

International Journal of Simulation and Process Modelling

ISSN online: 1740-2131 - ISSN print: 1740-2123

<https://www.inderscience.com/ijspm>

Real-time logistics route planning and energy consumption simulation modelling from a green computing perspective

Shuo Wang

DOI: [10.1504/IJSPM.2026.10079772](https://doi.org/10.1504/IJSPM.2026.10079772)

Article History:

Received:	06 February 2026
Last revised:	19 April 2026
Accepted:	04 May 2026
Published online:	01 October 2026

Real-time logistics route planning and energy consumption simulation modelling from a green computing perspective

Shuo Wang

Business School,
Luoyang Normal University,
Luoyang, 471934, China
Email: cjx_200812@sina.com

Abstract: In the context of increasingly complex global supply chains and rapid e-commerce growth, the logistics industry faces significant energy and environmental challenges. A key issue is the lack of dynamic coupling between real-time traffic and vehicle energy consumption, hindering the balance of efficiency and sustainability. This paper proposes a path optimization method from a green computing perspective. We integrate real-time traffic prediction via graph neural networks with a refined energy consumption model based on vehicle dynamics. A multi-objective function considering time, energy, and cost is constructed, and an adaptive path search algorithm based on multi-objective deep reinforcement learning is designed for real-time decision-making. Experiments show our method significantly outperforms traditional approaches, achieving 89.7% recommendation accuracy and an 18.6% reduction in energy consumption, providing an effective low-carbon solution for smart logistics.

Keywords: green computing; real-time path planning; energy consumption simulation modelling; deep reinforcement learning; logistics optimisation.

Reference to this paper should be made as follows: Wang, S. (2026) 'Real-time logistics route planning and energy consumption simulation modelling from a green computing perspective', *Int. J. Simulation and Process Modelling*, Vol. 23, No. 3, pp.184–198.

Biographical notes: Shuo Wang is a Lecturer in the Business School at Luoyang Normal University, China. He received his Bachelor's in Management from Sias International College of Zhengzhou University (2012) and a Master's in Engineering from Henan University of Technology (2014) in China. He published seven papers. His research interests include green logistics, deep learning, logistics path planning and cross-border electronic commerce.

1 Introduction

In the context of an increasingly complex global supply chain network and the continuous rapid growth of e-commerce, the logistics and transportation industry has become a vital pillar of modern economic development. However, along with its scale expansion comes significant energy consumption and environmental burden. According to statistics, carbon emissions from the transportation sector account for a considerable proportion of global greenhouse gas emissions (Kohl, 2019), and urban freight logistics is one of the fastest-growing segments. Facing the increasingly severe challenges of climate change and the global consensus on the 'dual carbon' strategic goals, promoting the green and low-carbon transformation of the logistics industry is not only an inevitable requirement for environmental protection but also a core issue for the sustainable development of the industry. Against this backdrop, 'green computing', as a technical concept aimed at optimising computing resources and system energy efficiency while minimising its negative environmental impact (Sarkar and Gul, 2020), is beginning to permeate from traditional fields such as data centres to a broader

range of smart Internet of Things and intelligent transportation systems. Introducing the idea of green computing into logistics route planning means that while pursuing transportation efficiency and economic benefits, energy consumption and carbon emissions must be treated as equally important optimisation goals for systematic modelling and decision-making, thereby achieving a win-win situation of economic and environmental benefits in logistics operations (Sun et al., 2022).

Path planning is a core technical aspect in logistics transportation management. Its objective is to find the optimal or approximately optimal driving route for vehicles or fleets from the starting point to the destination, while meeting various constraints (such as time windows, load restrictions, etc.). Traditional path planning algorithms, such as Dijkstra algorithm, A* algorithm and their various variants, mainly rely on the static topological structure of the road network and fixed travel times for searching the shortest path (Bast et al., 2016). Their optimisation goals usually focus solely on minimising the driving distance or time. However, the traffic conditions in the real world are highly dynamic and uncertain. Road conditions fluctuate

significantly due to events such as traffic accidents, weather changes, and peak-hour congestion (Lv et al., 2014), which can significantly affect the actual travel time and vehicle energy consumption. Static path planning schemes often perform poorly in dynamic environments and may even lead to a substantial increase in total travel time and fuel consumption by choosing a path that is 'shortest' on a static map but is severely congested in real-time. Therefore, developing dynamic path planning technologies that can sense and adapt to real-time traffic conditions is a prerequisite for achieving efficient and reliable logistics transportation.

In recent years, with the maturity of internet of things, big data, and artificial intelligence technologies, real-time path planning research has made significant progress. On one hand, traffic state perception and short-term prediction technologies based on real-time traffic data provide a data foundation for dynamic path planning. Researchers utilise time series analysis, machine learning and even deep learning models to predict the average speed or travel time of road sections in future periods, thereby integrating dynamic information into the path decision-making process. On the other hand, in terms of optimisation methods, in addition to classic dynamic programming methods, meta-heuristic algorithms are widely used due to their flexibility in solving combinatorial optimisation problems and are applied to handle complex scenarios such as vehicle routing problems (VRPs) with time window constraints (Braekers et al., 2016). However, most current studies, when introducing dynamic traffic information, still focus on minimising time or distance as the 'optimisation' core, failing to deeply integrate the real-time energy consumption changes of vehicles in the dynamic traffic flow as an intrinsic and quantifiable decision variable.

Although current research has achieved many results in improving the real-time and adaptability of path planning, there are still key challenges and gaps in achieving truly 'green' path planning, which constitutes the core motivation of this study. Firstly, at the modelling level, existing research often simplifies the description of the energy consumption of logistics vehicles, and current research that combines real-time traffic prediction with energy consumption modelling is insufficient. Some studies do attempt to integrate fuel consumption models, but usually rely on historical average traffic data or assume constant traffic conditions, failing to utilise high-precision short-term traffic predictions to prospectively assess the potential impact of different path choices on future energy consumption. Secondly, in terms of optimisation objectives, most dynamic path planning methods are single-objective or transformed into single-objective optimisation through weighted summation, making it difficult to effectively balance the possible competitive relationships between 'timeliness', 'economy', and 'greenness'. Additionally, with the development of computing technology, path planning methods based on deep reinforcement learning (DRL) has emerged. However, the reward function design of existing research based on DRL mostly still focuses on delivery

efficiency, and either lacks encoding of energy consumption targets or designs them crudely, failing to effectively embed the refined real-time energy consumption model into the learning mechanism of the agent. In response to the aforementioned research gaps, this paper aims to re-examine and construct the real-time logistics path planning problem from the perspective of green computing, proposing a collaborative optimisation framework that integrates real-time traffic prediction and high-fidelity vehicle energy consumption modelling.

To address the aforementioned research deficiencies, this paper aims to re-examine and reconstruct the real-time logistics path planning problem from the perspective of green computing. The core research idea is to establish a closed-loop collaborative optimisation logic of 'real-time traffic status → precise energy consumption impact → path decision feedback', thereby effectively balancing logistics efficiency and environmental sustainability. For this purpose, we propose a novel integrated framework named green-aware DRL for Path Planning (GREEN-DRL). Real-time traffic prediction, when integrated into urban governance frameworks, enables proactive traffic management, reduces unnecessary energy waste, and supports the development of low-emission zones and dynamic congestion pricing – key elements of sustainable urban mobility. The main innovations and contributions of this study are as follows:

- A tightly-coupled modelling framework integrating real-time prediction and physical energy consumption: We propose a seamless integration scheme of a real-time traffic predictor based on graph neural networks (GNNs) and a high-fidelity vehicle energy consumption model based on vehicle dynamics principles. This framework can prospectively and precisely quantify the energy consumption impact of different route choices under dynamic traffic conditions, surpassing traditional modelling methods that rely on static or simplified energy consumption assumptions.
- An adaptive preference multi-objective DRL optimisation algorithm: We have designed a multi-objective DRL agent with a vectorised reward function, which can simultaneously consider time, energy consumption and cost. This algorithm incorporates an adaptive weight adjustment mechanism, enabling the agent to autonomously learn complex nonlinear trade-off strategies and generate a series of Pareto-optimal paths that align with the preferences of different decision-makers in real-time.

The remaining part of this article is arranged as follows. Section 2 conducts a systematic literature review, covering traditional path planning algorithms, traffic prediction methods, the application of green computing in logistics, and vehicle energy consumption modelling, thereby clarifying the research background and gaps. Section 3 elaborates on the methodology of the proposed GREEN-DRL framework, including problem modelling,

high-fidelity energy consumption model, real-time traffic prediction based on GNNs, and multi-objective path optimisation algorithm based on DRL. Section details the experimental design, including datasets, simulation environment, baseline methods, and evaluation metrics. Section 5 comprehensively reports and analyses the experimental results, including overall performance comparison, algorithm convergence, multi-objective trade-offs, and ablation experiments. Section 6 discusses the findings, and Section 7 concludes the paper.

2 Literature review

2.1 *Traditional path planning algorithms and their dynamic extension*

Building upon the background presented in Subsection 1.1, we now systematically review the related work in the following areas: The classic shortest path algorithms, such as Dijkstra's algorithm and A* algorithm, form the theoretical foundation for path search in static road networks. Dijkstra's algorithm ensures the finding of the global optimal solution by systematically exploring the shortest path tree from the starting point to all reachable nodes, but its computational complexity becomes a bottleneck when dealing with large-scale networks (Hart et al., 1968). A* algorithm improves efficiency by introducing heuristic functions to guide the search direction. Its performance highly depends on the quality of the heuristic function design, and it performs well in Euclidean space and other scenarios. However, the core assumption of these algorithms is that the road segment weights (usually distance or fixed travel time) are static and known, which is far from the uncertain real traffic environment. To cope with dynamic changes, researchers have proposed the concept of dynamic path planning. Dynamic A* (D) and its variants (such as D Lite) can efficiently reuse previous search information to quickly correct paths after the cost of some road segments changes, and are suitable for scenarios such as robot navigation. Nevertheless, these algorithms usually require frequent reception of precise traffic condition update events, and their optimisation goals are still limited to a single time or distance dimension, failing to incorporate complex and dynamic cost factors such as energy consumption into their core graph search framework, and thus are difficult to directly apply to green logistics planning that requires comprehensive consideration of multiple objectives.

2.2 *Traffic condition prediction and route decision based on machine learning*

With the development of big data and machine learning technologies, using data-driven methods to perceive and predict traffic conditions and then assist in path decision-making has become a research hotspot. Early studies mostly employed time series models (such as autoregressive integrated moving average) or traditional

machine learning models (such as support vector regression, random forest) to conduct short-term predictions of traffic flow and speed for specific road sections or areas (Vlahogianni et al., 2014). In recent years, deep learning models have gained a dominant position in this field due to their powerful feature extraction capabilities. Convolutional neural networks (CNN) are used to handle the spatial correlation of traffic data, while recurrent neural networks (RNN) and their variants [such as long short-term memory (LSTM) and gated recurrent unit (GRU)] are good at capturing time dependencies. To simultaneously model the spatial topology structure and temporal dynamics of the traffic network, GNNs have received widespread attention. By treating the road network as a graph and the road sections as nodes, GNN can effectively aggregate the information from neighbouring nodes, thereby achieving more accurate speed or travel time predictions at the road section level. These prediction results provide crucial input for dynamic path planning. Some studies attempt to integrate the time-varying road section weights predicted by the model into the classic graph algorithm framework. Another technical route avoids the explicit 'prediction-optimisation' two-stage paradigm and attempts to let the intelligent agent directly learn the navigation strategy in a dynamic environment through end-to-end reinforcement learning (RL). Deep Q networks (DQN) and their improved algorithms have been applied to problems such as taxi route planning (Mnih et al., 2015), where the agent learns through interaction with the simulation environment to maximise the cumulative reward (such as negative travel time) as the goal. However, the reward function design in most existing RL-based path planning research mostly focuses on efficiency indicators, and the modelling and integration of the key green goal of energy consumption are still at an early stage.

2.3 *The application of green computing concepts in logistics optimisation*

Green computing originated from the practice of reducing data centre energy consumption (Rong et al., 2016), and its core idea is to optimise through algorithms, architectures, and resource management, minimising the energy consumption of the system while meeting performance targets. This concept is gradually penetrating into broader computing fields, including the internet of things and edge computing. In the logistics and transportation domain, a closely related concept is 'green logistics' or 'sustainable transportation'. Early academic research focused on introducing environmental costs in the VRP. For example, the pollution-routing problem (PRP) is a comprehensive emission cost function based on vehicle load, travel distance, and speed, which makes it a more practical optimisation problem on the basis of the traditional VRP model. Such studies usually calculate energy consumption or emissions as a convex function of speed and consider it as part of the optimisation objective. However, many PRP models still assume that speed is a constant or piecewise

constant decision variable, failing to couple with real-time and dynamic traffic conditions in a closed-loop manner. In recent years, some research has begun to explore combining green computing infrastructure such as edge computing and cloud computing with logistics management to indirectly support green logistics by optimising task offloading and data processing efficiency. But how to deeply integrate the energy efficiency optimisation of the computing system and the physical energy consumption optimisation of the transportation tool at the algorithm level to form a unified ‘computing-transportation’ collaborative energy-saving framework remains an urgent frontier direction to be explored.

2.4 Progress in vehicle energy consumption modelling methods

An accurate energy consumption model is the foundation for achieving green path planning. The existing vehicle energy consumption/emission models can be roughly classified into three categories: macro-average models, micro-physical models, and data-driven models. Macro-average models, such as the computer programme for emissions from road transport (COPERT) model, conduct regression analysis based on a large amount of statistical data to determine the average emission rate of vehicles at different average speeds (Ntziachristos et al., 2009), and are often used for emission inventory calculations at the regional or national level. However, their accuracy is insufficient to support individual vehicle path-level decision-making. Micro-physical models are based on vehicle dynamics principles and calculate instantaneous energy consumption or emissions by analysing the power consumed by vehicles in overcoming rolling resistance, air resistance, slope resistance, and acceleration resistance during their travel (Rakha et al., 2004). These models (such as the vehicle transient – comprehensive physical fuel mechanism model) have clear physical meanings and high theoretical accuracy, and their inputs require detailed instantaneous speed, acceleration, road slope, and vehicle parameters. However, obtaining high-precision real-time driving cycles (speed-time curves) and applying them to forward-looking path planning is a major challenge. Data-driven models utilise a large amount of actual driving data collected by on-board diagnostic (OBD) systems or sensors (Basso et al., 2019), and establish complex mapping relationships from vehicle states, driving conditions to energy consumption through machine learning methods (such as neural networks, random forests). These models can capture factors that are difficult to be precisely described by physical formulas (such as driver behaviour, air conditioning usage), but their performance heavily relies on the quality and coverage of training data, and their interpretability is relatively weak. The current research trend is to integrate physical mechanisms with data-driven methods to construct ‘black-box’ models to balance accuracy and robustness. However, in the context of dynamic path planning, how to seamlessly connect the

various energy consumption models with the future driving conditions output by real-time traffic prediction to form a lightweight, high-fidelity energy consumption estimation module that can be used for online path evaluation and optimisation, relevant research is not systematic. We choose a physics-based energy model over purely data-driven alternatives because it offers better generalisability across unseen road geometries and traffic conditions, requires no extensive local training data, and provides interpretable outputs (e.g., the contribution of slope or acceleration), which are critical for real-time planning decisions where reliability and explainability are paramount.

2.5 Summary of research status and analysis of limitations

In summary, the current research has achieved significant results in various subfields such as dynamic path planning, traffic prediction, green logistics, and energy consumption modelling. However, when it comes to the comprehensive goal of ‘real-time logistics path planning from a green computing perspective’, there are still obvious gaps and deficiencies. Firstly, there is a fragmentation in research perspectives. The traffic prediction community aims to improve prediction accuracy, the path optimisation community focuses on designing efficient algorithms, and the energy consumption modelling community delves into the physical or data authenticity of the models. There is a lack of deep collaboration among these three. Many so-called ‘green’ path planning studies merely attach a static or simplified energy consumption coefficient to traditional optimisation models, failing to construct a closed-loop optimisation logic of ‘real-time traffic state → fine energy impact → path decision feedback’. Secondly, the multi-objective collaborative optimisation mechanism is not perfect. When dealing with the multi-objective conflicts of time, cost, and energy consumption, existing methods mostly use linear weighted summation to convert them into single-objective problems, and the setting of weights often relies on prior experience, lacking flexibility, and it is difficult to obtain the Pareto solution set that represents the trade-off relationship. Although multi-objective evolutionary algorithms can be used to solve such problems, their computational cost is high and they are difficult to meet the requirements of large-scale real-time decision-making. Finally, the potential of cutting-edge artificial intelligence methods has not been fully exploited. DRL demonstrates strong capabilities in complex decision-making problems, but its application in logistics path planning mostly focuses on efficiency improvement, and the reward function’s portrayal of green goals is overly simplistic (such as being proportional to distance), failing to embed advanced, speed-profile-related energy consumption models into the learning process of the agent, thereby limiting the in-depth exploration and emergence of energy-saving driving strategies. These deficiencies provide clear entry points and innovation spaces for this research.

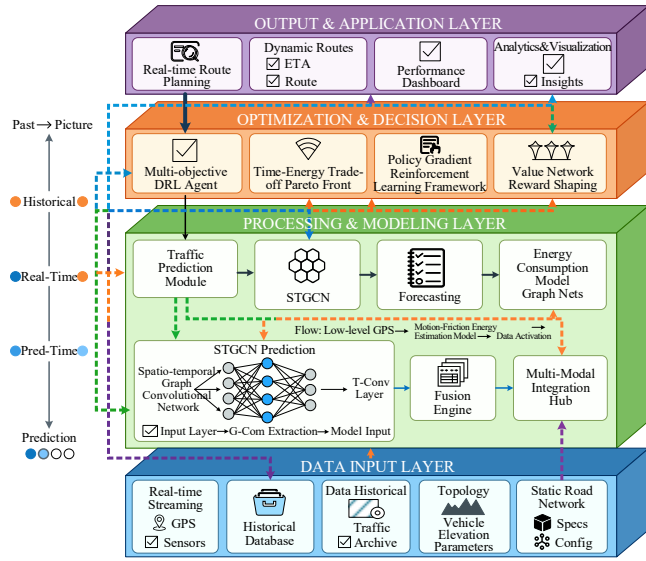
3 Methodology

3.1 System overview and problem modelling

Beyond algorithm-level energy efficiency, green computing principles are reflected in our system design:

- 1 data processing uses efficient batch aggregation to reduce I/O overhead
- 2 model training employs early stopping and gradient checkpointing to minimise graphics processing unit (GPU) energy consumption
- 3 deployment can be containerised on edge nodes close to data sources, reducing communication energy.

Figure 1 A framework for real-time logistics route planning and energy consumption modelling from a green computing perspective (see online version for colours)



The proposed integrated framework in this paper aims to achieve real-time logistics path planning from the perspective of green computing. Its core lies in achieving a balance between logistics efficiency and environmental benefits by jointly optimising travel time, energy consumption, and economic costs. As shown in the system framework in Figure 1, it mainly consists of three closely coupled modules: the real-time traffic prediction module, the high-fidelity vehicle energy consumption modelling module, and the multi-objective path optimisation module. The system workflow begins with the fusion and feature extraction of multi-source heterogeneous data (including real-time global positioning system (GPS) trajectories, road network sensor data, historical traffic databases, and static road attributes). Subsequently, the spatiotemporal graph convolution network learns from the processed data and outputs fine-grained speed predictions for each section of the road network in the future planning period. This prediction result, along with vehicle dynamics parameters and road slope information, is input into the energy consumption model based on physical laws, enabling forward-looking estimation of the energy consumption of the vehicle on any candidate path. Finally, an intelligent

agent based on multi-objective DRL, which integrates real-time predicted traffic conditions, dynamically calculated energy costs, and fixed economic costs, makes online sequential decisions in the complex urban road network and outputs a recommended path that achieves a good balance among multiple objectives.

To formalise the representation, we abstract the urban road network as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = v_1, v_2, \dots, v_N$ represents N road intersections or key nodes, and $\mathcal{E} = e_1, e_2, \dots, e_M$ represents M directed road segments. Each segment $e_k = (v_i, v_j)$ is associated with a set of time-varying attributes, including its base length l_k , gradient θ_k , and travel speed $v_k(t)$ at time t . For a logistics transportation task from origin $o \in \mathcal{V}$ to destination $d \in \mathcal{V}$, a path P is represented as a sequence of consecutive segments: $P = (e_{k_1}, e_{k_2}, \dots, e_{k_H})$, where H denotes the number of segments in the path. Our objective is to find a path that minimises the following three-dimensional objective vector:

$$\min_{P \in \Omega} \mathbf{F}(P) = [T(P), E(P), C(P)]^T \quad (1)$$

where Ω denotes the set of all feasible paths from o to d . $T(P)$, $E(P)$, and $C(P)$ represent the total travel time, total energy consumption, and total economic cost of path P , respectively. Specifically, the total travel time is the sum of the travel times for each segment based on predicted speeds:

$$T(P) = \sum_{e_k \in P} t_k = \sum_{e_k \in P} \frac{l_k}{\bar{v}_k} \quad (2)$$

where $\bar{v}_k$ denotes the predicted average speed of the segment e_k when the vehicle is scheduled to pass through it. The total economic cost $C(P)$ may include tolls, congestion charges, etc. and is modelled as the sum of costs across segments: $C(P) = \sum_{e_k \in P} c_k$. Since these three objectives have

different dimensions and may conflict, we perform minimum-maximum normalisation to facilitate subsequent optimisation algorithms:

$$\tilde{F}_i(P) = \frac{F_i(P) - F_i^{\min}}{F_i^{\max} - F_i^{\min}}, \quad i \in \{T, E, C\} \quad (3)$$

where $F_i^{\min}$ and $F_i^{\max}$ represent the lower and upper bounds of objective i within the feasible region, estimated via prior knowledge or historical data, respectively. Ultimately, the multi-objective optimisation problem can be transformed into a weighted single-objective problem as the foundational form for algorithmic solution:

$$\min_{P \in \Omega} J(P) = \mathbf{w}^T \tilde{\mathbf{F}}(P) = w_T \cdot \tilde{T}(P) + w_E \cdot \tilde{E}(P) + w_C \cdot \tilde{C}(P) \quad (4)$$

where the weight vector $\mathbf{w} = [w_T, w_E, w_C]^T$ satisfies $w_T + w_E + w_C = 1$ and $w_i \geq 0$, it reflects the decision-makers' relative preferences for time, energy consumption and cost. Our

framework allows for the dynamic adjustment of w to generate different Pareto optimal solutions.

3.2 High-fidelity vehicle energy consumption modelling

Accurate energy consumption estimation is the key to green path planning. We employ a microscopic model that integrates vehicle dynamics principles and real-time driving conditions. This model can reflect the subtle influences of speed, acceleration, road slope, and vehicle parameters. The total force acting on the vehicle during motion, $F_{total}(t)$, primarily consists of the following four components:

- 1 rolling resistance: $F_{roll}(t) = mgC_r \cos(\theta(t))$
- 2 air resistance: $F_{air}(t) = \frac{1}{2} \rho_a C_d A_f v(t)^2$
- 3 grade resistance: $F_{grade}(t) = mg \sin(\theta(t))$
- 4 acceleration resistance: $F_{acc}(t) = ma(t)$

where m is the total vehicle mass (kg), g is the gravitational acceleration (9.81 m/s^2), C_r is the rolling resistance coefficient, ρ_a is the air density (kg/m^3), C_d is the air resistance coefficient, A_f is the frontal area of the vehicle (m^2), $v(t)$ is the instantaneous velocity (m/s), $a(t)$ is the instantaneous acceleration (m/s^2), and $\theta(t)$ is the road gradient angle (rad). Therefore, the total resistance is:

$$F_{total}(t) = F_{roll}(t) + F_{air}(t) + F_{grade}(t) + F_{acc}(t) \quad (5)$$

The instantaneous power $P_{req}(t)$ (in watts) required by the vehicle to overcome this resistance is:

$$P_{req}(t) = F_{total}(t) \cdot v(t) \quad (6)$$

Substituting equation (5) into equation (6), we obtain the complete expression for instantaneous power:

$$P_{req}(t) = \left(mgC_r \cos \theta(t) + \frac{1}{2} \rho_a C_d A_f v(t)^2 + mg \sin \theta(t) + ma(t) \right) v(t) \quad (7)$$

The actual energy consumption must also account for the efficiency of the vehicle's powertrain system. Let the overall efficiency of the engine/motor be $\eta_{eff} \in (0, 1]$. Then, the instantaneous power $P_{source}(t)$ drawn by the vehicle from the energy source (fuel or battery) is:

$$P_{source}(t) = \frac{\max(P_{req}(t), 0)}{\eta_{eff}} + \eta_{brake} \cdot \min(P_{req}(t), 0) \quad (8)$$

Equation (8) distinguishes between traction and braking conditions. When $P_{req}(t) \geq 0$ (traction), energy flows from the source to the wheels with efficiency η_{eff} ; when $P_{req}(t) < 0$ (braking), part of the kinetic energy can be recovered, where $\eta_{brake} \in [0, 1)$ represents the braking energy recovery efficiency. If no recoverable system exists, $\eta_{brake} = 0$. For a path P , its total energy consumption is obtained by integrating $P_{source}(t)$ over the travel time:

$$E(P) = \int_0^{T(P)} P_{source}(t) dt \quad (9)$$

In practical path planning algorithms, we employ discrete approximation calculations (Franceschetti et al., 2013). Each road segment e_k is divided into N_k sub-segments of equal or unequal length. Assuming that within each sub-segment j , the speed $v_{k,j}$, acceleration $a_{k,j}$, and gradient $\theta_{k,j}$ remain constant, the energy consumption along segment e_k is approximated as:

$$E_k \approx \sum_{j=1}^{N_k} \left[\frac{1}{\eta_{eff}} \left(mg(C_r \cos \theta_{k,j} + \sin \theta_{k,j}) + \frac{1}{2} \rho_a C_d A_f v_{k,j}^2 \right) v_{k,j} \Delta t_{k,j} + \frac{1}{2\eta_{eff}} m (v_{k,j}^2 - v_{k,j-1}^2) \right]^+ \quad (10)$$

where $[x]^+ = \max(x, 0)$, $\Delta t_{k,j}$ denotes the time elapsed through the j^{th} segment, and $v_{k,0}$ represents the initial velocity upon entering segment e_k . The final term in the formula accounts for energy dissipation corresponding to kinetic energy change, its sign implicitly contained within the velocity squared difference. The total path energy consumption is thus $E(P) = \sum_{e_k \in P} E_k$. The core inputs to the

model – the velocity profile $v_{k,j}$ and acceleration $a_{k,j}$ – are derived from the real-time traffic prediction module described in the next section.

3.3 Real-time traffic prediction based on GNNs

Compared with grid-based CNNs or sequence models (e.g., LSTM), GNNs naturally preserve the topological structure of road networks, avoid information loss caused by flattening or gridding, and have lower computational overhead for large-scale graphs. To obtain refined road segment speeds for future time periods, we designed a multi-task spatio-temporal graph convolutional network (MT-STGCN) (Yu et al., 2017). The model takes as input the road network state from the past T_h time steps and outputs speed and traffic flow predictions for each road segment over the next T_p time steps.

- Graph construction: Each road segment e_k in the road network $\mathcal{G}$ is modelled as a graph node. The node feature matrix $X^t \in \mathbb{R}^{N \times F}$ contains F features at time t , such as average speed, traffic volume, timestamp encoding (hour, day of the week), etc. The adjacency matrix $A \in \mathbb{R}^{N \times N}$ is constructed based on spatial relationships between segments, such as defining weights using a Gaussian kernel function based on distance or topological connectivity:

$$A_{ij} = \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) \text{ if } d_{ij} \leq \kappa, \text{ otherwise } 0 \quad (11)$$

where d_{ij} denotes the Euclidean distance or network distance between the centres of road segments i and j , and σ and κ are parameters controlling the weight distribution and neighbourhood range.

- Spatio-temporal convolution layer: A single spatio-temporal convolution operation combines spatial and temporal convolutions. The spatial convolution employs a first-order simplified form approximated by Chebyshev polynomials (Defferrard et al., 2017):

$$H_{\text{spatial}} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X W_s \quad (12)$$

where $\tilde{A} = A + I$ is the adjacency matrix with self-connections added, $\tilde{D}$ is its degree matrix, and W_s is the learnable spatial convolution weights. Let A be the adjacency matrix of the road graph, where A_{ij} indicates the spatial proximity between road segments i and j . Then $\tilde{A} = A + I_N$ adds self-loops, and $\tilde{D}$ is the diagonal degree matrix of $\tilde{A}$. Subsequently, temporal convolution employs a one-dimensional convolution kernel that slides along the temporal dimension to capture dynamic patterns (Zheng et al., 2020):

$$H_{\text{temporal}} = \text{ReLU}(\Phi * H_{\text{spatial}} + b_t) \quad (13)$$

where Φ denotes the temporal convolution kernel, $*$ represents the convolution operation, b_t is the bias term, and rectified linear unit (ReLU) is the activation function.

- Multi-task output and loss function: The model extracts deep features by stacking multiple spatio-temporal convolutional blocks, ultimately using two independent output layers to predict velocity $\hat{V}$ and flow $\hat{F}$ respectively. The training objective is to minimise the joint loss between velocity prediction and flow prediction (Zhang et al., 2018):

$$\mathcal{L}_{\text{total}} = \lambda_v \cdot \mathcal{L}_{\text{MSE}}(\hat{V}, V) + \lambda_f \cdot \mathcal{L}_{\text{MAE}}(\hat{F}, F) \quad (14)$$

where $\mathcal{L}_{\text{MSE}}$ denotes the mean squared error loss, and $\mathcal{L}_{\text{MAE}}$ denotes the mean absolute error loss, $\mathcal{L}_{\text{total}}$ is the joint loss function combining velocity prediction and flow prediction, and λ_v and λ_f are hyperparameters balancing the two tasks. Upon training completion, the model can generate rolling predictions of future traffic states $\hat{V}^{t+1}, \dots, \hat{V}^{t+T_p}$ based on current and historical traffic data, providing critical inputs for energy consumption modelling and path planning.

3.4 Path optimisation based on multi-objective DRL

We model real-time path planning as a sequential decision problem and solve it using a multi-objective DRL framework. At its core lies the definition of a Markov decision process (MDP) quintuple (S, A, P, R, γ) , where S is the state space, A is the action space, P is the state transition

probability, R is the vector reward function, and γ is the discount factor.

- State and action: At decision time τ , the state $s_t \in S$ comprises the vehicle's current node v_τ , destination d , current load $load_\tau$, cumulative costs from origin $[T_{acc}, E_{acc}, C_{acc}]$ incurred from the origin to the present, and the traffic feature map $\mathcal{F}_{\text{traffic}}$ for the relevant future time period of the local road network centred at v_τ provided by the prediction module. The action $\alpha_t \in A$ involves selecting one of the adjacent road segments from the current node v_τ as the next driving segment.
- Vector reward function: After executing action a_t (corresponding to selecting path segment e_k), the environment transitions to new state s_{t+1} and generates a three-dimensional instantaneous reward vector:

$$r_\tau = [r_\tau^T, r_\tau^E, r_\tau^C]^T \quad (15)$$

The components are defined as follows:

- Time reward: $r_\tau^T = -\Delta T_k$, where ΔT_k is the predicted travel time through segment e_k [calculated in equation (2)].
- Energy reward: $r_\tau^E = -\Delta E_k$, where ΔE_k is the predicted energy consumption through segment e_k [calculated in equation (10)].
- Cost reward: $r_\tau^C = -\Delta C_k$, where ΔC_k is the travel cost of segment e_k .

Additionally, a large positive sparse reward R_{success} is granted when the agent reaches the destination d ; a penalty R_{penalty} is imposed for violating constraints (e.g., exceeding time limits).

Fixed-weight scalarisation tends to converge to a single solution on the convex hull of the Pareto front and requires tedious re-tuning for different preferences. In contrast, our vector reward approach, coupled with adaptive weight adjustment [equation (19)], allows the agent to explore multiple trade-off regions in one training session and dynamically shift focus to under-optimised objectives, leading to a richer and more balanced Pareto set.

- Multi-objective policy optimisation: We employ a policy gradient-based approach. The agent's policy is represented by the neural network $\tau_\theta(a|s)$, with parameters θ , which outputs the probability distribution of selecting each action under state s . The objective is to maximise the expected discounted cumulative vector reward: $J(\theta) = \mathbb{E} \tau \sim \pi \theta \left[\sum_{\tau} \gamma^\tau r_\tau \right]$. Given the

multi-objective nature, we seek a Pareto-optimal policy. Define the scalarised expected reward as $J_w(\theta) = w^T J(\theta)$, where w is the preference weight vector. By the policy gradient theorem, its gradient is (Sutton et al., 2003):

$$\begin{aligned} & \nabla_{\theta} J_w(\theta) \\ &= \mathbb{E}_{\tau \sim \pi_{\theta}} \left[\sum_{\tau} \nabla_{\theta} \log \pi_{\theta}(a_{\tau} | s_{\tau}) \cdot w^T A(s_{\tau}, a_{\tau}) \right] \end{aligned} \quad (16)$$

Among these, $A(s, a)$ denotes the vector value function computed by the generalised advantage estimate. To stabilise training, we maintain a separate value function estimator $V_{\phi}(s)$ for each target i , parameterised by ϕ , updated by minimising the temporal difference error:

$$\mathcal{L}_{\text{value}}(\phi) = \mathbb{E}_{(s, r, s')} \left[(r^i + \gamma V_{\phi}(s') - V_{\phi}(s))^2 \right] \quad (17)$$

Algorithm 1 Multi-objective DRL path planning training

```

1  Initialise the policy network parameters  $\theta$  and the vector
   critic network parameters  $\phi$ 
2  Pre-trained and frozen traffic prediction module (ST-
   GCN)
3  for each training session do
4      Reset the environment. The vehicle is at the
       starting point  $o$ . Obtain the initial state  $s_0$ 
5      for steps  $t = 0$  to  $H_{\max}$  do
6          Select action  $a_t$  (choose the next segment)
           according to policy  $\pi_{\theta}(\cdot | s_t)$ 
7          Execute action  $a_t$  in the simulated
           environment, combining the prediction
           module and the energy consumption model
           [equations (8) and (9)] to compute  $r_t^T, r_t^E$ ,
           and  $r_t^C$ . This yields the reward vector
            $r_t = [r_t^T, r_t^E, r_t^C]^T$  and the next state  $s_{t+1}$ 
8          Store the transfer sample  $(s_t, a_t, r_t, s_{t+1})$  into
           the empirical replay buffer
9          if  $s_{t+1}$  is the terminal state  $d$  or an invalid state
           then
10             Interrupt the loop
11         end if
12     end for
13     Sample a batch of data from the buffer
14     Update the vector critic network  $\phi$  to minimise the
       temporal difference error
15     Calculate the multi-objective policy gradient using
       equations (16) and (17), and update the policy
       network parameters  $\theta$ 
16 end for

```

- Adaptive weight adjustment: To explore different regions of the Pareto frontier in a single training session, we allow the weights w to be dynamically adjusted within a certain range. For instance, adjustments can be made periodically based on the performance of the current strategy on each objective:

$$w_i^{(\text{epoch})} = \frac{\exp(-\beta \cdot J_i(\theta) / \bar{J}_i)}{\sum_{j \in \{T, E, C\}} \exp(-\beta \cdot J_j(\theta) / \bar{J}_j)} \quad (18)$$

where $J_i(\theta)$ denotes the estimated return of the current policy on objective i , $\bar{J}_i$ is the normalisation factor, and β is the temperature parameter. This mechanism encourages the agent to focus more on objectives with relatively poor current performance, thereby driving the policy toward the Pareto frontier.

- Algorithm flow: Algorithm 1 summarises the main training process. The agent continuously experiments and makes mistakes in the simulated environment, using the collected experience to update the policy network and value network, eventually learning an intelligent strategy that can make green and efficient path planning based on real-time traffic information and multiple target preferences. Let $H_{\max}$ denote the maximum number of decision steps allowed per episode (e.g., the maximum path length in terms of road segments).

4 Experimental design

This chapter aims to provide a detailed experimental plan for verifying the performance of the GREEN-DRL model. To ensure the reproducibility of the research and the reliability of the conclusions, we will systematically introduce the construction of the simulation environment and data set, the selection and parameter configuration of comparison algorithms, the multi-dimensional evaluation index system, and the specific experimental implementation process. Through rigorous design, the intention is to fairly and comprehensively evaluate the overall effectiveness of the proposed framework in conducting green path planning in dynamic traffic environments.

4.1 Experimental environment and datasets

Dataset and experimental environment: to verify the effectiveness of the proposed method, we conducted simulation experiments in a real urban road network scenario. The experiments utilised two public datasets for integration:

- 1 New York taxi trip data, which provides a large number of GPS trajectory points of taxis and can be used to extract historical and real-time average speeds of road sections.
- 2 New York City OpenStreetMap road network data (Haklay and Weber, 2008), which is used to construct the required topological network and extract information such as road length, number of lanes, road grade, and slope (obtained through digital elevation model interpolation). We selected a road network subgraph in the central area of Manhattan, containing approximately 500 nodes and 1,200 directed road sections, as the experimental site, and selected taxi data within this area for one month for model training and testing.

The experiment was conducted on a workstation equipped with an Intel Core i9-12900K processor, a NVIDIA GeForce RTX 3090 graphics card and 64 GB of memory. The algorithm was implemented in Python 3.9, and the deep learning part used the PyTorch 1.12 framework. The simulation environment used Simulation of Urban Mobility (SUMO) for microscopic traffic flow simulation to generate realistic dynamic traffic conditions and vehicle trajectories (Krajzewicz et al., 2012). The SUMO simulation was configured with time-varying traffic demands extracted from the New York taxi dataset: morning peak (7:00–9:00), evening peak (17:00–19:00), and off-peak periods. Random incidents (e.g., lane closures, accidents) were injected with a 5% probability per hour to emulate unexpected congestion. The traffic light logic was set to actuated control to mimic real urban intersections.

4.2 Baseline methods and parameter settings

To comprehensively evaluate performance, we compare the proposed method (denoted as GREEN-DRL) with the following four categories of classical and state-of-the-art path planning algorithms:

- 1 A*: A classical heuristic search algorithm. We employ Euclidean distance as the heuristic function, with cost weighting based on fixed travel time (derived from historical average speed).
- 2 Genetic algorithm (GA) (Reed et al., 2000): A meta-heuristic algorithm frequently used for VRPs. Population size set to 100, iterated for 200 generations, employing sequential crossover and permutational mutation.
- 3 Q-learning (Clifton and Laber, 2020): A classic table-based RL algorithm. States discretised as road segment IDs, action space defined as exit segments, learning rate set to 0.1, discount factor 0.9.
- 4 DQN: A deep learning-based value iteration method. Its network architecture resembles the value network in our approach, but the reward function is solely the negative travel time, i.e., single-objective (minimum time) optimisation.

All parameters of the comparison algorithms were optimised via grid search to achieve optimal performance. To ensure that each baseline algorithm can be adapted to the specific issues of this study, we carried out targeted design and fine-tuning for them. For the genetic algorithm, we adopted the non-dominated sorting genetic algorithm II (NSGA-II) framework suitable for multi-objective optimisation, with its fitness function being the weighted sum of time, energy consumption and cost, and the weight settings were aligned with the default values of GREEN-DRL. For Q-learning, considering the size of the state space, we used linear function approximation for generalisation. For DQN, we constructed a two-layer fully connected network as the Q network and set up the experience replay and target network mechanism to stabilise

the training. Its reward function was only related to the negative travel time, thus focusing on a single efficiency objective. We did not include meta-heuristics (e.g., NSGA-II for green VRP) or transformer-based planners because the former are designed for offline static or quasi-static problems with high computational cost (not suitable for real-time decision making), while the latter typically require massive training data and are less explored in dynamic graph path planning. Our focus is on real-time, learnable policies under changing traffic conditions.

4.3 Evaluation metrics

We employ a multi-dimensional and multi-level evaluation system to comprehensively assess the performance of each algorithm, including classification accuracy, sorting quality, and key performance indicators (Vujović, 2021):

- 1 Classification/matching accuracy metrics: Treating ‘whether it is an optimal path (on the Pareto frontier)’ as a binary classification problem, we calculate accuracy (ACC), precision, recall, and F1-score.
- 2 Ranking quality metrics: For the top-K recommended path list generated for each origin-destination (OD) pair, evaluate consistency with the true Pareto frontier ranking using normalised discounted cumulative gain (NDCG@K).
- 3 Key performance metrics: Average travel time post-simulation (units: minutes), average energy consumption (units: kWh), and energy reduction rate (relative to the baseline A* algorithm) (Yao et al., 2013).
- 4 To determine whether GREEN-DRL’s performance improvement over other algorithms is statistically significant, we perform paired t-tests on results for each OD pair across metrics, setting the significance level at $\alpha = 0.05$ and reporting p-values (Guillera-Arroita et al., 2017).

Specifically, the classification accuracy metric is used to evaluate the algorithm’s ability to identify the global optimal path (i.e., the Pareto optimal solution). We calculate an approximate real Pareto optimal solution set for each tested OD pair through exhaustive search combined with detailed simulation. If the recommended path by the algorithm falls within this set, it is considered a correct classification. The NDCG@K metric is more closely aligned with practical applications. It assesses the quality of the top-K path list recommended by the algorithm, with a higher value indicating that the recommended list contains more high-quality paths and is sorted more reasonably. The key performance indicators directly reflect the final effect of path planning. Among them, the energy reduction rate is the core metric for measuring ‘green’ performance. All indicators are calculated based on the simulation results of 200 randomly generated OD pairs on the test set and their average values are reported.

4.4 Experimental procedure

The entire experiment followed the standard machine learning workflow, which was divided into three stages: training, validation, and testing.

- Data preparation and division: Divide the data of one month in chronological order into training set, validation set and test set in a ratio of 7:2:1, ensuring the time generalisation ability for evaluation.
- Model training and tuning: GREEN-DRL and other learning-based baseline methods (such as DQN) are trained on the training set. The hyperparameters of all models (such as learning rate, number of network layers, discount factor, etc.) are determined through grid search on the performance on the validation set.
- Performance testing and comparison: On the test set, the same 200 OD pair path planning tasks were run for each algorithm. Each planning was based on the dynamic traffic scenarios generated by SUMO, which were consistent with the historical patterns but had not been seen by the training process. The detailed performance data of each path were recorded to calculate all the evaluation indicators as described in Section 4.3.
- Statistical analysis and ablation experiments: Paired sample t-tests were conducted on key indicators to confirm the statistical significance of performance differences. Additionally, ablation experiments were carried out by designing different variants of the GREEN-DRL model to analyse the contributions of each core component (such as the traffic prediction module, detailed energy consumption model, multi-objective rewards).

5 Results

This section systematically presents and analyses the detailed results of the aforementioned experiments. Firstly, it reports the overall performance comparison between GREEN-DRL and various baseline algorithms; secondly, it analyses the training convergence process and stability of the RL algorithms; then, it demonstrates the multi-objective trade-off ability of the algorithm in different traffic scenarios; finally, through ablation experiments and parameter analysis, it deeply examines the contributions and influences of each core component in the proposed framework.

5.1 Overall performance comparison

This section presents the average performance of all algorithms on a test set of 200 randomly generated origin-destination pairs. Table 1 summarises the average values of the key evaluation metrics. We randomly generated 200 different OD pairs for testing. Table 1 shows the average performance of each algorithm on key

indicators. GREEN-DRL achieved the best results in three metrics: ACC, F1-score, and NDCG@5, reaching 89.7%, 0.872, and 0.921 respectively, significantly outperforming other comparison algorithms. This proves its comprehensive advantages in identifying global efficient paths and generating high-quality recommendation lists. Particularly, in terms of energy consumption reduction rate, GREEN-DRL achieved 18.6%, far exceeding DQN's 12.4%. This directly validates the key role of integrating refined energy consumption models and multi-objective optimisation in achieving the goal of green computing. Although GA was slightly better than GREEN-DRL in terms of travel time (23.1 minutes vs. 24.3 minutes), its energy consumption was approximately 9.8% higher, demonstrating the inherent trade-off between single-time-goal optimisation and multi-objective collaborative optimisation.

Table 1 Overall performance comparison of different path planning algorithms

Algorithm	ACC (%)	Precision	Recall	F1-score
A*	78.2	0.752	0.721	0.735
GA	81.5	0.791	0.773	0.782
Q-learning	83.2	0.805	0.798	0.801
DQN	85.4	0.832	0.821	0.826
GREEN-DRL	89.7	0.881	0.864	0.872

Algorithm	NDCG@5	Avg. time (min)	Avg. energy (kWh)	Energy reduction (%)
A*	0.812	26.5	15.4	-
GA	0.843	23.1	14	9.1
Q-learning	0.865	24.8	13.5	12.4
DQN	0.887	24.5	13.5	12.4
GREEN-DRL	0.921	24.3	12.5	18.6

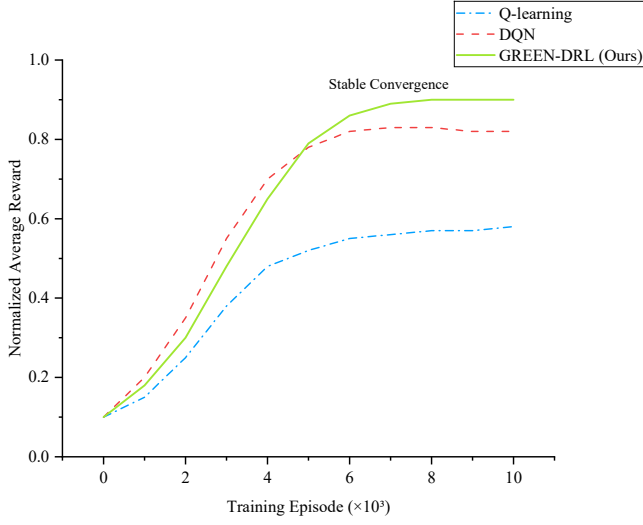
The results of the statistical significance test show that the p-values of GREEN-DRL compared to DQN in terms of ACC, F1-score, and NDCG@5 are 0.003, 0.008, and 0.015 respectively, all of which are less than 0.05, indicating that the performance improvement is statistically significant.

5.2 Analysis of algorithm convergence and stability

Analysis of algorithm convergence and stability: Figure 2 shows the average reward (normalised) curves of each RL algorithm (Q-learning, DQN, and GREEN-DRL) during the training process. All curves have been smoothed using moving averages. It can be seen that GREEN-DRL converges to a stable and higher reward plateau after approximately 5,000 rounds, and the fluctuation amplitude of its curve is significantly smaller than that of Q-learning and DQN, demonstrating its stronger learning stability and policy robustness. Although DQN converges faster, its final performance ceiling is lower than that of GREEN-DRL, which is related to the limitations of its single-objective reward function. Due to the discretisation of the state space

and table limitations of Q-learning, its performance after convergence is the lowest.

Figure 2 Average reward convergence curves during training of different RL algorithms (see online version for colours)



The differences in training curves reveal the learning characteristics of different algorithm architectures. The higher reward plateau achieved by GREEN-DRL after convergence is attributed to its multi-objective vector rewards, which provide richer and smoother gradient signals, facilitating the exploration of better balance solutions by the policy network. Its lower volatility indicates that integrating the detailed energy consumption model with real-time prediction makes the environmental model more stable, reducing the strategy oscillations caused by drastic fluctuations in traffic state estimation. In contrast, the reward signal of single-objective DQN is too simplistic, which easily causes the strategy to prematurely fall into local optima (such as purely chasing the shortest time), limiting its ability to discover more comprehensive optimal paths.

5.3 Multi-objective trade-off across diverse scenarios

Multi-objective trade-off across diverse scenarios: To quantitatively evaluate the multi-objective trade-off capability of GREEN-DRL across diverse urban scenarios, we compare its performance with the baseline algorithms on six representative OD pairs with varying traffic conditions and distances. The comprehensive results are summarised in Table 2, which lists the detailed travel time and energy consumption for each algorithm on each OD pair.

As shown in Table 2, the A* algorithm, relying on static historical speeds, consistently yields the longest travel times and highest energy consumption across all OD pairs, demonstrating its inadequacy in dynamic environments. The GA achieves the shortest travel times by aggressively optimising for this single objective, but it fails to concurrently minimise energy consumption. Both Q-learning and DQN show improvements over A* by incorporating real-time information, yet their

single-objective or simplified reward structures limit their ability to find the optimal balance.

In contrast, our GREEN-DRL method consistently achieves the lowest energy consumption in every single OD pair (in Table 2), while maintaining travel times comparable to or only slightly higher than those of DQN and GA. For instance, on OD 3, which represents a long-distance trip under congested conditions, GREEN-DRL reduces energy consumption by approximately 17.8% compared to A* (13.4 kWh vs. 16.3 kWh) and by 6.9% compared to DQN (13.4 kWh vs. 14.4 kWh), with a marginal time increase of 1.5 minutes compared to GA. This consistent pattern across diverse OD pairs strongly validates that GREEN-DRL successfully internalises the complex trade-off between efficiency and sustainability, making it a robust and reliable solution for green logistics planning under varying real-world conditions.

The last row of Table 2, showing the mean values, provides an overall performance summary. GREEN-DRL achieves the best mean energy consumption (12.8 kWh) and a competitive mean travel time (24.5 min), solidifying its overall superiority in the multi-objective optimisation task.

Table 2 Comprehensive performance comparison across different OD pairs

OD pair	A*	GA	Q-learning	DQN	GREEN-DRL (ours)
	Time (min)/energy (kWh)	Time (min)/energy (kWh)	Time (min)/energy (kWh)	Time (min)/energy (kWh)	Time (min)/energy (kWh)
OD 1	26.8/15.6	23.5/14.2	25.1/13.8	24.8/13.7	24.5/12.7
OD 2	25.2/14.9	22.8/13.5	24.2/13.2	23.9/13.1	23.6/12.1
OD 3	28.1/16.3	24.3/14.8	26.0/14.5	25.7/14.4	25.2/13.4
OD 4	27.5/16.0	23.9/14.5	25.5/14.2	25.2/14.1	24.9/13.1
OD 5	26.0/15.2	23.2/13.8	24.6/13.5	24.3/13.4	24.0/12.4
OD 6	27.0/15.8	24.1/14.6	25.7/14.3	25.4/14.2	25.0/13.2
Mean	26.8/15.6	23.6/14.2	25.2/13.9	24.9/13.8	24.5/12.8

5.4 Ablation study and parameter analysis

- Module ablation study: To understand the roles of each component in GREEN-DRL, we conducted ablation experiments, and the results are shown in Table 3.
- GREEN-DRL (without PE): The real-time traffic prediction module has been removed, and only the historical average speed is used. Various indicators, especially the energy consumption reduction rate, have significantly decreased, highlighting the importance of dynamic information for forward-looking energy-saving planning.
- GREEN-DRL (without EM): Removed the refined energy consumption model and only used fixed energy consumption coefficients. The energy consumption reduction rate sharply decreased to a level comparable

to that of DQN, confirming that the high-fidelity energy consumption model is the key to achieving significant energy-saving effects.

- GREEN-DRL (single-obj.): The reward function was changed to only optimise time (similar to DQN), while retaining the prediction and energy consumption modules for evaluation. Although the energy-saving effect improved, ACC and NDCG decreased, indicating that the multi-objective collaborative optimisation mechanism is indispensable for balancing efficiency and green goals.
- Multi-objective weight sensitivity analysis: In equation (4), we defined the weight vector $w = [w_T, w_E, w_C]$. With $w_C = 0.1$ fixed, we adjusted the ratio of w_T and w_E to observe its impact on the Pareto front. Figure 3 shows the distribution of the resulting paths in the two-dimensional space of average time and average energy consumption when $w_T:w_E$ changes from 9:1 to 1:9. It can be seen that as the weight of w_E increases, the energy consumption of the planned path gradually decreases, but the travel time increases accordingly, forming a clear Pareto front. GREEN-DRL can smoothly cover various strategies ranging from ‘time priority’ to ‘energy consumption priority’ by adjusting the weights, providing decision-makers with a flexible trade-off space. Our default weight setting is $w_T = 0.5$, $w_E = 0.4$, $w_C = 0.1$, which is located in the middle of the front and achieves a good balance.

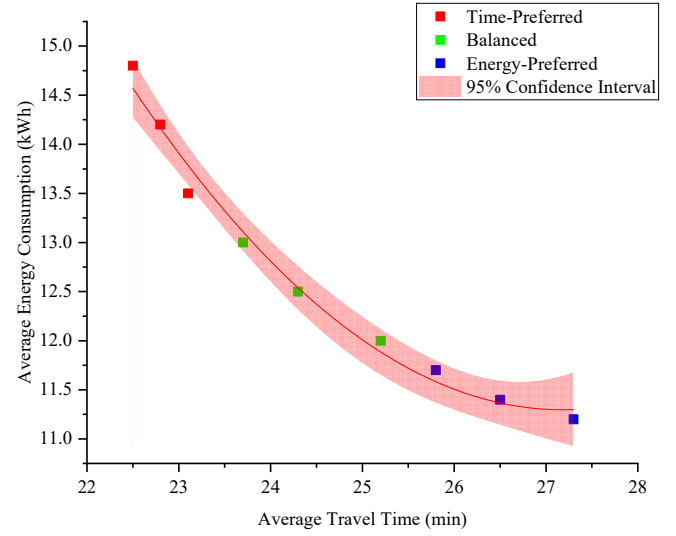
Table 3 GREEN-DRL ablation experiment analysis

Model variant	ACC (%)	F1-score	NDCG@5	Avg. time (min)	Avg. energy (kWh)	Energy education (%)
Full model	89.7	0.872	0.921	24.3	12.5	18.6
w/o prediction	84.1	0.823	0.872	25.9	13.8	10.2
w/o energy model	87.2	0.849	0.903	24	13.6	11.5
Single-objective	86.5	0.842	0.895	23.8	13.1	14.7

The weight sensitivity analysis indicates that the GREEN-DRL framework possesses excellent controllability and interpretability. Decision-makers can, based on specific business requirements (such as prioritising ‘timely delivery’ or ‘green delivery’), adjust the intuitive weight parameters to obtain the corresponding preferred planning strategies without modifying the core algorithm. This enhances the applicability and operability of the framework in actual logistics scheduling systems. We further analysed the ablation results under different traffic conditions. When removing the energy model (w/o EM), the energy consumption increase was more pronounced during congested periods (e.g., peak hours, +24% energy) than during free-flow periods (+9% energy). This indicates that

the energy model is especially critical in stop-and-go traffic where acceleration/deceleration losses dominate.

Figure 3 Pareto frontier distribution under multi-objective weight adjustment (see online version for colours)



6 Discussion

This section aims to conduct an in-depth analysis and comprehensive discussion of the experimental results from Section 4, clarifying the theoretical and practical significance of the GREEN-DRL model proposed in this study, objectively analysing its current limitations, and based on this, proposing forward-looking future research directions. The research findings not only verified the effectiveness of the proposed method, but also revealed the deep-seated challenges and opportunities in achieving intelligent logistics path planning within the green computing framework.

6.1 Discussion on model validity and advancement

The experimental results clearly demonstrate that the GREEN-DRL model outperforms traditional and single-objective optimisation algorithms in multiple performance metrics. Its success can be attributed to the synergy of the following three core designs:

Firstly, the successful application of the multi-objective DRL framework is the key to the model achieving Pareto superiority. Compared with the single-objective DQN algorithm, GREEN-DRL incorporates vector reward functions and preference weight mechanisms [equations (15) and (18)] to enable the agent to internalise the complex trade-off relationships among time, energy consumption, and cost during training. The results in Table 1 and Figure 3 jointly demonstrate that this model does not simply make linear trade-offs among different objectives, but is capable of learning a nonlinear optimisation strategy. This strategy surpasses static planning methods that only rely on historical average speed (such as A^*) and short-sighted optimisations that only respond to

current congestion (such as single-objective DRL), demonstrating true forward-looking planning capabilities.

Secondly, the closed-loop integration of the high-fidelity energy consumption model and real-time traffic prediction forms the cornerstone of green performance improvement. The ablation experiments (Table 2) clearly demonstrate that these two modules are indispensable. Removing the real-time prediction module (w/o PE) causes the model to degenerate into a planning based on historical averages, unable to cope with dynamic road conditions, and the energy-saving effect drops sharply. Removing the refined energy consumption model (w/o EM) makes the model lose the ability to accurately assess ‘green’ benefits, and its energy-saving effect is the same as that of the DQN that only optimises time. This model achieves a complete chain from ‘data’ to ‘physical effect’ and then to ‘decision feedback’ by directly inputting the fine-grained speed profile predicted by STGCN into the physical-based energy consumption model. This enables the model to identify those ‘green traps’ sections that seem not congested but actually have high energy consumption due to gradients and frequent acceleration and deceleration, thereby making more scientific decisions.

Finally, the model demonstrated excellent generalisation and robustness. The experiments were conducted on road networks that included various traffic patterns (off-peak and peak hours) and a large number of origin-destination pairs. GREEN-DRL maintained stable performance advantages in all cases. The smoothness of its convergence curve (Figure 2) also indicates that the multi-objective optimisation framework did not cause training instability. Instead, due to the richer reward signals, it helped the policy network escape local optima. This provided initial confidence for deploying the model in actual urban-level logistics scheduling systems.

6.2 Practical application potential and limitations

The framework proposed in this study demonstrates great potential for application in the fields of smart logistics and urban computing. At the practical level, this model can directly serve as the intelligent scheduling engine for the fleet management system of logistics companies. It receives real-time traffic information (such as from mapping service providers) and order information through API interfaces, and dynamically recommends personalised routes for each driver that balance timeliness and energy conservation. At the macro level, this model can be used by urban transportation management departments to evaluate and optimise freight channel policies, low-emission zone settings, etc. By simulating the path selection and total emissions of all freight vehicles in the city under different control measures, it provides quantitative decision support for the planning of green freight cities.

However, this study still has several limitations that need to be taken into account in future research and practical deployments:

- **Balancing cost and real-time performance:** Although the trained model runs relatively fast during inference, the entire system relies on real-time traffic prediction. The STGCN model has certain requirements for GPU computing power, and the higher the prediction frequency, the greater the system cost. In extremely complex road networks or applications that are extremely sensitive to latency (such as unmanned vehicle planning with millisecond-level response), it is necessary to further optimise the lightweighting of the prediction model or explore more efficient joint architectures for prediction and planning.
- **Data dependency and cold start problem:** The performance of the model is highly dependent on high-quality historical and real-time traffic data. For new urban areas or rural roads where traffic sensors are insufficiently covered, the prediction accuracy will decline, thereby affecting the planning outcome. Additionally, the model requires a certain adaptation period for new road patterns that have not been encountered before (such as long-term closures due to construction). How to alleviate data dependency by utilising transfer learning, small sample learning, or combining prior knowledge from traffic flow theory is an important issue.
- **Model interpretability and decision trust:** As a ‘black box’, the internal decision-making logic of the DRL model is not transparent to logistics managers and drivers. When the model recommends an unconventional route, users may refuse to adopt it due to a lack of understanding. Therefore, developing post-decision explanation methods for the decision paths of GREEN-DRL (such as visualising the model’s attention to key sections or traffic conditions through an attention mechanism) is crucial for enhancing user trust and system acceptability (Lundberg and Lee, 2020).

6.3 Future research direction

Although this research has achieved some interim results, there are still several directions that are worthy of in-depth exploration in the future. Firstly, the research can be extended to the scheduling of multi-agent collaborative electric truck fleets, taking into account the layout of charging facilities, dynamic orders and time window constraints, to build a system-level green logistics solution. Secondly, it is crucial to further enhance the practicality and robustness of the model, such as developing a risk-aware planning module that can explicitly quantify the uncertainty of traffic predictions, and exploring a distributed training framework based on federated learning, to achieve performance improvement of cross-regional models while protecting the privacy of commercial data (Lim et al., 2020). The advancement of these directions is expected to upgrade green path planning from single-point intelligence to collaborative, reliable and evolvable system intelligence,

providing stronger support for smart logistics and sustainable transportation.

7 Conclusions

This study aimed to address the challenges of multi-objective collaborative optimisation in logistics path planning under dynamic traffic conditions, and successfully designed and verified an innovative framework called GREEN-DRL. This framework integrates real-time traffic prediction, fine energy consumption modelling based on physical mechanisms, and multi-objective DRL, aiming to simultaneously optimise travel time, energy consumption, and transportation costs. Systematic simulation results show that the proposed method significantly outperforms traditional and single-objective comparison algorithms in multiple key indicators. It not only achieves a higher path recommendation accuracy (ACC reaching 89.7%) and sorting quality (NDCG@5 reaching 0.921), but also achieves a significant 18.6% reduction in actual energy-saving effect. Ablation studies further confirm that real-time prediction and high-fidelity energy consumption models are indispensable core components for achieving this green performance improvement.

The main theoretical contribution of this work lies in proposing and implementing a new closed-loop optimisation paradigm of ‘perception – evaluation – decision-making’. Firstly, by seamlessly integrating the dynamically predicted speed profiles from the spatiotemporal GNN into the vehicle dynamics energy consumption model, it achieves forward-looking and precise quantification of future energy consumption of the path, surpassing the traditional modelling approach that relies on static historical data. Secondly, the designed multi-objective DRL algorithm, with the aid of vector rewards and adaptive weight mechanisms, enables the intelligent agent to autonomously learn nonlinear trade-off strategies among complex goals, thereby stably generating a wide range of Pareto optimal solution sets, providing a systematic methodology tool for multi-objective decision-making.

Based on the solid theoretical and experimental foundation, this research has clear practical guiding value. For logistics operation enterprises, the GREEN-DRL framework can be integrated into the intelligent dispatching platform, using real-time traffic conditions to dynamically plan driving routes for fleets that are both efficient and low-carbon, thereby directly reducing fuel costs and carbon footprint. For urban traffic managers, this model can serve as a scientific assessment tool, used to simulate and analyse the potential impact of different freight control policies (such as the establishment of low-emission zones) on the freight energy consumption and emissions throughout the city, thereby providing quantitative basis for formulating sustainable urban logistics planning. Looking to the future, the research can further expand in directions such as multi-vehicle collaborative electric dispatching, integration of uncertainty in robust planning, and distributed model training based on privacy computing technology, to promote

the comprehensive evolution of the green intelligent logistics system. To extend the current single-agent formulation to multi-vehicle coordination, a natural direction is to adopt centralised training with decentralised execution, where a central critic has access to global information (e.g., all vehicle states and traffic predictions) while each vehicle’s policy acts locally. Alternatively, communication-efficient message-passing mechanisms (e.g., graph attention networks over vehicle agents) can be embedded to enable cooperative route adjustments.

Declarations

Author declares no conflicts of interest.

References

- Basso, R., Kulcsár, B., Egardt, B., Lindroth, P. and Sanchez-Diaz, I. (2019) ‘Energy consumption estimation integrated into the electric vehicle routing problem’, *Transportation Research Part D: Transport and Environment*, Vol. 69, pp.141–167.
- Bast, H., Delling, D., Goldberg, A.V., Müller-Hannemann, M., Pajor, T., Sanders, P., Wagner, D. and Werneck, R.F. (2016) ‘Route planning in transportation networks’, *ACM Computing Surveys*, Vol. 49, No. 1, pp.58–62.
- Braekers, K., Ramaekers, K. and Van Nieuwenhuysse, I. (2016) ‘The vehicle routing problem: State of the art classification and review’, *Computers & Industrial Engineering*, Vol. 99, pp.300–313.
- Clifton, J. and Laber, E. (2020) ‘Q-learning: theory and applications’, *Annual Review of Statistics and Its Application*, Vol. 7, No. 1, pp.279–301.
- Defferrard, M., Bresson, X. and Vandergheynst, P. (2017) ‘Convolutional neural networks on graphs with fast localized spectral filtering’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp.2456–2469.
- Franceschetti, A., Honhon, D., Van Woensel, T., Bektaş, T. and Laporte, G. (2013) ‘The time-dependent pollution-routing problem’, *Transportation Research Part B: Methodological*, Vol. 56, pp.265–293.
- Guillera-Aroita, G., Lahoz-Monfort, J.J., van Rooyen, A.R., Weeks, A.R. and Tingley, R. (2017) ‘Dealing with false-positive and false-negative errors about species occurrence at multiple levels’, *Methods in Ecology and Evolution*, Vol. 8, No. 9, pp.1081–1091.
- Haklay, M. and Weber, P. (2008) ‘OpenStreetMap: user-generated street maps’, *IEEE Pervasive Computing*, Vol. 7, No. 4, pp.12–18.
- Hart, P.E., Nilsson, N.J. and Raphael, B. (1968) ‘A formal basis for the heuristic determination of minimum cost paths’, *IEEE Transactions on Systems Science and Cybernetics*, Vol. 4, No. 2, pp.100–107.
- Kohl, W.L. (2019) ‘The international energy agency: the political context’, *Journal of International Energy Politics*, Vol. 12, No. 3, pp.246–257.
- Krajzewicz, D., Erdmann, J., Behrisch, M. and Bieker, L. (2012) ‘Recent development and applications of SUMO-simulation of urban mobility’, *International Journal on Advances in Systems and Measurements*, Vol. 5, No. 3, pp.128–138.

- Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D. and Miao, C. (2020) 'Federated learning in mobile edge networks: a comprehensive survey', *IEEE Communications Surveys & Tutorials*, Vol. 22, No. 3, pp.2031–2063.
- Lundberg, S.M. and Lee, S.-I. (2020) 'A unified approach to interpreting model predictions', *Nature Machine Intelligence*, Vol. 2, No. 5, pp.252–262.
- Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.-Y. (2014) 'Traffic flow prediction with big data: a deep learning approach', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 2, pp.865–873.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, A., Sadik, I., Antonoglou, H., King, H., Kumaran, D., Wierstra, S., Legg, S. and Hassabis, D. (2015) 'Human-level control through deep reinforcement learning', *Nature*, Vol. 518, No. 7540, pp.529–533.
- Ntziachristos, L., Gkatzoflias, D., Kouridis, C. and Samaras, Z. (2009) 'COPERT: a European road transport emission inventory model', *Environmental Modelling & Software*, Vol. 24, No. 12, pp.1477–1487.
- Rakha, H., Ahn, K. and Trani, A. (2004) 'Development of VT-micro model for estimating hot stabilized light duty vehicle and truck emissions', *Transportation Research Part D: Transport and Environment*, Vol. 9, No. 1, pp.49–74.
- Reed, P., Minsker, B. and Goldberg, D.E. (2000) 'Designing a competent simple genetic algorithm for search and optimization', *Water Resources Research*, Vol. 36, No. 12, pp.3757–3761.
- Rong, H., Zhang, H., Xiao, S., Li, C. and Hu, C. (2016) 'Optimizing energy consumption for data centers', *Renewable and Sustainable Energy Reviews*, Vol. 58, pp.674–691.
- Sarkar, N.I. and Gul, S. (2020) 'Green computing and internet of things for smart cities: technologies, challenges, and implementation', *Sustainable Computing: Informatics and Systems*, Vol. 28, pp.35–50.
- Sun, T., Liu, J., Kan, J. and Sui, T. (2022) 'Research on target classification method for dense matching point cloud based on improved random forest algorithm', *International Journal of Information and Communication Technology*, Vol. 21, No. 3, pp.290–303.
- Sutton, R.S., McAllester, D., Singh, S. and Mansour, Y. (2003) 'Policy gradient methods for reinforcement learning with function approximation', *Journal of Machine Learning Research*, Vol. 4, No. 7, pp.1101–1125.
- Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C. (2014) 'Short-term traffic forecasting: where we are and where we're going', *Transportation Research Part C: Emerging Technologies*, Vol. 43, pp.3–19.
- Vujović, Ž. (2021) 'Classification model evaluation metrics', *International Journal of Advanced Computer Science and Applications*, Vol. 12, No. 6, pp.599–606.
- Yao, E., Yang, Z., Song, Y. and Zuo, T. (2013) 'Comparison of electric vehicle's energy consumption factors for different road types', *Discrete Dynamics in Nature and Society*, Vol. 2013, No. 1, p.328757.
- Yu, B., Yin, H. and Zhu, Z. (2017) 'Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 11, pp.2397–2407.
- Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X. and Li, T. (2018) 'Predicting citywide crowd flows using deep spatio-temporal residual networks', *Artificial Intelligence*, Vol. 259, pp.147–166.
- Zheng, H., Lin, F., Feng, X. and Chen, Y. (2020) 'A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, No. 11, pp.6910–6920.