

Eco-Driving With Deep Reinforcement Learning at Signalized Intersections Considering On-the-Fly Queue Dissipation Estimation and Lane-Merging Disturbances

XINXING REN ¹ (Graduate Student Member, IEEE), CHUN SING LAI ¹ (Senior Member, IEEE),
GARETH TAYLOR ¹ (Senior Member, IEEE), AND YUJIE YUAN ² (Member, IEEE)

¹Department of Electronic and Electrical Engineering, Brunel University of London, UB8 3PH London, U.K.

²School of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China

CORRESPONDING AUTHORS: CHUN SING LAI; YUJIE YUAN (e-mail: chunsing.lai@brunel.ac.uk; yjyuan@cauc.edu.cn).

This work was supported by the Tianjin Municipal Science and Technology Bureau Science and Technology Program – Natural Science Foundation – Multi-Fund Program – Youth Project under Grant 24JCQNJC00280.

ABSTRACT Eco-driving research has grown significantly over the past decade, increasingly incorporating real-world traffic and road conditions such as road gradients, lane changes, and queue effects. However, most existing studies that account for queue effects are limited to single-lane scenarios, without considering lane-merging disturbances, and can only estimate queue length or discharge time within restricted regions. To address these limitations, this paper proposes a novel deep reinforcement learning (DRL) based eco-driving algorithm that simultaneously considers on-the-fly queue dissipation time estimation and lane-merging disturbances. The approach integrates a practical and cost-effective navigation-app-based traffic data sharing framework with a data-driven dissipation time estimation model, enabling the reinforcement learning agent to continuously receive accurate modified reference speeds that reflect both queueing and merging vehicle effects. Five comprehensive case studies, benchmarked against conventional and state-of-the-art eco-driving methods, were conducted to evaluate the effectiveness of the proposed approach. Simulation results demonstrate that the proposed method consistently achieves the best energy performance across all scenarios, reducing energy consumption by an average of 37.5% compared with the Intelligent Driver Model (IDM) baseline.

INDEX TERMS Eco-driving, dissipation time estimation, connected vehicle, deep reinforcement learning, deep learning.

NOMENCLATURE

Symbols and acronyms

A_t	Action space of DRL-based eco-driving.	$-c, c$	Action space limits of TD3 model.
a	Action of the Twin Delayed Deep Deterministic (TD3) model.	D_{e2s}	Distance from ego-vehicle to signalized intersection. (m)
a_1, a_2, a_3, a_4	Weighting factors in the reward function.	D_{e2l}	Distance from ego-vehicle to the last leading vehicle. (m)
a_t	Action of DRL-based eco-driving	D_{l2s}	Distance from the last leading vehicle to the signalized intersection. (m)
a_{IDM}	Acceleration provided by IDM. (m/s^2)	E_{bat}	Instantaneous electricity consumed by the battery pack. (W)
a_{lea}	Acceleration of the last leading vehicle. (m/s^2)	i	Index of Q-value functions of TD3 model.
		L	Loss function of the TD3 model.

N	Number of leading vehicles.
n	Index of leading vehicles.
Pos_m	Position of surrounding vehicle m . (m)
r	Instant reward of TD3 model.
Q_{θ_1} and Q_{θ_2}	Q-value functions of TD3 model.
R_t	Reward function of DRL agent
R_{ref}	Reward function of modified reference speed.
R_{ene}	Reward function of energy consumption.
R_{saf}	Reward function of safety.
R_{eff}	Reward function of traffic efficiency.
S_t	State space of the DRL agent.
s, s'	Current state and the next state of TD3 model.
t_{rem}	Remaining time of current phase. (s)
t_g	Duration of green phase. (s)
t_{l2t}	Dissipation time of the last leading vehicle. (s)
t_{cur}	Current time step. (s)
t_{pas}	Time step of the last leading vehicle pass signalized intersection. (s)
t_{rem2g}	Remaining time until the next green phase. (s)
V_{lim}	Speed limit. (m/s)
V_{ave}	Average speed of leading vehicles. (m/s)
V_{lea}	Speed of the last leading vehicle. (m/s)
V_{max}, V_{min}	Maximum and minimum reference speed, respectively. (m/s)
$V_{max}^{mod}, V_{min}^{mod}$	Modified maximum and minimum reference speed, respectively. (m/s)
V_m	Speed of surrounding vehicle m . (m/s)
V_n	Speed of leading vehicle n . (m/s)
V_t	Ego-vehicle speed. (m/s)
$VehID_m$	Vehicle ID of surrounding vehicle m .
y	Target Q value of the TD3 model.
β_1, β_2	Cut-of points.
γ	Discount factor in the TD3 model.
w_1, w_2	Weighting coefficients in the multi-objective function of eco-driving.
θ', θ''	Parameters of target networks of TD3 model.
μ	Policy of TD3 model.

I. INTRODUCTION

Driven by concerns over the gradual depletion of fossil energy and greenhouse gas emissions, electric vehicle (EVs) are progressively replacing traditional gasoline or diesel cars, becoming the dominant force in the vehicle market. Numerous countries, such as the United States, the United Kingdom, China, and other nations in Europe, have developed various strategies and dedicated substantial funding to encourage the widespread adoption of electric vehicles [1], [2]. Currently, with battery technology yet to mature, range anxiety remains a significant barrier to the widespread acceptance of electric vehicles [3]. In recent years, eco-driving, a technology aimed

at enhancing vehicle energy efficiency and lowering emissions, has attracted widespread interest and been extensively researched within the electric vehicle sector [4]. Galvin et al. [5] conducted EVs in relation to their acceleration and speed across eight common models on the market. The findings revealed that the optimal operating speed for electric vehicles ranges between 43.26 km/h to 60 km/h, with a preference for lower acceleration levels. Utilizing this principle, eco-driving strategies plan speed trajectories for EVs and avoid needless halts and idling [6], [7].

Eco-driving can be categorized into two scenarios: highway eco-driving and urban eco-driving. Highway eco-driving primarily focuses on maintaining a steady headway distance from the vehicle ahead to save energy [8], [9], [10], which is relatively straightforward. Urban eco-driving, on the other hand, involves not only considering car-following behavior but also taking into account the signal phase and timing (SPaT) information at signalized intersections [11], [12], [13], [14], [15]. Initial research indicates that the eco-driving method can achieve approximately a 40% energy savings compared to cruise control when considering SPaT information [16]. This approach aims to enable EVs to smoothly pass through signalized intersections without idling, thereby saving energy while ensuring traffic efficiency, so that eco-driving at signalized intersections are also known as eco-approach and departure (EAD) [17]. Meanwhile, numerous advanced studies have jointly considered energy management systems (EMS) and eco-driving [18], [19].

However, in practice, idling events are influenced not solely by SPaT information but also by the presence of vehicle queues at the approaches to signalized intersections [20]. An EV equipped with eco-driving technology that does not account for queue information will experience significant speed fluctuations when encountering queued vehicles, though it won't come to a complete stop before a signalized intersection as an EV without eco-driving would. However, the renowned ring road experiment [21] demonstrates that even minor speed oscillations intensify over time, preventing the maintenance of uniform traffic flow. Eventually, this leads to traffic disruption and more energy consumption.

Therefore, incorporating queue information into eco-driving algorithms is crucial, and recently, many researchers have begun to address this issue. Generally, eco-driving algorithms that consider queue information can be divided into two layers. The first layer predicts the queue length and queue discharge time, then calculates an adjusted window for passing through traffic lights or computes a modified reference speed considering predicted queue length and predicted discharge time. Subsequently, the second layer of speed planning algorithms takes this provided information to map out a speed trajectory for the EV.

Table 1 summaries the eco-driving algorithms considering the effect queue vehicles. It should be noted that in this paper, the term 'discharge time' refers to the time between when the

TABLE 1. A Summary of Eco-Driving Algorithms Considering the Effects of Queue Vehicles

Algorithm	Queue information source	Queue estimation methodology	Various queue type	Speed planning methodology	Car-following	Lane-Merging Disturbances	On-the-fly Queue Dissipation Estimation
[22]	Given	—	—	Mathematical Optimization (Unknow method to solve an approximation model)	—	—	—
[23]	Given	—	—	Mathematical Optimization (Unknow method to solve an approximation model)	—	—	—
[24]	V2V	Traffic flow theory/DKM	—	Analytical model	—	—	—
[20]	Unknown	Traffic flow theory	—	Mathematical Optimization (Global search)	—	—	—
[25]	Loop detector	Data-driven/Traffic flow theory	—	Mathematical Optimization (NLP)	—	—	—
[26]	Loop detector	Data-driven/Traffic flow theory	—	Analytical model	√	—	—
[27]	Loop detector	Traffic flow theory/DKM	—	Mathematical Optimization (DP/MPC)	√	—	—
[28]	Loop detector	Traffic flow theory	—	Mathematical Optimization (MPC/PMP)	√	—	—
[29]	Loop detector	Traffic flow theory	√	DRL	√	—	—
Proposed	Navigation-apps	Data-driven	√	DRL	√	√	√

green phase starts and the vehicle at the end of the queue begins to move. Meanwhile, the term 'dissipation time' denotes the time needed for a vehicle to pass through the upcoming signalized intersection. He et al. [22] integrated the position of the end of the queue and the discharge time of the last car in the queue into their eco-driving algorithm, replacing the distance of the ego vehicle to signalized intersections and the start time of the green light. They simplified the speed planning optimization problem into a multi-stage approximation model, enabling real-time computation. Building on this work, Wu et al. [23] conducted Vehicle-in-the-loop experiments, which further confirmed the reliability of the original research.

However, both studies assumed given queue information, an unrealistic scenario in real-world applications. Therefore, queue estimation has become a focal point in subsequent research efforts. Yang et al. [24] utilized vehicle-to-vehicle (V2V) communication to acquire the speed of each vehicle ahead, using a threshold of 5 mph to determine whether the preceding vehicle has stopped, thereby estimating the queue length. They determined the timing for vehicles to pass

through signalized intersections by combining traffic flow theory [30] with a deterministic kinematic model (DKM). Finally, they planned the speed trajectory for the ego vehicle using an analytical model, ensuring real-time capability. Yang et al. [20] estimated the dissipate time of vehicles ahead using traffic flow theory. Based on these estimations, they formulated a new ecodriving optimization problem and ultimately solved it using a global search method. Similarly, Sun et al. [25] employed an LSTM model to predict upcoming traffic flow. Based on traffic flow theory, they used the predicted traffic flow to estimate the time it takes for vehicles to pass through signalized intersections. From this, they derived a modified green light duration. This adjusted green light duration was then incorporated as a new constraint into the nonlinear programming (NLP) for speed planning.

However, relying merely on traffic lights and queue data for velocity optimization is not practical, since the energy efficiency of the ego-vehicle is also determined by the actions of the preceding vehicle [28]. Ye et al. [26] utilized a radial-based neural network (RB-NN) for predicting the speed of the leading vehicle, assessing whether it would halt at the

queue's end. Upon anticipation of a stop, the dissipate time of the leading vehicle is estimated via traffic flow theory, and the ego vehicle's speed trajectory is formulated through an analytical model. In instances where a stop is not projected, a rule-based car-following model is employed to maintain safety on the road. Dong et al. [27] acquired the length of the queue using loop detectors and estimated the dissipate time of the leading vehicle through a method that combines traffic flow theory and the Dynamic Kinematic Model (DKM). For speed planning, a two-layer optimization structure was adopted: initially employing dynamic programming (DP) to plan a reference speed trajectory, followed by the application of model predictive control (MPC) to plan the ego vehicle's acceleration at each time step while ensuring a safe distance from the leading vehicle. Similarly, Dong et al. [28] estimated the dissipate time of the leading vehicle using traffic flow theory. In speed planning, a combined approach of MPC and Pontryagin's Minimum Principle (PMP) was employed to dynamically calculate the ego vehicle's acceleration for each time step in real-time. However, the previously discussed methods assume that queued vehicles are already in place.

In reality, queues usually form slowly as the ego-vehicle approaches a traffic light, making it difficult to predict when a queue will start forming. Additionally, even during the green light phase, slow-moving traffic can affect the ego-vehicle's progress [29]. Li et al. [29] placed loop detectors at each signalized intersection and the midpoint of two signalized intersections. These detectors record the entry time and speed of each vehicle, which are then used to estimate the time each vehicle will arrive at the upcoming signalized intersection based on traffic flow theory. This estimated arrival time is compared with SPaT information to determine whether each vehicle will stop at the signalized intersection. If the last leading vehicle is anticipated to stop, modified traffic light position and discharge time will be further estimated, resulting in a modified speed reference range. If the last leading vehicle is not expected to stop, the estimated arrival time of the last leading vehicle will be used as dissipation time to directly determine the modified speed reference range. This modified speed range is then incorporated as part of the state space for a DRL agent to plan acceleration in real-time. With this design, their work takes into account three different queue scenarios.

However, this line of work has two major limitations: (1) it does not consider lane-merging disturbances, and (2) it cannot achieve on-the-fly queue dissipation estimation. Specifically, discharge time estimation only begins after the last leading vehicle passes the loop detectors located at the midpoint between two signalized intersections. Before this point, the DRL agent relies solely on the unmodified reference speed range as input, which may cause additional acceleration fluctuations and consequently lead to extra energy consumption. Moreover, the algorithm does not account for vehicles that merge into the ego lane after passing the loop detector in multi-lane scenarios, which may compromise the accuracy of the estimated queue length and discharge time. Additionally, the installation of extra loop detectors between two signalized

intersections incurs significant expenses. More importantly, all the aforementioned methods use traffic flow theory or the DKM to predict queues. Even when data-driven approaches are employed, they serve indirectly to provide reference information for these two methods. Both these two approaches treat vehicles in the queue as a single mass moving uniformly, a simplification that overlooks the dynamic nature of vehicles and the randomness of traffic in complex scenarios.

Building on the above analysis, this paper proposes a novel eco-driving algorithm that accounts for lane-merging disturbances while enabling on-the-fly queue dissipation estimation, as illustrated in Fig. 1. Real-time vehicle data, including position and speed, are collected via navigation apps and transmitted to a cloud platform managed by the transportation department. The ego vehicle can at any moment access information of leading vehicles in the same lane from the cloud, including vehicles that have recently merged. Leveraging this data, the dissipation time estimation model continuously updates the estimated dissipation time of the last leading vehicle, which is then used to determine the modified reference speed. The DRL agent subsequently exploits this modified reference speed—together with the states of the ego vehicle and its leading vehicles—to dynamically plan the ego vehicle's acceleration in real time. In this study, the ego-vehicle can obtain SPaT information from signalized intersections via Vehicle-to-Infrastructure (V2I) communication and accurately gather the status of the last leading vehicle through onboard sensors such as camera and Lidar. It is assumed that all vehicles are connected, however, it is not necessary for vehicles to be directly connected to each other through V2V communication systems. Instead, information can be shared through navigation apps on mobile devices, allowing the dissipation time estimation model to precisely predict the dissipation time of the last leading vehicle. This assumption is considered reasonable because navigation apps such as Baidu Maps now provides 'sub-meter level' positioning and 'lane-level' positioning [31]. Although there is no specific data on the prevalence of navigation apps among drivers across different regions, Google Maps was ranked seventh on the global iOS app download charts in 2023 [32], indicating widespread usage. More importantly, compared to the additional installation of loop detectors on roads, sharing traffic data through navigation apps to predict queue states is clearly a more cost-effective choice.

The main contribution of this work unfolds in three aspects: 1) A deep reinforcement learning-based eco-driving framework is developed that incorporates lane-merging disturbances, enabling robust dissipation time estimation in multi-lane scenarios, 2) An on-the-fly data-driven based queue dissipation estimation mechanism is proposed, which allows the ego vehicle to begin accurate estimation immediately after lane entry, reducing latency and avoiding unnecessary acceleration fluctuations and energy consumption, and 3) A practical data acquisition strategy is designed by leveraging real-time position and speed data from widely used navigation apps and transmitting them to a cloud platform, thus enabling accurate

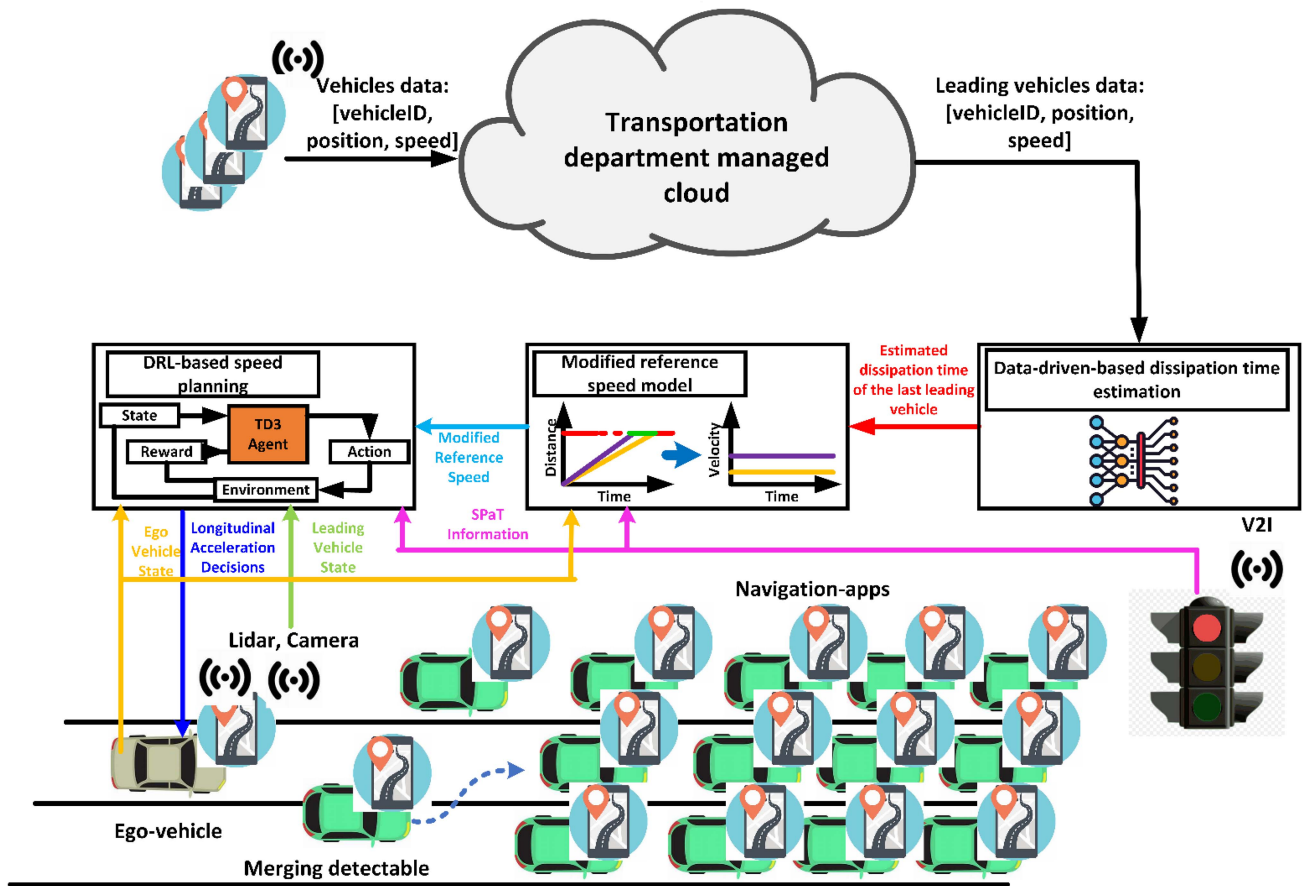


FIGURE 1. The depiction of the proposed eco-driving algorithm in a multi-lane scenario.

estimation without requiring additional roadside infrastructure or direct V2V connectivity.

The remainder of the paper is organized as follows: Section II presents the methodology of the proposed approach. Section III describes the simulation setup. Section IV provides case studies and discussion. Finally, Section V outlines the conclusions and future work.

II. ENERGY CONSUMPTION AND SURROUNDING VEHICLES MODELING

This section introduces the energy consumption model of the ego-vehicle and the dynamic modelling of surrounding vehicles.

A. ENERGY CONSUMPTION MODEL

To ensure precise performance evaluation of the proposed framework, it is essential to adopt a reliable and accurate energy consumption model for electric vehicles. In this study, the Electric Vehicle Emission Model (MMPEVEM), developed by the Mechatronics in Mobile Propulsion group at RWTH Aachen University, is utilized. This model incorporates detailed representations of the entire powertrain, enabling high-fidelity estimation of power demand. Validation against chassis dynamometer experiments on the Worldwide Harmonized Light Vehicles Test Cycle (WLTC) demonstrates

a root mean square error (RMSE) of only 4.99 kW, within a tested battery power range of -40 kW to 50 kW. Readers are referred to [38] for comprehensive descriptions of the model.

B. DYNAMIC MODELLING OF SURROUNDING VEHICLES

We employed the IDM [36] to describe the dynamics of all vehicles on the road except the ego vehicle. In the IDM configuration, the maximum acceleration and comfortable deceleration are both set to 3 m/s^2 , while the driver imperfection factor is specified as 0.5 to introduce stochastic variations in car-following behavior. Each vehicle is assigned a physical length of 5 m , and the maximum speed is limited to 20 m/s . To better reproduce realistic traffic conditions—including those of the leading vehicle—a speed factor is introduced, which follows a truncated normal distribution with mean 1 and standard deviation 0.1 , bounded between 0.2 and 2 . This parameter reflects inter-driver heterogeneity in desired speeds and allows for a more realistic simulation of mixed traffic flow.

II. PROPOSED ECO-DRIVING FRAMEWORK

This section will first introduce the reference speed model, followed by the navigation-apps-based traffic data sharing framework. It will then cover the data-driven-based on-the-fly dissipation time estimation model and the modified reference

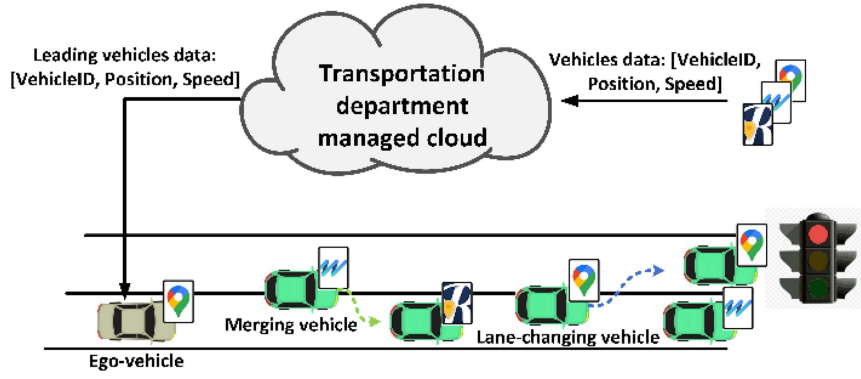


FIGURE 2. Illustration of navigation-apps-based traffic data sharing framework.

speed model, which provides input states for the DRL-based eco-driving algorithm introduced at the end.

A. REFERENCE SPEED MODEL WITHOUT CONSIDERING QUEUE EFFECT

For eco-driving that does not consider the queue effect, the ego-vehicle should adhere to the reference speed to ensure it passes through signalized intersections smoothly while maintaining traffic efficiency. Therefore, when the current phase is green, to promote the controlled ego-vehicle to pass the signalized intersection during the current green phase, $V_{\max}$ and $V_{\min}$ can be defined as follows:

$$\begin{aligned} V_{\max} &= V_{lim} \\ V_{\min} &= \frac{D_{e2s}}{t_{rem}} \end{aligned} \quad (1)$$

V_{lim} is determined as 20 m/s in this paper. When the current phase is red, to promote the controlled ego-vehicle to pass the signalized intersection during the next green phase, $V_{\max}$ and $V_{\min}$ can be defined as follows:

$$\begin{aligned} V_{\max} &= \frac{D_{e2s}}{t_{rem}} \\ V_{\min} &= \frac{D_{e2s}}{t_{rem} + t_g} \end{aligned} \quad (2)$$

B. NAVIGATION-APPS-BASED TRAFFIC DATA SHARING FRAMEWORK

To implement an eco-driving approach that accounts for the queue effect, it is necessary to modify the reference speed. This modification requires the development of an accurate method for estimating dissipation time. Intuitively, enhancing the stability and reducing acceleration fluctuations can further improve the energy efficiency of the eco-driving algorithm. Consequently, the dissipation time estimation needs to be continuously available and capable of integrating merge detection. In other words, continuous collection of traffic data, including the status of merged vehicles, is essential for updating the dissipation time estimation model in on-the-fly. Additionally, the practicality and cost implications of such a traffic data collection method must be assessed. To address

these considerations, this article introduces a navigation-apps-based traffic data sharing framework, illustrated in Fig. 2.

Vehicle data, including $VehID_m$, Pos_m , and V_m of each vehicle, will be transmitted to a cloud managed by the transportation department via various navigation apps. Simultaneously, the ego-vehicle can continuously access the data of leading vehicles traveling in the same lane, at any time, through this cloud via a navigation app. Due to the 'lane level positioning' capability of current advanced navigation apps, the vehicle data of merging and lane-changing vehicles can be dynamically updated—added or removed as necessary.

In this scenario, it is not necessary for vehicles to be directly connected to each other. Importantly, establishing a cloud managed by the transportation department and promoting various navigation apps to share their data is significantly more cost-effective than the procurement and installation of loop detectors at each signalized intersection.

C. DATA-DRIVEN-BASED ON-THE-FLY DISSIPATION TIME ESTIMATION METHOD

In this article, we operate under the assumption that the dissipation time of the last leading vehicle is influenced by its speed and position, as well as the number and speeds of the vehicles ahead and the SPaT information. To predict the dissipation time of the last leading vehicle, e.g., the time the last leading vehicle needs to travel to the upcoming signalized intersection t_{l2i} , we have designed fully connected neural network, illustrated in Fig. 3. This network receives 5 input features: V_{lea} , D_{l2s} , N , V_{ave} , and t_{rem2g} . Where V_{ave} is determined by:

$$V_{ave} = \frac{1}{N} \sum_{n=0}^N V_n \quad (3)$$

It's important to note that if the current phase is green, t_{rem2g} is equal to the remaining time of the current green phase, e.g.,

$$t_{rem2g} = \begin{cases} t_{rem}, & \text{Green} \\ t_{rem} + t_g, & \text{Red} \end{cases} \quad (4)$$

The neural network was designed to optimize training efficiency, prediction accuracy, and real-time performance. It

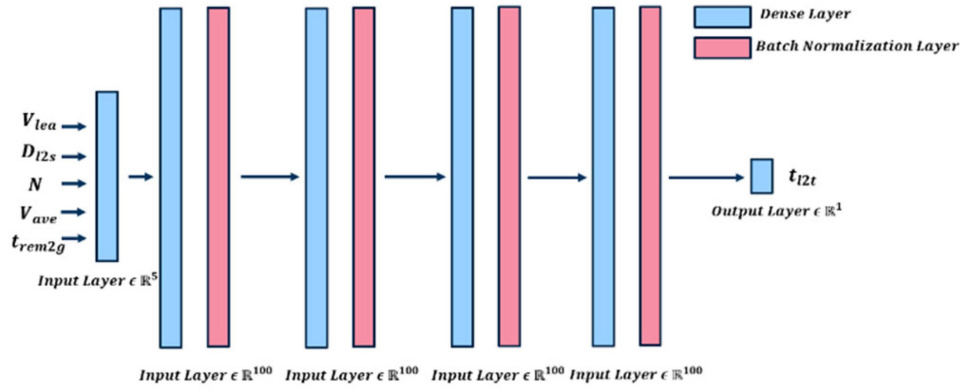


FIGURE 3. Illustration of the neural network for dissipation time estimation.

consists of 1 input layer, 5 hidden layers, and 1 output layer. Each hidden layer includes a Dense layer with 100 neurons, a batch normalization layer to enhance training efficiency and stabilize the training process, and a ReLU [33] activation function.

D. MODIFIED REFERENCE SPEED RANGE MODEL

Previous work on computing modified reference speeds or constructing modified time-space constraints that consider queue effects incorporated both the modified distance to the signalized intersection and the modified green phase time window [24], [25], [26], [27], [28], [29]. However, estimation errors in both the modified distance and the modified time have led to larger inaccuracies when determining the modified reference speed. In this study, the focus is solely on the modified time, which is determined by the estimated dissipation time of the last leading vehicle. This modified time is used to calculate the modified reference speed $[V_{\max_mod}, V_{\min_mod}]$, incorporating the unmodified distance of the ego-vehicle to the upcoming signalized intersection, as shown in (5).

$$V_{\max_mod} = \begin{cases} V_{\max_mod} = \frac{D_{e2s}}{t_{l2t}} \\ V_{\max_mod}, & V_{\max} > V_{\max_mod} \\ V_{\max}, & V_{\max} < V_{\max_mod} \\ V_{\min_mod} = V_{\min} \end{cases} \quad (5)$$

When the last leading vehicle has passed the signalized intersection, the ego-vehicle is permitted to follow immediately, and the modified maximum speed is determined accordingly. However, if the modified maximum speed exceeds the unmodified maximum speed, this indicates that the current phase is green, and the last leading vehicle is nearing passage through the intersection. In this case, the modified maximum speed should be capped to ensure driving safety and adherence to traffic rules.

E. DRL-BASED ECO-DRIVING ALGORITHM

In this paper, Twin Delayed Deep Deterministic Policy Gradient (TD3) model is selected as the DRL agent to plan acceleration for eco-vehicle. TD3 model is an advanced reinforcement learning algorithm introduced by Fujimoto et al

[34] as an improvement over the original Deep Deterministic Policy Gradient (DDPG) method [35]. This algorithm is specifically designed to address issues associated with function approximation errors that lead to overestimation of action values in DDPG and similar algorithms. TD3 incorporates three key innovations to improve the stability and performance of the training process: 1) Twin Critic Networks: TD3 utilizes two separate Critic networks (value functions), and the smaller of the two Q-values produced by these networks is used to compute the target value. This twin-critic design mitigates the overestimation bias inherent in Q-learning based algorithms. The Q-value functions, Q_{θ_1} and Q_{θ_2} , are updated using the following loss functions:

$$L(\theta_i) = \mathbb{E}_{(s,a,r,s') \sim D} \left[(Q_{\theta_i}(s, a) - y)^2 \right], \text{ for } i = 1, 2 \quad (6)$$

where $y = r + \gamma \min_{1,2} Q_{\theta_i}(s', \mu_{\theta'}(s'))$, 2) Target Policy Smoothing: This technique adds noise to the target action, smoothing out the value estimation across similar states and preventing sharp discrepancies in value estimations for similar actions. The modified target policy thus becomes:

$$\mu_{\theta'}(s') + \text{clip}(\epsilon, -c, c), \text{ where } \epsilon \sim \mathcal{N}(0, \sigma) \quad (7)$$

These improvements help TD3 achieve more reliable performance in various control tasks, demonstrating superior results compared to earlier DDPG implementations. TD3's approach to decoupling policy and value updates, along with its novel smoothing and twin critic mechanisms, set a new standard for continuous control tasks in reinforcement learning environments.

The DRL-based eco-driving algorithm aims to allow the ego-vehicle to pass through signalized intersections smoothly and minimize energy consumption while ensuring safety and traffic efficiency by determining acceleration a_t for the ego-vehicle at each time step. Therefore, the action space is represented by $A_t = a_t$, $a_t \in [-3, 3 \text{ m/s}^2]$. The reward function R_t for the DRL agent is defined as follows:

$$R_t = a_1 \cdot R_{ref} + a_2 \cdot R_{ene} + a_3 \cdot R_{saf} + a_4 \cdot R_{eff} \quad (8)$$

To enable the ego-vehicle to pass through signalized intersections without stop-and-go behavior, it is crucial to ensure that the ego-vehicle maintains a speed V_t within the reference range $[V_{\max}, V_{\min}]$ at each time step. Therefore, R_{ref} is defined as piecewise function in (9) shown at the bottom of this page.

R_{ref} will be 0 when V_t is within the modified reference speed range. If V_t exceeds $V_{\max}$, the reward function will increase gradually. Using the natural logarithm function ensures that both the value and gradient of the modified reference speed reward remain stable, even as the speed rapidly increases and surpasses the cut-off point β_1 . This method enhances the training stability of the DRL algorithm bottom of this page.

To further minimize energy consumption, the reward function for energy consumption is defined as follows:

$$R_{ene} = \begin{cases} E_{bat}, & E_{bat} \geq 0 \\ 0, & E_{bat} < 0 \end{cases} \quad (10)$$

R_{ene} increases when E_{bat} is greater than 0, while R_{ene} is set to 0 when E_{bat} is less than 0, such as during the regenerative braking process. This prevents the ego-vehicle from deliberately braking and stopping.

To ensure traffic efficiency, the reward function for traffic efficiency is defined as follows:

$$R_{eff} = \begin{cases} V_t^2, & V_{\min_mod} \leq V_t \leq V_{\max_mod} \\ 0, & \text{others} \end{cases} \quad (11)$$

When V_t is within the reference range, the higher V_t is, the higher R_{eff} becomes.

To ensure driving safety, the acceleration computed by the IDM a_{IDM} is considered, and the safety reward function is defined as follows:

$$R_{saf} = \begin{cases} (a_t - a_{IDM})^2, & a_{IDM} \leq a_t \\ 0, & a_{IDM} \geq a_t \end{cases} \quad (12)$$

R_{saf} will increase when the acceleration generated by the DRL agent exceeds that generated by the IDM. In terms of selecting the state space, it is important that the chosen space is closely related to the reward function. Therefore, the state space of the DRL agent is defined as follows:

$$s_t = [V_{\max_mod}, V_{\min_mod}, V_t, a_t, V_{lea}, a_{lea}, D_{e2s}, D_{e2l}] \quad (13)$$



FIGURE 4. Heatmap of correlation coefficients between variables.

III. SIMULATION SETUP

A. DATASET CONSTRUCTION AND TRAINING OF DISSIPATION TIME ESTIMATION MODEL

To enable the dissipation time estimation model to adapt to varying road lengths and traffic volumes, 12 single-lane roads with different lengths (300 m, 700 m, 1200 m, with traffic lights at the end of each road, to represent traffic area of commercial area, residential area and arterial area respectively) and traffic volumes (600 vehicles/hour, 800 vehicles/hour, 1000 vehicles/hour, 1200 vehicles/hour, to represent low, medium and high traffic volume respectively.) were constructed in the Simulator of Urban Mobility (SUMO) for dataset creation.

Each single-lane road underwent 25 simulations with randomly set initial timings for the signalized intersections and varying departure speeds for vehicles. For data collection efficiency, a simulation time step of one second was chosen. During each simulation time step, V_{lea} , D_{l2s} , N , V_{ave} , t_{rem2g} , and t_{cur} were recorded. When the last leading vehicle passed the signalized intersection, t_{pas} was recorded. t_{pas} was then broadcasted to previous collected data and used to calculate t_{l2t} by subtracting the previously recorded t_{cur} . As a result, totally 40131 data samples consisting of the elements $[V_{lea}, D_{l2s}, N, V_{ave}, t_{rem2g}, t_{l2t}]$ were constructed.

Before training the dissipation time estimation model, a correlation analysis was conducted to evaluate the relationship between the five assumed features and the dissipation time of the last leading vehicle. A correlation matrix depicting

$$R_{ref} = \begin{cases} \beta_1 + \ln((V_t - V_{\max_mod})^2), & V_{\max_mod} < V_t \text{ and } \beta_1 < (V_t - V_{\max_mod})^2 \\ (V_t - V_{\max_mod})^2, & V_{\max_mod} < V_t \text{ and } \beta_1 \geq (V_t - V_{\max_mod})^2 \\ 0, & V_{\min_mod} \leq V_t \leq V_{\max_mod} \\ (V_t - V_{\min_mod})^2, & V_{\min_mod} > V_t \text{ and } \beta_2 \geq (V_t - V_{\min_mod})^2 \\ \beta_2 + \ln((V_t - V_{\min_mod})^2), & V_{\min_mod} > V_t \text{ and } \beta_2 < (V_t - V_{\min_mod})^2 \end{cases} \quad (9)$$

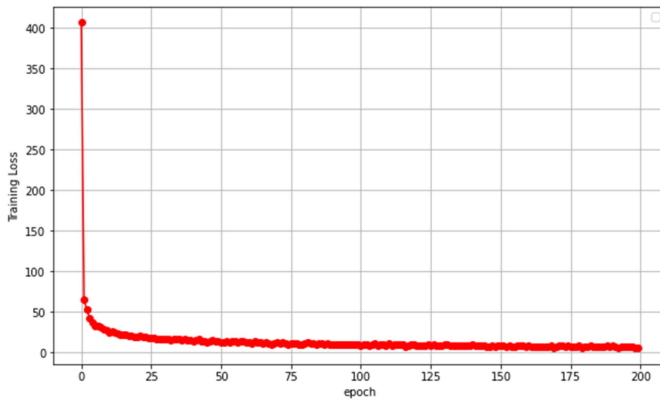


FIGURE 5. Training loss for 200 epochs of dissipation time estimation model.

TABLE 2. Comparison of Prediction Results Between Proposed Model and Li's Model

Model	Mean absolute error (MAE)	Root mean square error (RMSE)
Li's model [29] (Recorded at Time Step A)	3.51	4.45
Proposed model (Recorded at Time Step A)	2.31 (-51.95%)	3.08 (-44.48%)
Proposed model (Recorded at Time Step B)	2.17 (-61.75%)	2.59 (-71.81%)

the correlation among the assumed input features and the prediction target is displayed in Fig. 4. From the colors on the heatmap, green indicates stronger positive correlations. Lighter colors (yellow) suggest weaker relationships. It can be observed that D_{l2s} shows a strong correlation with t_{l2t} , at 0.742, while features V_{lea} and N as well as V_{ave} exhibit a moderate correlation, at 0.505, 0.420 and 0.391 respectively. Intuitively, one might expect t_{rem2g} to have a positive relationship with t_{l2t} . However, when a queue is formed, the last leading vehicle requires additional time to pass the signalized intersection. During this time, feature t_{rem2g} enters the next cycle, resulting in a negative relationship with t_{l2t} , at -0.135 .

The model was trained for 200 epochs using the Adam optimizer [37] with a learning rate of 0.001. The training loss is illustrated in Fig. 5. It can be seen that the training loss decreased significantly in the first 25 epochs, indicating satisfactory learning efficiency. After that, the training gradually converged after 150 epochs.

B. TRAINING OF DRL-BASED ECO-DRIVING ALGORITHM AND BASELINES

The principal parameters of the DRL agent are detailed in Table 2 in the appendix. Both the actor network and the critic network are structured with four hidden layers each. The actor network is constructed with a single deep network,



FIGURE 6. Convergence of the training of the DRL-based eco-driving algorithm.

whereas the critic network is built with two deep networks. A single-lane road 4 km in length was constructed in SUMO for the training process. This road features six signalized intersections positioned randomly, with each intersection having a static signal timing of 60 seconds for green and 60 seconds for red. The training encompassed 40 episodes, during which the initial signal timing at the six intersections and the departure speeds of vehicles were randomized for each episode. It is important to note that the simulation time interval for both training and subsequent case studies was set at 0.1 seconds to achieve higher accuracy. The training reward is displayed in Fig. 6. It can be observed that the model learned very quickly, beginning to converge as early as the 6th episode. There were some fluctuations between the 6th and 40th episodes due to various initial timing settings of the signalized intersections. These settings caused the ego-vehicle to travel at different speeds, which significantly influenced the reward function.

It is important to note that, to critically and fairly evaluate the performance of the proposed method against the following baselines, only one DRL-based eco-driving model was trained using the unmodified reference speed range as input features. This model was tested both individually and in conjunction with various queue estimation methods.

In this study, the baselines include the IDM model, which mimics the driving behavior of human drivers, and the DRL-based eco-driving model, representing standard eco-driving practices without accounting for queue effects. Additionally, the study compares the DRL-based eco-driving model enhanced with Li's queue estimation method [29], which embodies a state-of-the-art eco-driving approach that considers queue effects, and the DRL-based eco-driving model combined with the proposed model.

IV. SIMULATION RESULTS AND DISCUSSION

5 case studies were conducted to critically evaluate various aspects of the proposed method: its prediction accuracy, its

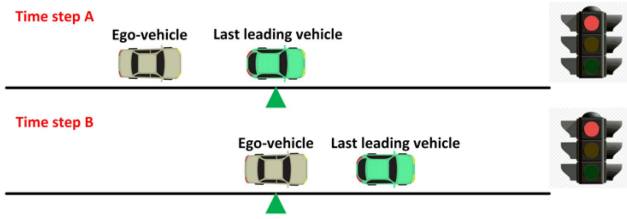


FIGURE 7. Two specific time steps are identified for collecting estimation results: 1) Time step A, which is the time when the last leading vehicle passes the midpoint between two signalized intersections, and 2) Time step B, which is the time when the ego vehicle passes the midpoint between two signalized intersections.

performance under scenario without lane-changing disturbances, its performance under scenario with lane-changing disturbances, its adaptability under different traffic density, and its adaptability under different partial connectivity. Meanwhile, two ablation studies were conducted to evaluate the impact of the reward function weighting scheme of the DRL agent and the network depth of the dissipation time estimation model.

A. EVALUATION OF THE PREDICTION ACCURACY OF THE PROPOSED METHOD

To assess the accuracy of the proposed dissipation time estimation method, this case study compares its estimation accuracy with that of Li's model. Li's model predicts the time when the last leading vehicle reaches the signalized intersections using a traffic-flow based approach. This method provides estimations only when the last leading vehicle passes the loop detectors, which is at the midpoint between two signalized intersections. In this study, this specific time step is designated as time step A, as illustrated in Fig. 7.

A single-lane road, 4 km in length, was constructed in SUMO to conduct this study. The locations of 6 signalized intersections were randomized and differed from those in the training environment. Both the proposed model and Li's model were executed 64 times in this scenario, each with different initial signal timings at the intersections. Although the proposed model is capable of providing estimation results continuously, to ensure fairness, the prediction results of the proposed model were only recorded at Time Step A. Consequently, both the proposed model and Li's model yielded 6 prediction results per run, totaling 384 prediction results for the entire study.

Additionally, to further assess the prediction accuracy of the proposed model, 384 more prediction results were recorded when the ego-vehicle passed the midpoint between two signalized intersections, designated as Time Step B. This additional data collection at Time Step B allows for a more comprehensive evaluation of the proposed model's performance.

The analysis of prediction results of proposed method and Li's method are shown in Table 2. To evaluate the prediction result critically, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used in this study, which are defined in the appendix. The MAE of the prediction results

from the proposed method at Time step A is 2.31, which represents a 51.95% improvement over Li's model, highlighting the superior accuracy of the proposed data-driven model. Additionally, the RMSE at the same time step is 3.08, 44.48% lower than that of Li's model, underscoring the stability of the proposed approach. Moreover, the proposed model can deliver predictions at any given time. Intuitively, as the last leading vehicle approaches the signalized intersection, the prediction accuracy increases. This is evidenced by the data collected at Time step B, where the MAE and RMSE are 2.17 and 2.59, respectively, further validating the effectiveness of the proposed data-driven model.

B. EVALUATION OF THE PROPOSED METHOD UNDER SCENARIOS WITHOUT LANE-MERGING DISTURBANCES

To evaluate the performance of the proposed model under traffic scenarios without lane-merging disturbances, a specific initial signal timing was adopted, and simulations were conducted on the same single lane road as in Case Study A. The distance, acceleration, and speed profiles of four models are illustrated in Fig. 8. Moreover, energy consumption and travel time were employed as performance metrics, with the results summarized in Table 2.

From Fig. 8(a), it is observable that the vehicle controlled by the IDM consistently followed the last leading vehicle and either stopped or braked at each signalized intersection. This behavior is characteristic of the IDM, which aims to adhere to the speed limit (set at 20m/s in this study) as closely as possible and decelerates in response to leading vehicles or a stop line ahead.

The vehicle controlled by the DRL, which used the unmodified reference speed as input states, attempted to pass through signalized intersections immediately after the phase turned green. However, the presence of leading queues ahead obstructed this behavior, resulting in the vehicle either stopping or braking at each intersection. The vehicle controlled by the combined DRL + Li's method initially used the unmodified reference speed as its input state until the last leading vehicle passed the loop detectors, after which it switched to a modified reference speed. This approach enabled the vehicle to pass through signalized intersections more smoothly. However, as indicated by the two pink arrows in Fig. 8(a), the vehicle initially followed the same reference speed as the vehicle controlled solely by DRL after passing the last signalized intersection. It then decelerated to avoid stopping at the end of queues once the last leading vehicle had passed the loop detectors, leading to noticeable acceleration fluctuations, which can be observed from Fig. 8(b). Additionally, as highlighted by the pink circles in Fig. 8(a), the vehicle controlled by DRL + Li's method still exhibited significant braking, particularly when traversing short sections.

The vehicle controlled by DRL + the proposed model demonstrated the most stable and smooth distance, acceleration, and speed trajectories among the four models, as evidenced by Fig. 8(a)–(c). This enhanced performance

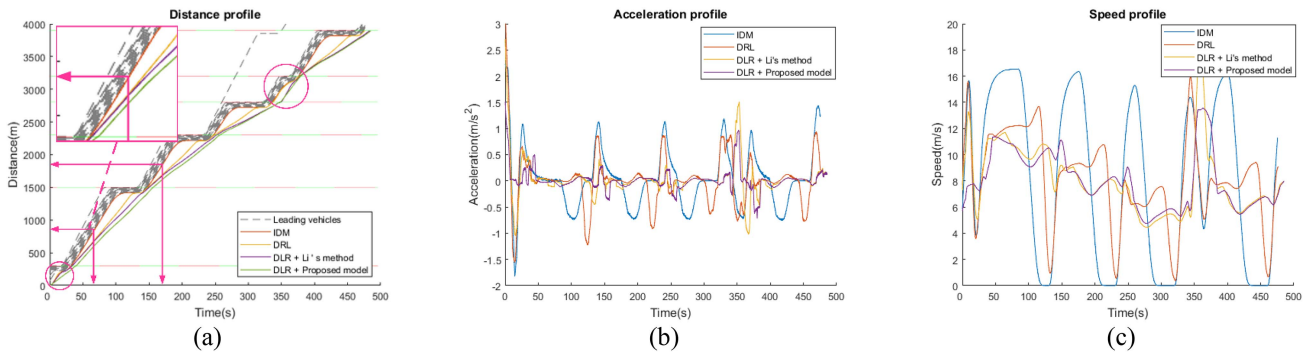


FIGURE 8. Vehicle trajectories in scenarios without lane-merging disturbances under four control models: (a) distance profile, (b) acceleration profile, and (c) speed profile.

TABLE 3. Simulation Results of Vehicles Controlled By Four Models in Scenarios Without Lane-Merging Disturbances

Model	Energy consumption (Wh)	Travel time (s)
IDM	523.7	475.6
DRL	420.6 (-24.51%)	476.7 (+0.2%)
DRL + Li's model [29]	381.4 (-37.31%)	483.7 (+1.7%)
DRL + proposed model	355.6 (-47.27%)	483.7 (+1.7%)

results from the proposed data-driven based on-the-fly dissipation time estimation model, which continuously updates the modified reference speed for the DRL agent, allowing it to effectively determine acceleration early on, taking into account the queue effect immediately after the vehicle passes the last signalized intersection. Furthermore, the smooth trajectories observed in the two pink circles of Fig. 8(a) while the vehicle traveled through short sections underscore the benefits of the accurate dissipation time predictions for the last leading vehicle predicted by the data-driven model.

Eventually, the superiority and effectiveness of the proposed model are clearly reflected in the energy consumption metrics. According to Table 3, the energy consumption of the vehicle controlled by DRL + the proposed model is 355.6 Wh, the lowest among the four models, and 47.27% lower than that of the vehicle controlled by the IDM model. This efficiency was achieved with only a minimal increase in travel time, which was 1.7% longer than that of the vehicle controlled by IDM model.

C. EVALUATION OF THE PROPOSED METHOD UNDER SCENARIOS WITH LANE-MERGING DISTURBANCES

To analyze and evaluate the performance of the proposed model under scenarios with lane-merging disturbances, the traffic scenario from Case Study B was expanded to include three lanes, and simulations were conducted within this setup. In this study, the ego-vehicle was restricted to the right-hand lane, while the surrounding vehicles were permitted to change lanes.

One key distinction between single-lane and multi-lane roads is the potential for surrounding vehicles to merge into the lane where the ego-vehicle is traveling. This dynamic can impact the effectiveness of the queue estimation and dissipation time estimation algorithms. Motivated by this, the primary focus of this case study is to observe and analyze the differences in results between vehicles controlled by the DRL + Li's model and the DRL + proposed model. Accordingly, the distance profiles and energy trajectories for vehicles using these 2 methods were recorded and are displayed in Fig. 9. Additionally, the simulation results for energy consumption and travel time across 4 baseline models are presented in Table 3 for further evaluation and discussion.

The grey dashed lines that appear before the second signalized intersection in both Fig. 9(a) and (b) represent the distance profiles of the merging vehicles. It is crucial to note that these vehicles merged into the lane where the ego-vehicle was traveling after passing the loop detector located between the first and second signalized intersections. In this scenario, the vehicle controlled by DRL + Li's model continued to follow the modified reference speed set when the previous last leading vehicle passed the loop detector. This model did not account for the newly merged vehicles as they were not detected by the loop detector. Consequently, the vehicle controlled by DRL + Li's model eventually stopped at the end of the queue.

Conversely, the vehicle controlled by DRL + the proposed model decelerated after the vehicles merged into its lane, ultimately passing the signalized intersection smoothly. This behavior is attributed to the superiority of the navigation-apps-based traffic data sharing framework and data-driven-based on-the-fly dissipation time estimation model, which allows the ego-vehicle to instantly obtain the status of merging vehicles and immediately update the modified reference speed to consider these vehicles.

To further analyze the impact of these behaviors on energy consumption, the pink dashed line in Fig. 9(c) illustrates the difference in energy consumption between the vehicles controlled by the two strategies as they passed the second signalized intersection. A significant difference of over 50 Wh can be observed, highlighting the energy efficiency benefits of the

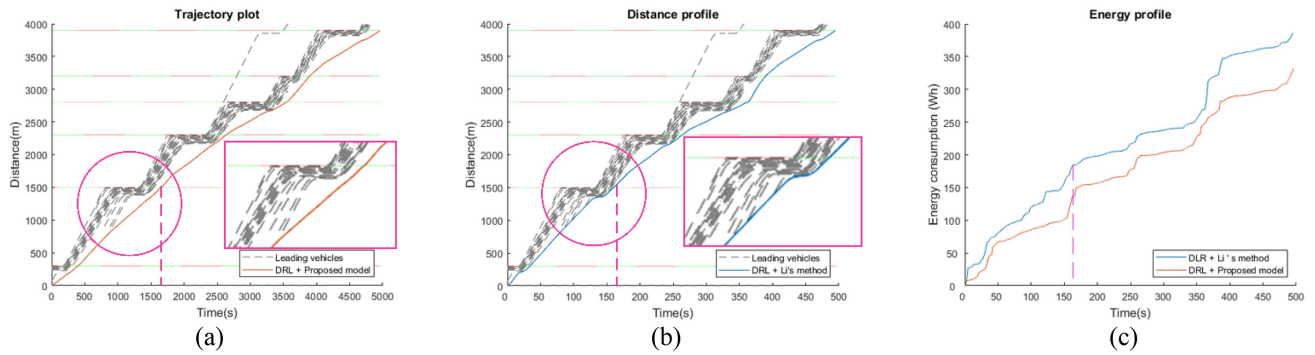


FIGURE 9. Vehicle trajectories under scenarios with lane-merging disturbances: (a) distance profile with DRL + proposed method, (b) distance profile with DRL + Li's method, and (c) comparison of the energy trajectories between DRL + proposed method and DRL + Li's method.

TABLE 4. Simulation Results of Vehicles Controlled By Four Models Under Scenarios with Lane-Merging Disturbances

Model	Energy consumption (Wh)	Travel time (s)
IDM	522.2	477.7
DRL	483.1 (-8.09%)	485.7 (+1.6%)
DRL + Li's model [29]	361.9 (-44.29%)	494.7 (+3.4%)
DRL + proposed model	324.4 (-60.97%)	494.4 (+3.5%)

ability of the proposed model to adapt to dynamically changing traffic conditions, e.g., vehicles merging into the lane.

A more quantified analysis is available from Table 4, where it is evident that the energy consumption of the vehicle controlled by the DRL + proposed model is the lowest, at 324.4 Wh, which is 60.97% lower than that of the vehicle controlled by the IDM model. Meanwhile, the travel time for the vehicle controlled by the proposed model is 3.5% higher than that of the IDM model. In a multi-lane scenario, the vehicle controlled by the DRL + proposed model decelerates when surrounding vehicles merge into its lane, leading to more vehicles attempting to overtake it. This behavior results in a slightly lower energy consumption, but a slightly higher travel time compared to single-lane road scenario.

D. EVALUATION OF THE ADAPTABILITY OF THE PROPOSED METHOD UNDER DIFFERENT TRAFFIC DENSITIES

To critically evaluate the adaptability of the proposed method under different traffic densities, simulations were conducted on the same 3-lane road setup as in Case Study C. For each baseline method, 64 simulation runs were performed under three traffic density levels—300, 700, and 1000 vehicles per lane—while varying the initial signal timings. The average results from these simulations for the 4 methods are illustrated in Table 5.

Overall, as traffic density increases, the energy consumption of all models exhibits a decreasing trend, while travel time correspondingly increases. At each individual traffic density

TABLE 5. Average Simulation Results of Vehicles Controlled By Four Models Under Different Traffic Densities

Model	Energy consumption (Wh)	Travel time (s)
Traffic Density: 300		
IDM	586.7	406.6
DRL	474.2 (-23.72%)	407.0 (+0.10%)
DRL + Li's model [29]	454.9 (-28.97%)	409.1 (+0.61%)
DRL + proposed model	418.2 (-40.29%)	410.7 (+1.00%)
Traffic Density: 700		
IDM	564.1	412.8
DRL	463.5 (-21.7%)	420.4 (+1.8%)
DRL + Li's model [29]	448.1 (-25.9%)	429.3 (+4.0%)
DRL + proposed model	410.4 (-37.5%)	432.0(+4.65%)
Traffic Density: 1000		
IDM	532.4	414.3
DRL	452.2 (-17.7%)	427.4 (+3.16%)
DRL + Li's model [29]	440.0 (-21.0%)	440.2 (+6.25%)
DRL + proposed model	400.7 (-32.9%)	445.2 (+7.46%)

level, the IDM consistently shows the highest energy consumption and the lowest travel time, whereas the DRL + proposed model achieves the lowest energy consumption but with the longest travel time. Across the three traffic density settings (300, 700, and 1000 vehicles per lane), the DRL + proposed model reduces energy consumption relative to IDM by approximately -40.29%, -37.5%, and -32.9%, respectively. However, it also incurs additional travel time of +1.00%, +4.65%, and +7.46%. This occurs because the DRL + proposed model tends to decelerate when being merged into, which in turn encourages more vehicles to attempt merging. The effect becomes more pronounced as traffic density increases.

TABLE 6. Average Simulation Results of Vehicles Controlled By the Proposed Model Under Different Partial Connectivity

Partial Connectivity	Energy consumption (Wh)	Travel time (s)
100%	410.4	432.0
80%	432.9 (+5.5%)	432.1
50%	432.2 (+5.33%)	432.2
30%	450.2 (+9.82%)	432.8

TABLE 7. Average Simulation Results of Vehicles Controlled By DRL Agent Under Different Reward Function Weightings

Reward Function Weightings	Energy consumption (Wh)	Travel time (s)
$a_1 = 8, a_4 = 0.15$	463.5	420.4
$a_1 = 8, a_4 = 0.10$	454.6	425.8
$a_1 = 8, a_4 = 0.20$	472.1	420.2
$a_1 = 4, a_4 = 0.15$	467.9	420.3
$a_1 = 12, a_4 = 0.15$	460.4	421.7

E. EVALUATION OF THE ADAPTABILITY OF THE PROPOSED METHOD UNDER DIFFERENT PARTIAL CONNECTIVITY

To validate the effectiveness of the proposed model under different levels of partial connectivity, 64 simulation runs were conducted in the Case Study C setting with a traffic density of 700 vehicles per lane, where the initial signal timings were varied. The partial connectivity of surrounding vehicles was set to 100%, 80%, 50%, and 30%, with the important assumption that the last leading vehicle was always connected. As shown in Table 6, travel time remains nearly unchanged across the four connectivity levels, while energy consumption increases as partial connectivity decreases. Nevertheless, even under the 30% connectivity condition, the energy consumption is only 9.82% higher than that under full connectivity, which is comparable to the consumption of DRL + Li at the same traffic density in Case Study D. These results demonstrate the robustness of the proposed method under varying degrees of partial connectivity.

F. IMPACT OF REWARD FUNCTION WEIGHTING IN THE DRL AGENT

In this ablation study, we investigate the impact of different reward function weightings on the performance of the DRL agent. The baseline weights are set to $a_1 = 8, a_2 = 0.1, a_3 = 3$, and $a_4 = 0.15$. This configuration was selected to balance energy consumption and traffic efficiency, given that the efficiency-related term is naturally larger in magnitude than the reference term. The ablation primarily focuses on varying a_1 and a_4 , as they directly influence the trade-off between energy efficiency and traffic performance. Fig. 10 illustrates the convergence of DRL agent training under different weight settings, while Table 7 reports the average simulation results obtained from 64 runs in the Case Study C setting with a traffic density of 700 vehicles per lane, where initial signal timings were varied. The results show that increasing a_4 accelerates convergence and leads to higher final rewards,

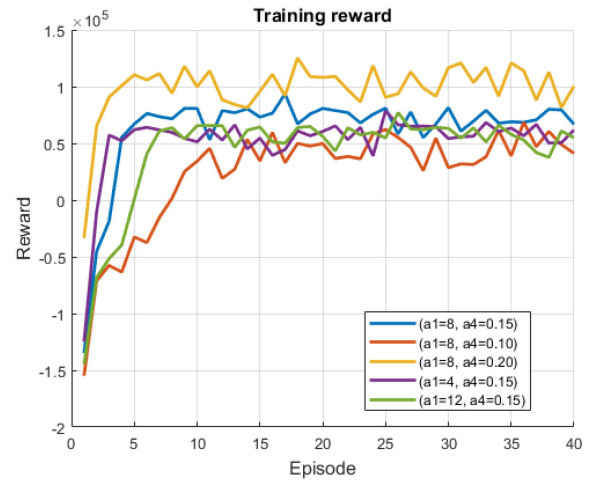


FIGURE 10. Convergence of DRL agent training under different reward function weightings.

TABLE 8. Comparison of Prediction Results of Dissipation Time Estimation Model Under Different Network Depth

Layer	Mean absolute error (MAE)	Root mean square error (RMSE)	Inference time (ms)
3	2.88	3.85	0.10
5	2.31 (-24.68%)	3.08 (-25.00%)	0.16 (+37.5%)
7	2.30 (-25.22%)	3.07 (-25.41%)	0.21 (+52.38%)

with traffic efficiency largely unchanged but at the cost of higher energy consumption. Conversely, decreasing a_4 slows convergence and reduces the final reward, resulting in lower energy consumption but degraded traffic efficiency. Adjusting a_1 also has a noticeable effect: larger values yield slightly slower convergence and lower final rewards, reducing energy consumption but at the expense of efficiency, while smaller values produce the opposite trend with only marginal changes in efficiency. Overall, these findings confirm that the chosen baseline weights achieve a reasonable balance between traffic efficiency and energy saving.

G. IMPACT OF NETWORK DEPTH IN THE DISSIPATION TIME ESTIMATION MODEL

An ablation study was conducted on the network depth of the dissipation time estimation model while keeping the number of neurons per layer unchanged. Models with 3, 5, and 7 layers were trained and then evaluated under the same experimental setup as in Case Study A. At Time Step A, 384 data points were collected for testing, and the results are summarized in Table 2. When the depth increased from 3 to 5 layers, the MAE and RMSE decreased by 24.68% and 25.00%, respectively, while the computation delay increased by 37.5% to 0.16 ms (measured on an 11th Gen Intel(R) Core (TM) i7-11700 @ 2.50 GHz processor). However, further increasing the depth from 5 to 7 layers yielded only marginal improvements in MAE and RMSE, accompanied by a significant rise in computation delay.

TABLE 9. The Principal Parameters of the TD3 Agent

Parameter	Value
Actor network layer	4
Critic network layer	4
Hidden layer width	48
Actor network learning rate	0.0001
Critic network learning rate	0.005
Activation function of hidden layer	ReLU
Activation function of output layer	Tanh
Batch size	512
Discount factor	0.999
Gaussian exploration noise	0.1
Target network update rate	0.005
Range to clip target policy noise	0.5
Critic update noise	0.5
Frequency of delayed policy updates	2

V. CONCLUSION

This research introduced a pioneering deep reinforcement learning-based eco-driving algorithm capable of handling lane-merging disturbances while incorporating on-the-fly queue dissipation time estimation. By integrating a practical and cost-effective navigation-app-based traffic data sharing framework with a data-driven dissipation time estimation model, the proposed approach enables the reinforcement learning agent to continuously receive accurate modified reference speeds that account for both queueing and merging vehicle effects. Through four detailed case studies, involving comparisons with the conventional IDM, standard eco-driving strategies, and state-of-the-art methods that consider queue effects, the proposed framework and estimation method demonstrated high accuracy, operational continuity, and effective merging detection. The results highlighted the superior energy efficiency of the proposed model over all baselines, achieving an average energy reduction of 37.5% compared to IDM under lane-merging scenarios, albeit with a slight increase in travel time. Furthermore, additional case studies under varying traffic densities and different levels of partial connectivity confirmed the robustness of the proposed method.

Future research will address the reliance on SUMO-based simulations by pursuing real-world validations and tackling potential challenges such as GPS inaccuracies in navigation apps and cloud latency. In addition, future work will extend the modeling framework to incorporate more complex surrounding traffic conditions, including heterogeneous vehicle types, to further enhance the generalizability of the proposed approach.

APPENDIX

Table 9 shows principal parameters of the TD3 agent.

The equations for MAE and RMSE, using J as the number of observations, are:

$$\text{MAE} = \frac{1}{J} \sum_{j=1}^J |f_j - \hat{f}_j|$$

$$\text{RMSE} = \sqrt{\frac{1}{J} \sum_{j=1}^J (f_j - \hat{f}_j)^2}$$

where j is the number of observations, y_j is the actual value for j -th observation, $\hat{f}_j$ is the predicted value for the j -th observation, $|f_j - \hat{f}_j|$ represents the absolute error for each prediction, and $(f_j - \hat{f}_j)^2$ is the squared error for each prediction.

ACKNOWLEDGMENT

The authors appreciate for Dr Zekun Guo for his constructing suggestions and comments on the paper.

REFERENCES

- [1] C. Fiori and V. Marzano, "Modelling energy consumption of electric freight vehicles in urban pickup/delivery operations: Analysis and estimation on a real-world dataset," *Transp. Res. Part D: Transport Environ.*, vol. 65, pp. 658–673, 2018.
- [2] X. Yang, W. Lin, R. Gong, M. Zhu, and C. Springer, "Transport decarbonization in big cities: An integrated environmental co-benefit analysis of vehicles purchases quota-limit and new energy vehicles promotion policy in Beijing," *Sustain. Cities Soc.*, vol. 71, 2021, Art. no. 102976.
- [3] S. Tian, C. Li, Q. Lv, and J. Li, "Method for predicting the remaining mileage of electric vehicles based on dimension expansion and model fusion," *IET Intell. Transport Syst.*, vol. 16, no. 8, pp. 1074–1091, 2022.
- [4] C. Sun, J. Guanetti, F. Borrelli, and S. J. Moura, "Optimal eco-driving control of connected and autonomous vehicles through signalized intersections," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 3759–3773, May 2020.
- [5] R. Galvin, "Energy consumption effects of speed and acceleration in electric vehicles: Laboratory case studies and implications for drivers and policymakers," *Transp. Res. Part D: Transport Environ.*, vol. 53, pp. 234–248, 2017.
- [6] M. A. S. Kamal, M. Mukai, J. Murata, and T. Kawabe, "Ecological vehicle control on roads with up-down slopes," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 783–794, Sep. 2011.
- [7] Y. Chen, X. Li, C. Wiet, and J. Wang, "Energy management and driving strategy for in-wheel motor electric ground vehicles with terrain profile preview," *IEEE Trans. Ind. Informat.*, vol. 10, no. 3, pp. 1938–1947, Aug. 2014.
- [8] H. Lim, C. C. Mi, and W. Su, "A distance-based two-stage ecological driving system using an estimation of distribution algorithm and model predictive control," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6663–6675, Aug. 2017.
- [9] H. Lim, W. Su, and C. C. Mi, "Distance-based ecological driving scheme using a two-stage hierarchy for long-term optimization and short-term adaptation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 1940–1949, Mar. 2017.
- [10] X. T. Yang, K. Huang, Z. Zhang, Z. A. Zhang, and F. Lin, "Eco-driving system for connected automated vehicles: Multi-objective trajectory optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 12, pp. 7837–7849, Dec. 2021.
- [11] S. Wang and X. Lin, "Eco-driving control of connected and automated hybrid vehicles in mixed driving scenarios," *Appl. Energy*, vol. 271, 2020, Art. no. 115233.
- [12] Z. Nie and H. Farzaneh, "Real-time dynamic predictive cruise control for enhancing eco-driving of electric vehicles, considering traffic constraints and signal phase and timing (SPaT) information, using artificial-neural-network-based energy consumption model," *Energy*, vol. 241, 2022, Art. no. 122888.
- [13] M. Wegener, L. Koch, M. Eisenbarth, and J. Andert, "Automated eco-driving in urban scenarios using deep reinforcement learning," *Transp. Res. Part C: Emerg. Technol.*, vol. 126, 2021, Art. no. 102967.
- [14] J. Li, X. Wu, M. Xu, and Y. Liu, "Deep reinforcement learning and reward shaping based eco-driving control for automated HEVs among signalized intersections," *Energy*, vol. 251, 2022, Art. no. 123924.
- [15] D. Xie, J. Guo, Y. Jiang, Z. Hou, and J. Deng, "Energy-efficient route and velocity planning for electric vehicles: A hierarchical eco-driving framework integrating traffic and road information," *IEEE Open J. Veh. Technol.*, vol. 6, pp. 1317–1332, 2025.

[16] B. Asadi and A. Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 3, pp. 707–714, May 2011.

[17] W. Li, H. Ding, N. Xu, and J. Zhang, "Toward carbon-neutral transportation electrification: A comprehensive and systematic review of eco-driving for electric vehicles," *IEEE Trans. Transport. Electrific.*, vol. 10, no. 3, pp. 6340–6360, Sep. 2024.

[18] Y. Li, H. He, Y. Chen, and H. Wang, "A cloud-based eco-driving solution for autonomous hybrid electric bus rapid transit in cooperative vehicle-infrastructure systems: A dynamic programming approach," *Green Energy Intell. Transp.*, vol. 2, no. 6, 2023, Art. no. 100122.

[19] H. Tong, L. Chu, Z. Chen, Y. Liu, Y. Zhang, and J. Hu, "Multi-objective autonomous eco-driving strategy: A pathway to future green mobility," *Green Energy Intell. Transp.*, vol. 4, 2025, Art. no. 100279.

[20] H. Yang, H. Rakha, and M. V. Ala, "Eco-cooperative adaptive cruise control at signalized intersections considering queue effects," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1575–1585, Jun. 2017.

[21] Y. Sugiyama et al., "Traffic jams without bottlenecks—Experimental evidence for the physical mechanism of the formation of a jam," *New J. Phys.*, vol. 10, no. 3, Art. no. 033001, 2008.

[22] X. He, H. X. Liu, and X. Liu, "Optimal vehicle speed trajectory on a signalized arterial with consideration of queue," *Transp. Res. Part C: Emerg. Technol.*, vol. 61, pp. 106–120, 2015.

[23] X. Wu, X. He, G. Yu, A. Harmandayan, and Y. Wang, "Energy-optimal speed control for electric vehicles on signalized arterials," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2786–2796, Oct. 2015.

[24] Z. Yang, Y. Feng, X. Gong, D. Zhao, and J. Sun, "Eco-trajectory planning with consideration of queue along congested corridor for hybrid electric vehicles," *Transp. Res. Rec.*, vol. 2673, no. 9, pp. 277–286, 2019.

[25] C. Sun, C. Zhang, H. Yu, W. Liang, Q. Ren, and J. Li, "An eco-driving approach with flow uncertainty tolerance for connected vehicles against waiting queue dynamics on arterial roads," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5286–5296, Aug. 2022.

[26] F. Ye, P. Hao, X. Qi, G. Wu, K. Boriboonsomsin, and M. J. Barth, "Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1378–1389, Apr. 2019.

[27] H. Dong, W. Zhuang, B. Chen, G. Yin, and Y. Wang, "Enhanced eco-approach control of connected electric vehicles at signalized intersection with queue discharge prediction," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 5457–5469, Jun. 2021.

[28] S. Dong, H. Chen, B. Gao, L. Guo, and Q. Liu, "Hierarchical energy-efficient control for CAVs at multiple signalized intersections considering queue effects," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11643–11653, Aug. 2022.

[29] J. Li, A. Fotouhi, W. Pan, Y. Liu, Y. Zhang, and Z. Chen, "Deep reinforcement learning-based eco-driving control for connected electric vehicles at signalized intersections considering traffic uncertainties," *Energy*, vol. 279, 2023, Art. no. 128139.

[30] X. Wu and H. X. Liu, "A shockwave profile model for traffic flow on congested urban arterials," *Transp. Res. Part B: Methodological*, vol. 45, no. 10, pp. 1768–1786, 2011.

[31] "baidu auto solution," 2024. [Online]. Available: <https://mapauto.baidu.com/solution>

[32] D. Curry, "Most popular apps (2023)," 2024. [Online]. Available: <https://www.businessofapps.com/data/most-popular-apps/2024>

[33] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.

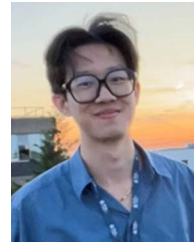
[34] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *Proc. Int. Conf. Mach. Learn.. PMLR*, 2018, pp. 1587–1596.

[35] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," in *Proc. 4th Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Juan, Puerto Rico, May 2–4, 2016.

[36] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E Statist. Phys. Plasmas Fluids Related Interdiscip. Topics*, vol. 62, no. 2 Pt A, pp. 1805–1824, 2000.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 7–9, 2015.

[38] L. Koch et al., "Accurate physics-based modeling of electric vehicle energy consumption in the SUMO traffic microsimulator," in *Proc. IEEE Int. Intell. Transport. Syst. Conf.*, 2021, pp. 1650–1657.



XINXING REN (Graduate Student Member, IEEE) received the B.Eng. degree with first-class honors in electronic and communications engineering in 2022 from Brunel University of London, London, U.K., where he is currently working toward the Ph.D. degree. He is currently a Member of the Brunel Interdisciplinary Power Systems (BIPS) Research Centre, Brunel University of London. His research interests include connected vehicles, reinforcement learning, imitation learning, end-to-end autonomous driving, and the integration of large

language models into transportation systems.



CHUN SING LAI (Senior Member, IEEE) received the B.Eng. (First Class Hons.) in electronic and electrical engineering from Brunel University of London, London, U.K., in 2013, and the D.Phil. degree in engineering science from the University of Oxford, Oxford, U.K., in 2019. He is currently a Senior Lecturer with the Department of Electronic and Electrical Engineering and Course Director of MSc Electric Vehicle Systems with the Brunel University of London. His research interests are in power system optimization and electric vehicle systems. He was a Technical Program Co-Chair for 2022 IEEE International Smart Cities Conference. He is currently the Vice-Chair of the IEEE Smart Cities Publications Committee. He is also an Associate Editor for IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON CONSUMER ELECTRONICS and *IET Energy Conversion and Economics*. He is the Working Group Chair for IEEE P2814 and P3166 Standards, an Associate Vice President, Systems Science and Engineering of the IEEE Systems, Man, and Cybernetics Society (IEEE/SMCS) and Co-Chair of the IEEE SMC Intelligent Power and Energy Systems Technical Committee. He was the recipient of the 2022 Meritorious Service Award from the IEEE SMC Society for "meritorious and significant service to IEEE SMC Society technical activities and standards development". He is a Chartered Engineer and Fellow of the Higher Education Academy.



GARETH TAYLOR (Senior Member, IEEE) is currently a Professor and Director of the Brunel Interdisciplinary Power Systems (BIPS) Research Centre, London, U.K., and was the Head of the Department of Electronic and Electrical Engineering from 2019 to 2023 with the Brunel University of London. He has contributed to more than 300 research publications. His research interests include power system operation, smart grids, renewable energy systems, energy and power system information systems and communications. He was also project coordinator for a 4.2M euro H2020 Energy project from 2017 to 2020/774500 entitled 'Coordination of Transmission and Distribution data eXchanges for renewables integration in the European marketplace through Advanced, Scalable and Secure ICT Systems and Tools' TDX-ASSIST.



YUJIE YUAN (Member, IEEE) received the B.Sc. degree in management from Keuka College, Keuka Park, NY, USA, in 2015, the second B.Sc. degree in project management from the Yunnan University of Finance and Economics, Yunnan, China, in 2015, the M.Sc. degree in science in management from the Keuka College, in 2016, and the Ph.D. degree in control system and engineering from the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China in 2023. She is currently a Lecturer with Civil Aviation University of China, Tianjin, China. Her research interests include system engineering, transport engineering, high-speed rail and civil aviation competition, and system control.