Y. Cheng, and Z. Hu, "Multi-time simulation of electric taxicabs' charging demand based on residents' travel characteristics," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Nov. 2017, pp. 1–6.
[24] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.



**YANG YOU** received the B.E. degree in electrical engineering from the South China University of Technology, Guangzhou, China, in 2016, where she is currently pursuing the Ph.D. degree in electrical engineering. Her research interests include load modeling and optimal dispatching for plug-in electric taxis in a power grid.

**JISONG ZHU** received the B.A. degree in electrical engineering from Zhengzhou University, Zhengzhou, in 2015, and the M.Sc. degree from the School of Electric Power Engineering, South China University of Technology, Guangzhou, in 2018. He is currently pursuing the Ph.D. degree. His research interests include the simulation and modeling of energy systems and power markets.

**YICHUAN HUANG** received the M.Sc. degree in control engineering from the South China University of Technology, Guangzhou, China, in 2017, where he is currently pursuing the Eng.D. degree with the Smart Grid and Its Automation Team of Energy Research Institute. He is also a member of the Smart Grid and Its Automation Team of Energy Research Institute, South China University of Technology. His research interests include the simulation and modeling of energy systems, intelligent control, and electrical engineering.

**ZHAOXIA JING** (Member, IEEE) received the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2003. She is a currently a Professor with the School of Electric Power Engineering, South China University of Technology. Her research interests include electricity markets, integrated energy system optimization, and electric vehicles.