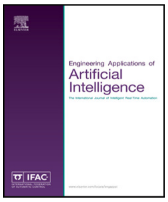
Engineering Applications of Artificial Intelligence 149 (2025) 110392

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Reinforcement learning based multi-perspective motion planning of manned electric vertical take-off and landing vehicle in urban environment with wind fields

Songyang Liu [a], Weizi Li [b], Haochen Li [c], Shuai Li [a,*]

[a] ESSIE - Engineering School of Sustainable Infrastructure and Environment at University of Florida, Gainesville, FL, USA
[b] Min H. Kao Department of Electrical Engineering and Computer Science at University of Tennessee, Knoxville, TN, USA
[c] Department of Civil and Environmental Engineering at University of Tennessee, Knoxville, TN, USA

ARTICLE INFO

Keywords:
Reinforcement learning
Electric vertical take-off and landing
Motion planning
Urban wind field

ABSTRACT

Electric vertical-takeoff and landing (eVTOL) aircraft, known for their maneuverability and flexibility, offer a promising alternative to traditional transportation systems. However, these aircraft face significant challenges from various perspectives, including the need to increase energy efficiency, enhance passenger experience, and mitigate noise impact on urban environments. While mathematical modeling-based approaches have been employed for flight motion planning, they often struggle to adapt to dynamic and complex environments. In this work, we introduce a three-dimensional motion planning method based on deep reinforcement learning (DRL), tailored for manned eVTOL flights through urban wind fields. Our approach considers three crucial aspects: aircraft energy consumption, passenger experience, and noise impact on urban environment. We modify the Proximal Policy Optimization (PPO) algorithm and design comprehensive reward function that considers these objectives. By incorporating energy efficiency, passenger experience, and noise impact into our reward function, our method demonstrates improved policy learning compared to existing approaches. Comparative experiments conducted under various wind conditions show that our method outperforms commonly used techniques, effectively optimizing multiple objectives in challenging urban environments. Code of our work are available at https://github.com/cgchrfchscyrh/eVTOL_RL/tree/main.

1. Introduction

Urban air mobility (UAM) is an efficient transportation system where everything from small package-delivery drones to passenger-carrying air taxis operate over urban areas (Kelsey, 2023). The advent of eVTOL aircraft heralds a promising future for UAM, aiming to revolutionize transportation by reducing congestion, environmental pollution, and travel time. The continuous growth in urban populations and the corresponding increase in road traffic underscore the urgent need for innovative transportation solutions. Moreover, the integration of eVTOL into the urban fabric requires meticulous planning and optimization across multiple domains, including energy consumption, flight duration, noise levels, and payload management, to ensure operational efficiency and public acceptance (Kleinbekman et al., 2018). Companies are making progress towards government approval and real-world implementation of eVTOL (Anon, 2023, 2024).

Unlike ground vehicles, which are restricted by roads and traffic, eVTOL has the freedom to choose from a wide array of flight paths. However, this flexibility introduces variability in energy consumption and flight durations depending on the chosen route. Additionally, the complex wind patterns generated by urban landscapes and terrain (Ware and Roy, 2016), along with other unpredictable environmental factors, further complicate the task of planning and optimizing eVTOL flight paths (Hong et al., 2021a).

Traditional mathematical models have been applied to aircraft flight path planning (Forkan et al., 2022; Chen et al., 2023), but they often face challenges in addressing risk assessment and multi-objective optimization, especially in densely populated urban areas (Babu et al., 2022; Wang et al., 2023a). Moreover, these models can struggle to quickly adapt to changing environmental conditions. In contrast, machine learning-based approaches offer greater adaptability, robustness, and efficiency. They are particularly effective in managing dynamic and uncertain environments, adjusting to real-time changes, and are better suited for scaling in complex urban scenarios (Ramezani et al., 2023; Tu and Juang, 2023; Maciel-Pearson et al., 2019) (see Fig. 1).

[*] Correspondence to: P O Box 116580, 460C Weil Hall, Gainesville, FL 32611, USA.
E-mail addresses: liusongyang@ufl.edu (S. Liu), weizili@utk.edu (W. Li), hli111@utk.edu (H. Li), shuai.li@ufl.edu (S. Li).

https://doi.org/10.1016/j.engappai.2025.110392
Received 12 August 2024; Received in revised form 25 January 2025; Accepted 21 February 2025
Available online 10 March 2025

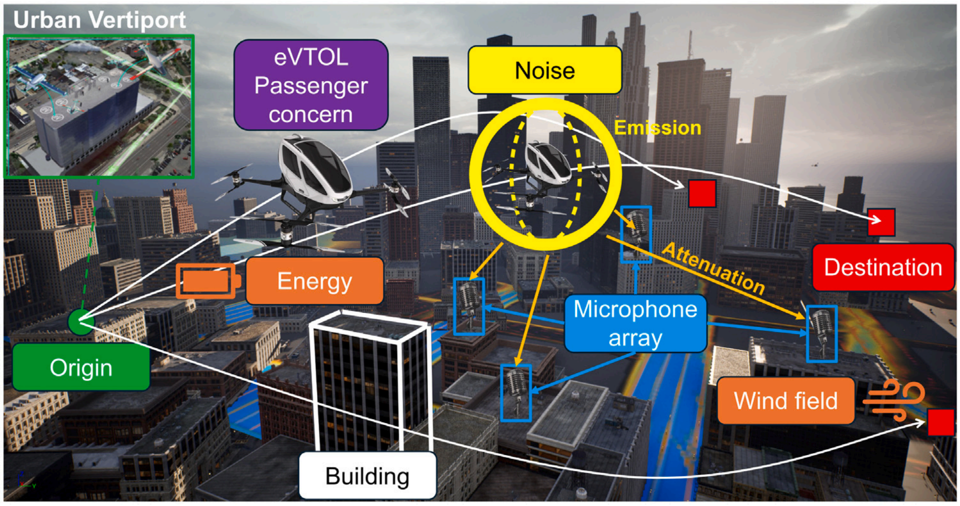


Fig. 1. Multi-objective motion planning of eVTOL aircraft flight, from origin to destination, through urban wind fields enabled by deep reinforcement learning. We consider eVTOL aircraft energy consumption and efficiency, passenger concern and noise impact on urban environment as objectives. We only study the flight phase, not takeoff and landing phase of eVTOL. Partial image taken from Rizzi et al. (2020), Pradeep and Wei (2018). Figure not to scale.

In this paper, our application scenario is eVTOL aircraft carries passenger from an origin to a destination in a simulated urban environment with wind fields. We do not consider the takeoff and landing phases of eVTOL flight because the planning flexibility for these phases is limited compared to the free-flight portion, where more extensive planning and decision-making are required. We use a deep reinforcement learning (DRL) method to plan the eVTOL flight from three different perspectives: eVTOL aircraft energy, passenger concerns and environmental impact. We explain the importance of the three chosen perspectives in following paragraphs.

Energy consumption and efficiency are critical for eVTOL flights in urban environments. Lower energy use extends flight range and reduces operational costs, while high efficiency ensures that flights are both sustainable and economically viable. In dense urban settings, optimizing these factors is key to minimizing environmental impact and making eVTOL operations practical and reliable.

It is a critical challenge to ensure the security of the electric and flying vehicle systems (Alqahtani and Kumar, 2024). For passenger concerns with respect to a new transportation option, the most important factors are safety, followed by reliability, time savings, convenience and comfort (Edwards and Price, 2020). We choose the safety, time savings and comfort to be the three objectives for passenger concerns perspective. We explain the integration of these factors into our DRL method in Section 3.

For urban environment, noise is important. New noise exposure and annoyance from autonomously controlled vehicles could limit the success of integrating UAM into the transportation system (Rizzi et al., 2020; National Academies of Sciences and Division on Engineering and Physical Sciences and Aeronautics and Space Engineering Board and Committee on Enhancing Air Mobility A National Blueprint, 2020; Holmes et al., 2017; Holden and Goel, 2016). If eVTOL noise is not properly considered, there is a risk that the rapid uptake of UAM aircraft may outpace the regulatory framework and ultimately fail due to a public backlash (Jackson and Bardell, 2023). Limiting the noise impact on the population at an early stage is crucial, even more so because the opinion of the population on drones still seems to be forming (Schäffer et al., 2021).

The three perspectives mentioned above are essential to the practical deployment and broad public acceptance of eVTOL, particularly when operating in complex urban wind fields. However, there is a research gap in the realm of eVTOL flight motion multi-objective optimization. Many existing studies focus on one or two perspectives (Bhalla et al., 2024; Su et al., 2024; Nagashima et al., 2022), while overlooking other crucial perspectives. To address this gap, we propose a multi-objective optimization approach that simultaneously considers the three important perspectives including energy consumption, flight time, safety, passenger comfort, and noise. Our integrated strategy not only seeks high performance in technical terms but also accounts for the social and environmental impacts of eVTOL operations. Unlike traditional approaches that isolate these objectives, our method integrates them into a balanced solution, recognizing the intricate trade-offs and interdependencies among often competing targets. Our integrated approach represents a step towards practical and large-scale eVTOL deployment in complex populated urban areas, ensuring that both technical and societal requirements are effectively met.

The main contributions of this paper are listed as follows:

- We introduce a three-dimensional motion planning method for manned eVTOL navigating through urban wind fields. Our approach, enabled by DRL, considers three important perspectives: aircraft energy, passenger concerns and noise impact on environment.
- We modify Proximal Policy Optimization (PPO) (Schulman et al., 2017) and tailor it to our purpose. We compare our modified PPO with the vanilla PPO to demonstrate our advantage.
- We design a RL reward function architecture for minimizing eVTOL energy consumption and maximizing energy efficiency. We compare our design with existing approaches to demonstrate our advantage. We further perform an ablation study to demonstrate the effectiveness of our reward function components.
- We analyze the factors need to be considered for eVTOL flight with passenger and design quantitative metric for evaluating the passenger experience and noise impact on environment. Then, we incorporate the energy, passenger concerns and noise impact into our reward function design.
- We conduct comparative experiments under various wind fields and RL algorithms, including the state of the art model-based and model-free RL algorithms. The results show that our method has advantages over other commonly used methods.

2. Related work

2.1. Motion planning

During movement, path planning plays an important role in achieving the autonomous operation of the vehicle (Zhao et al., 2024). Motion planning for aerial vehicles typically relies on two main approaches: mathematical modeling and machine learning. Mathematical models often utilize heuristic algorithms, such as ant colony optimization for enhancing search capabilities in aircraft routing (Al-Habob et al., 2021; Wan et al., 2023; Huan et al., 2021), or cuckoo search for designing energy-efficient paths in wireless networks (Zhu et al., 2021). Other examples include Chodnicki et al.'s aircraft model that considers forces and moments (Chodnicki et al., 2022), and the use of mixed-integer linear programming for multi-vehicle path planning in urban environments with varying obstacle heights (Bahabry et al., 2019). Although these models can yield precise solutions, their computational intensity and complexity limit their practicality in real-time, dynamic situations (Yao et al., 2014; Yang et al., 2020; Zhang et al., 2020; Bai et al., 2021; Liu et al., 2021; Sandino et al., 2022; Zhou et al., 2023). On the other hand, machine learning methods are more adaptable and efficient, particularly in dynamic and uncertain settings, due to their ability to adapt to dynamic obstacles and varying conditions (Ramezani et al., 2023; Tu and Juang, 2023; Maciel-Pearson et al., 2019).

Among machine learning techniques, RL has proven effective for aircraft motion planning. For instance, Xu et al. developed a DQN-based algorithm to navigate around obstacles (Xu et al., 2022), while Li et al. introduced a stepwise DQN approach to identify common features across navigation targets (Li and Liu, 2022). Wang et al. implemented a D3QN method for real-time aerial vehicle navigation (Wang et al., 2022), and Luna et al. showcased DQN's ability to achieve optimal mission coverage (Luna et al., 2022). Recent advancements have shifted towards Policy Gradient (PG) methods, which offer faster convergence and better adaptability. Techniques like Deep Deterministic PG (DDPG) have been used to adjust flight altitudes for aircraft (Qiu et al., 2022), and Twin Delayed DDPG (TD3) has been applied to optimize responses for obstacle avoidance (Liu et al., 2022; Zhang et al., 2023a,b; Hu et al., 2023). Given TD3's sensitivity to hyperparameters, PPO has become the preferred choice due to its stability and efficiency in policy adjustments, making it well-suited for our task (Xu et al., 2024). Further, we implement the state-of-the-art model-based RL algorithm named TD-MPC2 (Hansen et al., 2023) and a recent model-free RL algorithm named Average-TD3 (Luo et al., 2024), to solve the motion planning with multi-objective optimization problem. To the best of our knowledge, this is the first implementation of the TD-MPC2 algorithm for multi-objective optimization in eVTOL flight motion planning. Olivares et al. (2024) implement TD-MPC2 on fixed-wing UAV attitude control under varying wind conditions. However, their approach did not account for the impact of noise generated by UAV flight on urban environments.

2.2. Multi-objective optimization

Many real-world tasks involve multiple, possibly competing, objectives. For instance, choosing a financial portfolio requires trading off between risk and return; controlling energy systems requires trading off performance and cost; and autonomous cars must trade off fuel costs, efficiency, and safety (Abdolmaleki et al., 2020). One common approach to multi-objective sequential decision-making problem is to adopt an axiomatic approach in which the optimal solution set is assumed to be the Pareto front (Akbari et al., 2014). However, this set is typically large, and may be prohibitively expensive to retrieve (Hayes et al., 2022).

Another common approach, the scalarization method, makes the multi-objective function create a single solution and the weight is determined before the optimization process. The scalarization method incorporates multi-objective functions into a scalar fitness function as shown in the following equation (Gunantara, 2018; Murata et al., 1996):

$$F(x) = w_1 f_1(x) + w_2 f_2(x) + \cdots + w_n f_n(x)$$

Such methods have been demonstrated to be effective in using RL methods to deal with aircraft control tasks. For instance, researchers design and integrate linearly combined RL reward function into an autonomous system that can race drones at the level of the human world champions (Kaufmann et al., 2023). In this work, we choose to use the linear scalarization method. To deal with the multiple objectives of our application, we combine all the important aspects together into a single scalar additive reward function. We assign numerical rewards or penalties to events that can occur in the environment.

Many researchers have studied MOO problem in the realm of aircraft. Ye et al. formulate three optimization problems: a sum-throughput maximization problem, a total-time minimization problem, and a total-energy minimization problem in rotary-wing unmanned aerial vehicle (UAV)-enabled full-duplex wireless-powered Internet-of-Things (IoT) networks (Ye et al., 2020). Yu et al. jointly optimize three objectives: maximization of sum data rate, maximization of total harvested energy and minimization of UAV's energy consumption in UAV-assisted IoT network (Yu et al., 2021). Wu et al. study the tradeoff between the energy and time consumption for UAV-enabled wireless-powered communication network (Wu et al., 2019). Song et al. propose an evolutionary multi-objective RL algorithm to minimize the task delay and the UAV's energy consumption, and maximize the number of tasks collected by the UAV in a mobile edge computing system (Song et al., 2023). However, none of these works have treated the noise production of aircraft as one objective to optimize, which can harm the passenger comfort and the health of nearby urban residents as shown in Table 1.

2.3. Reward function design for energy

The reward function design is critical for RL method. To lower the energy consumption, there are two common approaches: "Direct energy consumption" (Hong et al., 2021b; Yu et al., 2021; Song et al., 2022, 2024; Liu et al., 2023; Zhang and Cao, 2022; Guo et al., 2023) and "Energy efficiency" (Abedin et al., 2020; Liu et al., 2019, 2018; Omoniwa et al., 2022; Dai et al., 2021; Chen et al., 2020; Fu et al., 2021; Li et al., 2021; Qi et al., 2020; Nie et al., 2020). "Direct energy consumption" means directly using the product of a predefined weight w and the value of energy consumption in current training step E to be the reward signal R:

$$R = w * E$$

"Energy efficiency" means dividing the gain G by the energy consumption E:

$$R = G/E$$

The definition of the gain can be adapted to different application scenarios. For example, for communication related work, the gain can be the wireless data received by drone in a short time period. In this work, the gain is defined as the distance (m) that the agent can approach the destination for every unit of energy consumed (kWh).

Moreover, some researchers have proposed unique design different from the two common approaches, such as normalizing energy consumption with specific math equation, introducing virtual energy queue and compute the reward based on the state-of-charge level (Qi et al., 2019; Arani et al., 2021; Do et al., 2021). We name the reward function design in Arani et al. (2021) as "Math". The reward function for each UAV is defined as follows. $\eta_{u,c}$ and $\eta_{u,e}$ denote the weight parameters that indicate the impact of throughput and energy consumption. ϕ, φ, ϵ are the adjustable control parameters for the Gompertz function $G(t)$.

$$Y_u(t) = \eta_{u,c} F\left(C_u(t)\right) + \eta_{u,e} G\left(E_u^{\text{Total}}(t)\right)$$

$$G(t) = \phi - \phi e^{-\varphi e^{-\epsilon t}}$$

We refer to the reward function design in Do et al. (2021) as "Queue". The objective function is defined as a weighted sum of the system reliability-of-learning R_t^s and the energy consumption E_t at the

Table 1
Optimization objectives comparison between reference using RL method to deal with aircraft problems. Minimize the energy consumption of aircraft flight is the most common objective.

Objective	Ours	Hong et al. (2021b)	Yu et al. (2021)	Song et al. (2022)	Song et al. (2024)	Liu et al. (2023)	Zhang and Cao (2022)	Guo et al. (2023)
Number of objectives	6	1	3	3	2	2	3	2
Min energy consumption	✓	✓	✓	✓	✓	✓	✓	✓
Max energy efficiency	✓						✓	
Max sum data rate			✓					
Max harvested energy			✓					
Min delayed tasks				✓		✓		
Max collected tasks				✓	✓			
Max system throughput							✓	
Max system achievable rate								✓
Min time consumption	✓							
Min noise production	✓							
Max passenger safety	✓							
Max passenger comfort	✓							

Table 2
Algorithm comparison between reference dealing with aircraft problems. DDPG algorithm is commonly used to deal with aircraft problems. Dijkstra and RRT algorithm are used as baseline to compare with RL algorithms.

	Algorithm	Ours	Hong et al. (2021b)	Yu et al. (2021)	Zhang et al. (2022)	Song et al. (2022)	Song et al. (2024)	Liu et al. (2023)	Zhang and Cao (2022)	Guo et al. (2023)	An et al. (2023)
RL	PPO	✓									
	TD3	✓	✓								
	DDPG	✓	✓	✓					✓	✓	
	Particle swarm optimization with RL				✓						
	Evolutionary MORL					✓					✓
	Envelope MO Q-learning						✓				
	Double/Dueling DQN							✓			
	TD-MPC2	✓									
	Average TD3	✓									
Non-RL	Dijkstra		✓								
	RRT		✓								

UAV as follows. A virtual energy queue ψ_t for the UAV is introduced. E_b is the allowable energy budget of the UAV and T is the length of a time horizon.

$$\max \mathbb{E}\left[\sum_{t=1}^{T}\gamma^{t-1}\left(R_t^s - \delta\,\psi_t E_t\right)\right]$$

$$\psi_{t+1} = \max\left\{\psi_t + E_t - \frac{E_b}{T}, 0\right\}$$

However, these approaches limit the RL agent to explore better flight actions with lower energy consumption. We propose a new reward function design architecture and demonstrate its advantages over the two common approaches, ‘Direct’ and ’Efficiency,’ as well as the other two approaches, ‘Math’ and ’Queue,’ through comparative experiments. We use DDPG algorithm as a major part of our experiments because DDPG is commonly used in related works as shown in Table 2. The experiment details are explained in Section 4.

3. Methodology

Our work pipeline is shown in Fig. 2. We begin by explaining our problem formulation and learning techniques, followed by the introduction of simulation data source and details.

3.1. Multi-objective eVTOL aircraft motion planning

We formulate our problem with six objectives categorized into three perspectives: passenger concerns, eVTOL aircraft concerns and environment concern over a task period. Fig. 3 shows the interdependencies highlighting the complexity of eVTOL flight motion planning to consider energy, noise impact, and passenger concerns. Passenger concerns include: ❶ maximization of safety, ❷ minimization of traveled time, ❸ maximization of comfort. eVTOL concern includes: ❹ minimization of energy consumption, ❺ maximization of energy efficiency. Environment concern includes: ❻ minimization of noise produced by eVTOL. We explain the definition and calculation method of the objectives in the observation space part.

We formulate our task as a Partially Observable Markov Decision Process (POMDP) represented by a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, T, \Omega, \mathcal{O})$ where: $\mathcal{S}$ is the state space; $\mathcal{A}$ is the action space; $\mathcal{P}(s'|s,a)$ is the transition probability function; $\mathcal{R}$ is the reward function; $\gamma \in (0,1]$ is the discount factor; T is the episode length (horizon); Ω is the observation space; and $\mathcal{O}$ is the probability distribution of retrieving an observation $\omega \in \Omega$ from a state $s \in \mathcal{S}$. At each timestep $t \in [1,T]$, an eVTOL uses its policy $\pi_\theta(a_t|o_t)$ to take an action $a_t \in \mathcal{A}$, given the observation $o_t \in \mathcal{O}$. Next, the environment provides feedback on action a_t by calculating a reward r_t and transitioning the agent into the next state s_{t+1}. The eVTOL's goal is to learn a policy π_θ that maximizes the discounted sum of rewards, i.e., return, $R_t = \sum_{i=t}^{T}\gamma^{i-t}r_i$.

3.1.1. Action space

The action space consists of continuous actions $\in [-1,1]$ along the X-, Y-, Z-axis, respectively. When $a_x > 0$, the eVTOL advances in the positive X-direction; if $a_x = 0$, the aircraft remains stationary along the X-axis; if $a_x < 0$, the aircraft moves towards the negative X-direction. The same goes for Y- and Z-axis.

$$A = \{(a_x, a_y, a_z)\}, a_x, a_y, a_z \in [-1,1].$$

The agent determines the eVTOL's velocity by multiplying the actions by the maximum velocity V_{max}, $v_x = a_x * V_{max}$. We set the maximum velocity to be 60 m/s. The agent acquires the state and executes the action every second, $T_{step} = 1\ s$. Upon executing an action, the aircraft transitions from one 3D coordinate to another.

$$V = \{(v_x, v_y, v_z)\}, v_x, v_y, v_z \in [-V_{max}, V_{max}].$$

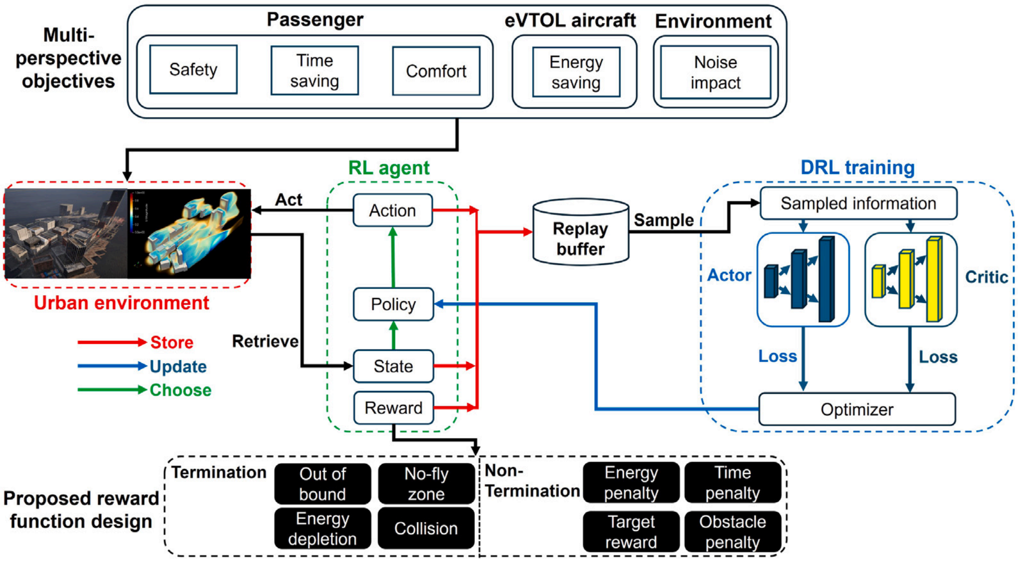


Fig. 2. Overview of the RL framework for eVTOL flight motion planning in urban environment. The framework addresses multi-perspective objectives from passengers, eVTOL aircraft, and the environment, focusing on safety, time-saving, comfort, energy-saving, and noise impact. The RL agent interacts with the urban environment, storing actions, updating policies, and retrieving states to maximize rewards and minimize penalties. The design incorporates a DRL training process with actor-critic methods, optimizing the policy through sampled information and loss minimization. The proposed reward function design includes termination conditions (out of bounds, energy depletion) and non-termination penalties (energy, time, obstacle penalties) along with target rewards.

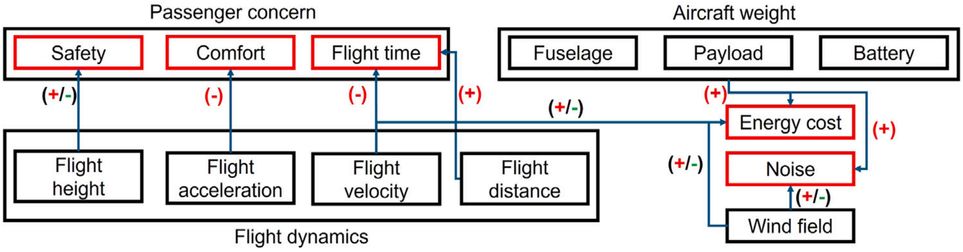


Fig. 3. Interrelated factors of eVTOL flight with passenger concerns in urban wind fields. Plus and minus signs indicate factors that increase or decrease the respective factor. Red-bordered boxes indicate the factors we focus on this study. Key passenger concerns include safety, comfort, and flight time. Aircraft weight factors – fuselage, payload, and battery – directly affect energy cost and flight dynamics. The wind field can have both positive and negative impacts on energy cost. Flight speed and distance influence both energy cost and noise, where higher flight speeds reduce flight time but increase noise, and longer distances raise energy consumption.

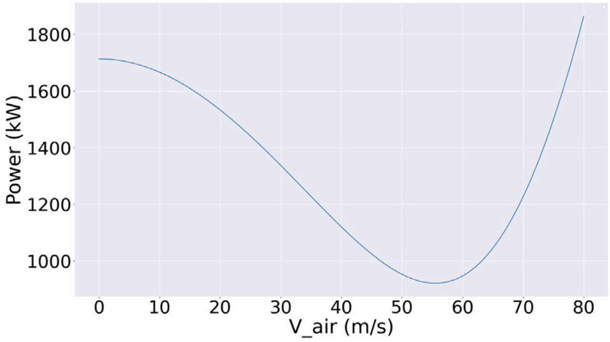


Fig. 4. eVTOL aircraft power consumption versus horizontal airspeed V_{air}.

Table 3
Modified value of parameters in the power model.

Parameter	Value
k_1	0.8554
k_2	2.774
c_2	10.1664
c_3	0
c_4	0.444
c_5	0.6696
c_6	0
m	305 kg

3.1.2. Observation space

To enable an RL policy to generalize across a variety of scenarios, we transform the conditions each eVTOL observes into a fixed-length representation, which includes the following.

- Position info: We use p^t to represent the 3D coordinates of aircraft. We further define des as the position of the flight destination and det as the distance to the nearest building or environment boundary in six directions: front, back, left, right, up, down:

$$det = \{d_{\text{front}}, d_{\text{back}}, d_{\text{left}}, d_{\text{right}}, d_{\text{up}}, d_{\text{down}}\}.$$

- Wind field: The wind vector is defined as:

$$W = (x, y, z, u, v, w),$$

where (x, y, z) represents a point in the simulation area and (u, v, w) represents the wind velocity at (x, y, z).

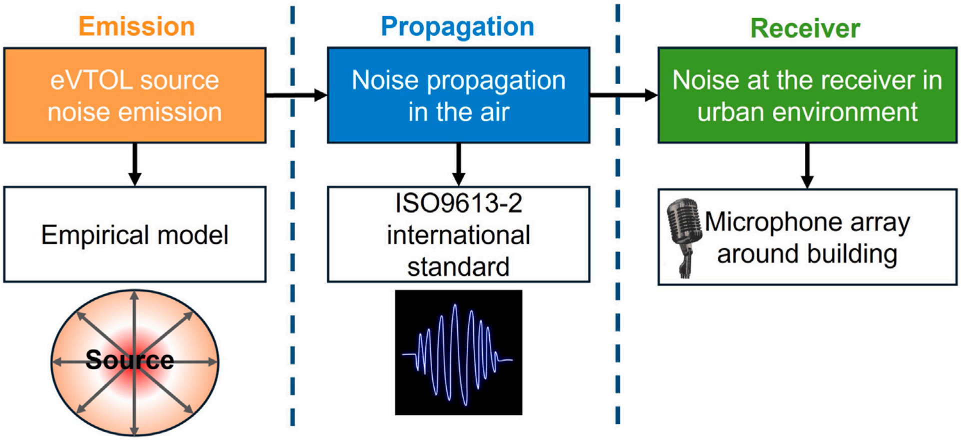

Emission
Propagation
Receiver
eVTOL source noise emission
Noise propagation in the air
Noise at the receiver in urban environment
Empirical model
ISO9613-2 international standard
Microphone array around building
Source

Fig. 5. Noise emission, propagation, and receiver framework for eVTOL flight in urban environment. In the emission stage, noise generated by the eVTOL source is modeled using existing empirical method. During propagation, noise travels through the air according to the ISO9613-2 international standard, which details the attenuation of sound during outdoor propagation. We note that we do not consider the effects of local wind, atmospheric conditions, and reflections due to reflective structures (e.g., buildings) on sound propagation. In the receiver stage, the noise level at various points in the urban environment is measured using a microphone array arranged around buildings.

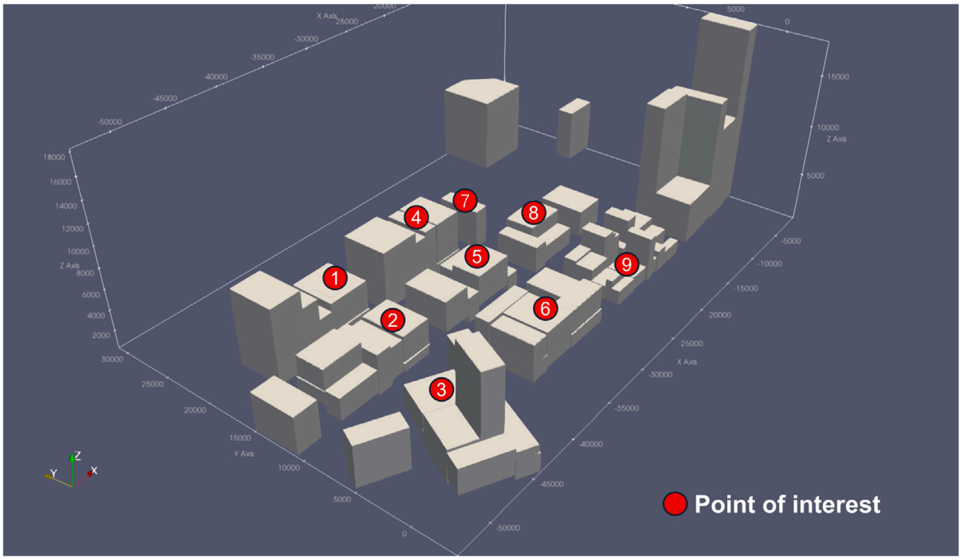


Fig. 6. The locations of chosen POIs. The nine POIs are placed dispersedly on the top of buildings, to cover the majority area of urban environment.

- Passenger concerns: For passenger concerns with respect to a new transportation option, the most important factor is safety, followed by reliability, time savings, convenience and comfort (Edwards and Price, 2020). We choose the safety, time saving and comfort as the three objectives as passenger concerns.
  - Safety: First, if the acceleration exceeds the set safety threshold, a negative reward is applied to reduce the impact of the acceleration change on ascent safety and comfort. Second, conflict detection and avoidance is considered as one part of passenger perceptions of safety (Edwards and Price, 2020). If the distance between eVTOL and the building in any direction is below the safety threshold, a negative reward is applied. Third, if the wind speed exceeds the set maximum safe wind speed threshold, a negative reward is applied to ensure that eVTOL can remain stable in high wind speeds.
  - Time saving: The objective is to minimize the flight time consumption from origin to destination.
  - Comfort: First, we use longitudinal acceleration (xy axis) < 0.4–$0.6\ g$, vertical acceleration $< 0.1\ g$ to measure the comfort factor C. Every $0.02/3 = 0.007\ g$ increases by 10% uncomfority. Second, interior noise for UAM cabins could become an issue for passenger comfort (Rizzi et al., 2020), so we also consider the source noise produced by the eVTOL aircraft as one part of the passenger comfort.

- eVTOL concern: We calculate the energy consumption E_t at timestep t of eVTOL flying with air resistance by following Liu et al. (2017) proposed power P model:

$$E_t = P * T_{step}$$

$$P = P_i(F_T, V_{vert}) + P_p(F_T, V_{air}) + P_{par}(V_{air})$$

$$P_i(F_T, V_{vert}) = k_1 T \left[\frac{V_{vert}}{2} + \sqrt{\left(\frac{V_{vert}}{2}\right)^2 + \frac{F_T}{k_2^2}} \right]$$

$$P_p(F_T, V_{air}) = c_2 F_T^{3/2} + c_3 (V_{air} \cos\alpha)^2 F_T^{1/2}$$

$$P_{par}(V_{air}) = c_4 V_{air}^3$$

$$F_T = \sqrt{(mg - (c_5 (V_{air}\cos\alpha)^2 + c_6 F_T))^2 + (c_4 V_{air}^2)^2}$$

$$V_{air} = \|\mathbf{V}_{air}\| = \|\mathbf{V}_{ground} - \mathbf{V}_{wind}\|$$

where P_i, P_p, P_{par} are the induced power, profile power and parasite power, respectively. F_T is the thrust, m is the total takeoff mass of the eVTOL aircraft, g is the gravity acceleration (i.e., 9.8 m/s^2), α is the angle of attack. V_{vert} is the vertical speed, V_{air} is the horizontal airspeed, $\mathbf{V}_{air}$, $\mathbf{V}_{ground}$, $\mathbf{V}_{wind}$ are the horizontal air velocity, ground velocity, and wind velocity, respectively. $k_1, k_2, c_2, c_3, c_4, c_5, c_6$ are dimensionless parameters to be identified.

We note that Liu et al. (2017) identifies the parameters of the power model by flying a small drone, we do not directly use the value of these identified parameters from Liu et al. (2017) because our envisioned application scenario requires a much larger aircraft capable of taking passenger, rather than small package. We follow the analytical expression of the parameters in Liu et al. (2017) and modify the value based on eVTOL aircraft related work (Pradeep and Wei, 2018). The modified value of parameters are shown in Table 3. Fig. 4 shows the power curve calculated by modified parameters.

- Environment concern: As discussed in Section 1, noise production of UAM needs to be considered. We consider the noise produced by eVTOL aircraft and received by buildings as the environment concern. Fig. 5 shows the noise calculation process. The equivalent continuous downwind octave-band sound pressure level at a receiver location $L_{fT}(DW)$ (International Organization for Standardization, 1996) is defined as:

$$L_{fT}(DW) = L_w + D_c - A$$

where L_w is the octave-band sound power level (SPL), in decibels, produced by the point sound source relative to a reference sound power of one picowatt; For an omnidirectional point sound source radiating into free space, $D_c = 0$ dB; A is the octave-band attenuation, in decibels, that occurs during propagation from the point sound source to the receiver.

  - Emission L_w: The SPL produced by eVTOL aircraft flight is defined as Schmähl et al. (2021):

$$L_w = 10 \cdot \log_{10}\left(10^{\left(L_{directivity+thrust}+L_{spreading}\right)\frac{1}{10}} + 10^{\left(L_{background}\right)\frac{1}{10}}\right)$$

$$L_{directivity+thrust}(\bar{\phi}, \vartheta, P_{el}, \vec{p}) = p_1 \cdot \log(P_{el}) \cdot \frac{f_{directivity}(\bar{\phi}, \vartheta, \vec{p})}{\left[f_{directivity}(\bar{\phi}, \vartheta, \vec{p})\right]_{max}}$$

$$f_{directivity}(\bar{\phi}, \vartheta, \vec{p}) = \left(1 + p_2 \cdot \sin(\bar{\phi} - p_3) \cdot \left(\vartheta + \frac{\pi}{2}\right)\left(\frac{\pi}{2}\right)\right) \cdot \left(1 + p_4 \cdot \sin(\vartheta - p_5)\right)$$

$$L_{background}(\vec{p}) = p_6$$

  where ϕ, ϑ is the azimuth and polar angle between POI and eVTOL aircraft, P_{el} is the eVTOL power consumption, and $\vec{p}$ is the model input parameter vector, detailed in Schmähl et al. (2021).

  We note that we set $L_{spreading}$ to be zero because the propagation loss is already included in attenuation A. Since Schmähl et al. (2021) does not give a detailed calculation method for P_{el}, we set $P_{el} = \frac{P}{5}$, to match the noise level produced by manned aircraft in real life.
  - Attenuation A: We only consider the geometrical divergence A_{div}, accounts for spherical spreading in the free field from a point sound source, making the attenuation.

$$A = A_{div} + A_{others}$$

$$A_{div} = [20lg(d/d_0) + 11]dB$$

  where d is the distance from the source to receiver, in meters, d_0 is the reference distance (= 1 m).
  - Receiver: Fig. 6 shows the nine point of interest (POI) on the top of buildings to measure the eVTOL flight's effect on urban environment. Each POI receives same background noise $L_{background}$ for simplicity and evaluation of eVTOL flight noise.

Overall, the observation space of an eVTOL at t is:

$$o^t = \langle p^t \rangle \oplus \langle des \rangle \oplus \langle det \rangle \oplus \langle W \rangle \oplus \langle E_t \rangle \oplus \langle S \rangle \oplus \langle C \rangle \oplus \langle L_{fT} \rangle.$$

3.1.3. Reward function

Our reward function consists of non-terminating reward r_{NT} (for intermediate steps) and terminating reward r_T (for the terminate step).

r_{NT} is designed to optimize multiple objectives, including energy consumption, time consumption, noise impact, and passenger comfort and safety.

$$r_{NT} = R_e + \alpha_1 T_{step} + \alpha_2 R_{dest} + \alpha_3 L_{fT} + \alpha_4 S + \alpha_5 C,$$

where each term is defined as:

- R_e: rewards eVTOL for minimizing energy consumption, as described in the subsequent section.
- T_{step}: penalizes the time taken per step. It rewards faster progress towards the destination, thus minimizing the time cost for passengers.
- R_{dest}: rewards eVTOL moving towards the destination. The reward is a fixed positive value if the difference in distance to the target between steps is greater than 0, and negative if it is less than 0. This encourages the eVTOL to consistently move closer to the destination.
- L_{fT}: penalizes eVTOL for generating noise, aiming to minimize the environmental impact. The noise impact is calculated by measuring the noise levels received at the uniformly distributed receivers placed in the urban environment. At each step, the average noise level from the receivers is computed, and a penalty is applied based on this average. Lower average noise levels result in higher rewards, encouraging quieter flight.
- S: penalizes eVTOL unsafe maneuvers to ensure passenger safety. If the distance to the nearest obstacle below eVTOL is below a safety threshold, a penalty is applied. The penalty is proportional to how much the distance is below the threshold, encouraging eVTOL to maintain a safe distance from obstacles.
- C: penalizes eVTOL excessive accelerations and noise that could cause discomfort. It ensures that the flight is smooth and comfortable for passengers by rewarding minimal accelerations and lower noise levels.

We terminate training when the eVTOL is ❶ out-of-bounds (exiting simulation), ❷ depleting energy, ❸ exceeding a predefined time limit, or ❹ successfully reaching the destination. For case ❹, we set $r_T = 1000$, and for all other cases, $r_T = -1000$. We determine the weight settings through multiple experiments, gradually identifying the appropriate values for $\alpha1$, $\alpha2$, …, $\alpha5$ by exploring various weight combinations and observing their impact on multi-objective optimization, including energy consumption, time, noise, safety, and passenger comfort. In multi-objective problems, each objective's importance must be reflected through careful trade-offs; for this reason, $\alpha1$ and $\alpha2$, associated with energy consumption and time efficiency, are assigned relatively higher values, while $\alpha3$, relating to noise levels, is chosen to minimize environmental impacts within mission constraints, and $\alpha4$ and $\alpha5$, corresponding to safety and passenger comfort, are similarly prioritized to enhance overall travel experience. These weights are selected based on iterative convergence of performance metrics and specific application requirements, resulting in a balanced configuration that effectively addresses all targeted objectives.

To ensure the RL agent not only reaches the destination successfully but also continually optimizes the flight path to minimize energy or time costs, we pursue reward shaping: an additional reward adjustment is introduced at the end of each successful episode. If the agent's energy or time cost in a successful episode is lower than the historical best, it earns additional rewards proportional to the cost difference. Conversely, if the cost exceeds the previous best, a penalty proportional to the excess cost is deducted. This setup encourages agents to continually seek more efficient flight paths. This reward mechanism does not apply to the first episode that successfully reaches the destination as there is no historical optimal value for comparison at that time. This dynamic reward adjustment effectively puts the focus of RL on continuous performance improvement rather than just completing the task itself.

3.1.4. Energy consumption reward function R_e

R_e includes several components to ensure the eVTOL aircraft minimizes its energy consumption, the values of $\alpha 1, \ldots, \alpha 10$ are detailed in the provided code repository.

$$R_e = \alpha_6 R_{efficiency} + \alpha_7 R_{naive} + \alpha_8 R_{best} + \alpha_9 R_{diff} + \alpha_{10} R_{terminal},$$

where each term is defined as:

- $R_{efficiency}$: rewards eVTOL to maximize the energy efficiency. Instead of simply rewarding the distance covered in each step, we define the energy efficiency as the difference in the portion of the total distance, D_{total}, to the destination traveled between the current and previous step, $d_{current}$ and d_{next}. This difference is then divided by the energy consumed during that step, E_t. Thus, the reward function rewards eVTOL to achieve greater efficiency by getting more closer to destination with less energy, promoting optimal energy usage throughout the flight.

$$R_{efficiency} = \frac{\frac{d_{current}}{D_{total}} - \frac{d_{next}}{D_{total}}}{E_t}$$

- R_{naive}: penalizes eVTOL every step according to the energy consumed.

$$R_{naive} = E_t$$

- R_{best}: penalizes eVTOL every step if the current episode's energy consumption exceeds the historically best energy consumption recorded.
- R_{diff}(Terminating Reward): penalizes eVTOL if the episode's total energy consumption exceeds the historically best energy consumption.
- $R_{terminal}$(Terminating Reward): rewards eVTOL maintaining the same energy level at the end of the episode as when it started, promoting overall energy efficiency.

3.1.5. RL algorithm

We use PPO (Schulman et al., 2017) to learn the optimal policy. The original PPO paper provided limited implementation details beyond the use of Generalized Advantage Estimation (GAE) for the advantage function calculation. The details of neural network architecture or activation function are left unspecified, allowing for customization based on the problem at hand. However, as Engstrom et al. (2020) suggest, even superficial or seemingly trivial changes in optimization methods or algorithmic tweaks can significantly impact PPO's performance. Our modification thus considers various factors: we use the tanh activation function; include LayerNorm, BatchNorm, and Dropout layers in both actor and critic networks; and adopt linearly decay learning rate. We further normalize the reward to mitigate impacts on the value function training caused by excessively large or small rewards. We record the standard deviation of a rolling discounted sum of rewards, $\sigma = std(\sum_{i=t}^{T} \gamma^{i-t} r_i)$, and normalize the current reward as r_t/σ. Table 4 shows the hyperparameters we use in experiment.

Table 4
Hyperparameters for PPO.

Parameter	Value
Batch size	65 536
Mini batch size	512
Hidden width	64
Actor learning rate	3e−4
Critic learning rate	3e−4
Gamma	0.99
Lamda	0.95
Epsilon	0.2
K epochs	10
Entropy coefficient	0.1

Further, we use original TD3 and DDPG to study the performance of our method under different RL algorithms. We also implement two more recent RL algorithms as model-based TD-MPC2 and model-free Average TD3. This allows us to examine the effectiveness of our design cross state-of-the-art model-based and model-free RL algorithms. We demonstrate the robustness of our method by conducting comparison experiments in Section 4.

3.2. Simulation data sources

The wind field simulation data we use in this paper comes from our own CFD simulation. We use an OpenStreetMap (OSM) model of city Atlanta, state GA, USA to simulate our city-scale wind field data. The simulated data details are a set of wind vectors W, where (x, y, z) represents a point in the simulation area and (u, v, w) represents the wind velocity at (x, y, z).

$$W = (x, y, z, u, v, w)$$

To simulate high-fidelity, city-scale wind fields (see Fig. 7), we use Reynolds-averaged incompressible Navier–Stokes equations (RANS) (Alfonsi, 2009) to simulate steady-state wind fields. The RANS simulations are carried out with an open-source finite-volume method (FVM) code of OpenFOAM (Weller et al., 1998). The RANS equations are defined in Eqs. (1) and (2).

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla P + \nabla \cdot (\nu \nabla \mathbf{u}) - \nabla \cdot \boldsymbol{\tau}, \tag{2}$$

where $\mathbf{u} = (u, v, w)$ is mean flow velocity, t is time, P is pressure, ρ is density, ν is kinematic viscosity. $\boldsymbol{\tau}$ is the Reynolds stress tensor and is approximated by the RANS turbulence models. The standard $k - \epsilon$ turbulence model is used along with the wall function approach. Such a combination provides a balance between performance and computational efficiency (Li and Sansalone, 2021). Detailed mathematical expressions of the $k - \epsilon$ turbulence model can be found in previous studies (Launder and Spalding, 1974).

In this project, we consider five wind speed $[4, 8, 12, 16, 20]$ m/s and for each speed we consider four wind directions $[0°, 90°, 180°, 270°]$. Depending on the wind direction, the Dirichlet boundary condition of wind speed is applied to the corresponding upstream domain boundary surface, and the Neumann boundary conditions are applied to the downstream boundary surface. No slip boundary conditions are applied to all building surfaces and ground. The free-shear boundary conditions are applied to the top boundary domain and two-side boundary surfaces. A 5% turbulence intensity is considered in the upstream boundary. The exact boundary conditions for turbulence quantities (i.e., k, ϵ) are less of a concern in this study because the flow solutions are dominated by the turbulent wakes generated by the buildings.

A Semi-Implicit Method for Pressure Linked Equations (SIMPLE) algorithm is used to solve the system of equations, i.e., Eqs. (1) and (2). A second-order upwind scheme is used for the advection terms in the mean flow and turbulence equations. For the diffusion terms

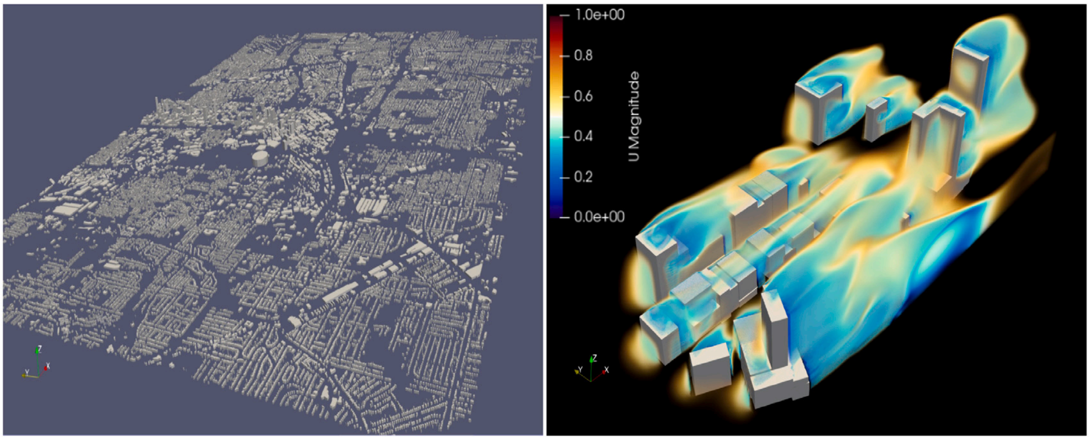


Fig. 7. City-scale wind field simulation. The left figure shows the OpenStreetMap (OSM) model of city Atlanta, state GA, USA. The simulation environment is imported into OpenFOAM (Anon, 2020a) via scripting. This allows for the visualization of the environment in Paraview (Anon, 2020b). In the right figure, one part of the result of wind field simulation is visualized as volumetric rendering of velocity field magnitude.

in the mean flow and turbulence equations, the second-order central-difference schemes are used. The simulations are considered as converged when the area-averaged turbulent kinetic energy (TKE) at the free-shear surface becomes asymptotic (i.e., relative difference < 0.1%), and the scaled residuals of all variables are below 10^{-5}.

4. Experiments and results

In this section, we first introduce our training strategies, and explain our experiment set-up. Following this, we present the overall results. Last, we present the result of our ablation study on reward function.

4.1. Evtol training strategies

We train four distinct training strategies. The first strategy prioritizes minimizing energy consumption, the second concentrates on improving passenger experience, and the third concentrates on reducing environmental impact, and the last aims to strike a balance between objectives.

4.1.1. Aircraft energy consumption

As mentioned in Section 2.3, we compare the performance of our reward function design with the two common reward function designs to minimize the energy consumption of RL agent: “Direct energy consumption” and “Energy efficiency”, and the other two reward function design “Math equation” and “Virtual energy queue”. The experiment design for energy consumption consists of ten parts, utilizing five RL algorithms: DDPG, TD3, PPO, Average TD3 and TD-MPC2. Each part focuses on comparing our reward function design with common approaches found in other research papers. The comparisons are conducted under the same RL algorithm and identical environment setup, with only the reward function design for energy consumption being modified.

- DDPG with Direct Reward Design Comparison: We use the DDPG algorithm to compare our reward function design against the direct reward function commonly used in other research.
- DDPG with Efficiency Reward Design Comparison: We use the DDPG algorithm but compares our reward function design against another common approach focusing on energy efficiency.
- PPO with Direct Reward Design Comparison: We employ the PPO algorithm to compare our reward function design with the direct reward function.
- TD3 with Efficiency Reward Design Comparison: We use the TD3 algorithm to compare our reward function design with the efficiency reward function.
- DDPG with Math Reward Design Comparison: We use the DDPG algorithm to compare our reward function design against the math equation reward function design.
- DDPG with Queue Reward Design Comparison: We use the DDPG algorithm to compare our reward function design against the virtual energy queue reward function design.
- TD-MPC2 with Direct Reward Design Comparison: We use the model-based TD-MPC2 algorithm to compare our reward function design against the direct reward function design.
- TD-MPC2 with Efficiency Reward Design Comparison: We use the model-based TD-MPC2 algorithm to compare our reward function design against the efficiency reward function design.
- Average TD3 with Direct Reward Design Comparison: We use the Average TD3 algorithm to compare our reward function design against the direct reward function design.
- Average TD3 with Efficiency Reward Design Comparison: We use the Average TD3 algorithm to compare our reward function design against the efficiency reward function design.

4.1.2. Passenger, environment concerns and all objectives

The second training strategy aims to increase the passenger safety, comfort, and to reduce the time during eVTOL flight. The third training strategy aims to reduce the noise generated by the eVTOL and received by the receivers in the urban environment. The last training strategy aims to optimize all the objectives involved in this study simultaneously: energy consumption, passenger safety, passenger comfort, time consumption and noise impact.

4.2. Experiment set-up

We set the OSM model of city Atlanta, state GA, USA to be our simulation environment. The origin and the destination of eVTOL are cross the city, 10 kilometers apart.

We report results about seven wind fields, namely D0-4, D0-8, D90-4, D180-4, D180-12, D270-16, D270-20. The wind field’s name consists of wind direction and speed, for example, D90-4 means that the angle between the wind direction and the positive X-axis is 90° and the wind speed is 4 m/s.

4.3. Computational complexity analysis

For one set of training, we utilize two NVIDIA RTX 2080Ti GPU to handle up to 10 million training steps over a period of approximately 24 h. Under these conditions, the system achieved roughly 2000 steps per second (SPS). For deployment stage in the future, the computational complexity will be reduced. Inference requires only a forward

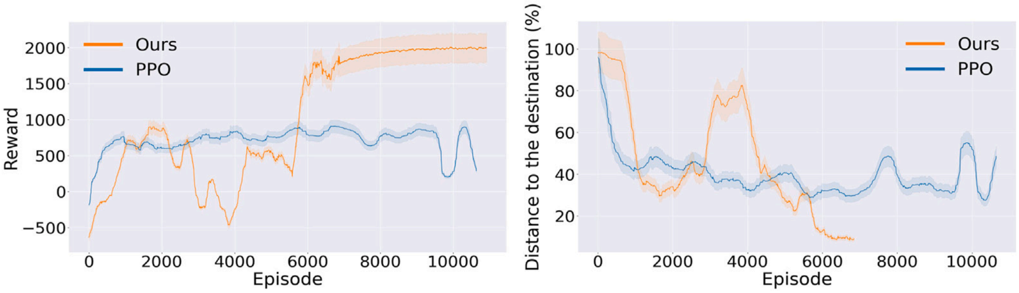


Fig. 8. Training performance comparisons. LEFT: Our method starts to outperform vanilla PPO (Schulman et al., 2017) starting around the 6000th episode. RIGHT: Our approach allows the eVTOL to reach its destination far earlier than vanilla PPO. These results demonstrate the effectiveness of our algorithm design.

pass through the trained neural network, and we plan to optimize the network architecture to run efficiently on low-power embedded devices at real-time speeds. As a result, the computational overhead at deployment is kept manageable, ensuring that the eVTOL control policy can be executed with low latency and without excessive computational demands.

4.4. Evaluation metric and results

In our RL training framework, we emphasize retaining the policy that demonstrates the best performance across critical metrics. Our evaluation involves comparing the extremum values of energy consumption, energy efficiency, noise levels, and passenger comfort against other designs. Unlike other methods that often focus on single-objective optimization, such as minimizing energy consumption or flight time, our approach distinguishes itself through a novel reward function design tailored for multi-objective optimization in eVTOL flight motion planning. This reward function incorporates broader objectives, including passenger safety, noise mitigation, and comfort, ensuring a balanced solution that addresses the social and environmental impacts of eVTOL operations. To provide a robust comparison, we conducted experiments using both model-based and model-free RL algorithms. While the RL training frameworks are consistent across methods, the primary distinction lies in our reward function's ability to generalize effectively across algorithms. Experimental results demonstrated that our method consistently outperformed others shown in the following sections. Possible disadvantages could involve additional complexity, leading to longer training times and requiring careful tuning of reward coefficients to balance competing objectives effectively. This integrated strategy highlights the effectiveness and practicality of our reward function design for advancing eVTOL flight motion planning in complex urban environments.

We first demonstrate the effectiveness of our modification on PPO algorithm as shown in Fig. 8 LEFT: we can see that our approach majorly outperforms PPO starting around the 6000th episode. Fig. 8 RIGHT shows our method can approach the destination around the 7000th episode, while PPO still has around 30% distance left to the destination around the 12000th episode.

4.4.1. Aircraft energy consumption

We compare our method with four existing approaches, demonstrating its superiority in reducing eVTOL flight energy consumption and enhancing energy efficiency as shown in Fig. 9. In subfigures (a) and (b), using the DDPG algorithm in the D0-8 wind field, our method can achieve lower energy consumption and higher energy efficiency, around episode 15000 th, than the direct, efficiency, math and queue methods. Subfigures (c) and (d) show that with the modified PPO algorithm in the D0-4 wind field, our method again excels, achieving the optimal policy around episode 14000 th. Subfigures (e) and (f) illustrate the TD3 algorithm's results in the D0-4 wind field, where our method continues to lead, with the optimal policy learned around episode 7000 th. Subfigures (g) and (h) display the TD-MPC2 algorithm's results in D270-16 wind field, comparing our method with the Direct and Efficiency method. Subfigures (i) and (j) display the Average TD3 algorithm's results in D270-20 wind field, comparing our method with the Direct and Efficiency method. For quantitative results, Table 5 shows the comparison of energy consumption and energy efficiency. Our method demonstrates overall advantage in both reducing energy consumption and enhancing energy efficiency, despite a few instances where it shows less advantage. In the DDPG algorithm with the D0-8 wind field, our method reduces energy consumption by 375.1% compared to the Direct method and by 25.1% compared to the Efficiency method, while also increasing energy efficiency by 75.22% and 17.58% respectively. Although there are cases like the TD-MPC2 algorithm with the D270-16 wind field, where our method shows a slight increase in energy consumption compared to the Direct method, the overall trend across various scenarios consistently favors our approach. On average, our method reduces energy consumption by 56.99% compared to the Direct method, by 31.56% compared to the Efficiency method, by 99.33% compared to the Math method and by 147.9% compared to the Queue method, highlighting its strong performance in minimizing energy usage. Also, our method improves energy efficiency by 15.66% compared to the Direct method and by 10.82% compared to the Efficiency method, by 36% compared to the Math method and by 42.5% compared to the Queue method, indicating that it not only saves energy but also improves the distance traveled per unit of energy consumed.

4.4.2. Passenger concerns

Fig. 10 shows the comparison of three passenger concerns between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. Fig. 10 shows that our method can learn better flight motions during RL training, and has an advantage in terms of passenger comfort, passenger safety and time consumption compared to DDPG. For our modified PPO algorithm, the comfort level steadily improves, reaching a value close to −200 around episode 8000 th. Although the DDPG algorithm can also learn better policy in the beginning, the performance decreases as training continues. Also, our modified PPO can achieve high passenger comfort comparable to the performance of recent model-based and model-free RL algorithms, TD-MPC2 and Average TD3.

4.4.3. Noise impact on urban environment

We record the average step noise as evaluation metric. For each successful episode, every step generates a noise value received by the nine POIs shown in Fig. 6. The sum of all noise values across all steps is divided by the total number of steps in the episode, yielding the average step noise. Fig. 11 shows the comparison of average step noise produced by eVTOL flight between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. Our modified PPO algorithm achieves lower noise level produced by eVTOL flight compared to DDPG and TD-MPC2 algorithm. Also, our performance is comparable to the recent model-free RL algorithm, Average TD3.

Table 5
Comparison of energy consumption and energy efficiency across five different RL algorithms and six wind fields. The first and second column lists the algorithms and corresponding wind fields. The table is divided into two main sections: energy consumption and energy efficiency. For each algorithm and wind field combination, we compare the performance of our method with four other methods: "Direct" and "Efficiency", "Math" and "Queue". The last row of the table shows the average percentage difference across all conditions for energy consumption and energy efficiency.

RL algorithm	Wind	Energy consumption (kWh)					Energy efficiency (m/kWh)				
	Field	Ours	Direct	Efficiency	Math	Queue	Ours	Direct	Efficiency	Math	Queue
DDPG	D0–8	49.71	236.2	62.2	177.51	123.34	231.37	57.29	190.69	73.69	99.5
	D90–4	99.4	52.36	435.29	118.48	78.48	146.16	240.76	33.8	108.52	182.69
	D180–12	149.34	406.95	103.75	424.07	370.88	81.51	37.19	112.0	34.36	37.23
	D270–16	98.36	131.77	112.55	144.59	234.52	126.71	93.1	105.54	87.26	55.03
	D270–20	154.92	129.06	332.08	138.41	659.2	89.77	99.10	42.11	93.28	27.43
Modified PPO	D0–4	82.65	97.47	N/A	N/A	N/A	149.47	129.1	N/A	N/A	N/A
TD3	D0–4	52.08	N/A	56.29	N/A	N/A	251.0	N/A	230.69	N/A	N/A
	D180–12	69.81	N/A	73.11	N/A	N/A	192.86	N/A	178.81	N/A	N/A
TD-MPC2	D0–8	57.44	87.72	58.75	N/A	N/A	203.84	143.61	198.9	N/A	N/A
	D90–4	60.15	62.47	60.94	N/A	N/A	197.0	189.51	198.48	N/A	N/A
	D180–12	95.88	112.45	111.07	N/A	N/A	130.61	115.83	114.46	N/A	N/A
	D270–16	69.04	86.22	83.44	N/A	N/A	173.09	144.07	147.99	N/A	N/A
	D270–20	99.67	89.95	100.51	N/A	N/A	124.77	136.37	117.78	N/A	N/A
Average TD3	D0–8	48.15	161.8	48.91	N/A	N/A	270.78	90.55	266.08	N/A	N/A
	D90–4	53.62	58.13	54.36	N/A	N/A	243.67	224.69	241.14	N/A	N/A
	D180–12	74.86	67.83	73.12	N/A	N/A	183.72	192.55	179.30	N/A	N/A
	D270–16	75.93	63.88	74.21	N/A	N/A	194.20	205.26	200.16	N/A	N/A
	D270–20	76.59	128.91	94.03	N/A	N/A	174.23	104.45	164.72	N/A	N/A
Average diff (%)			**56.99**	**31.56**	**99.33**	**147.9**		**15.66**	**10.82**	**36**	**42.5**

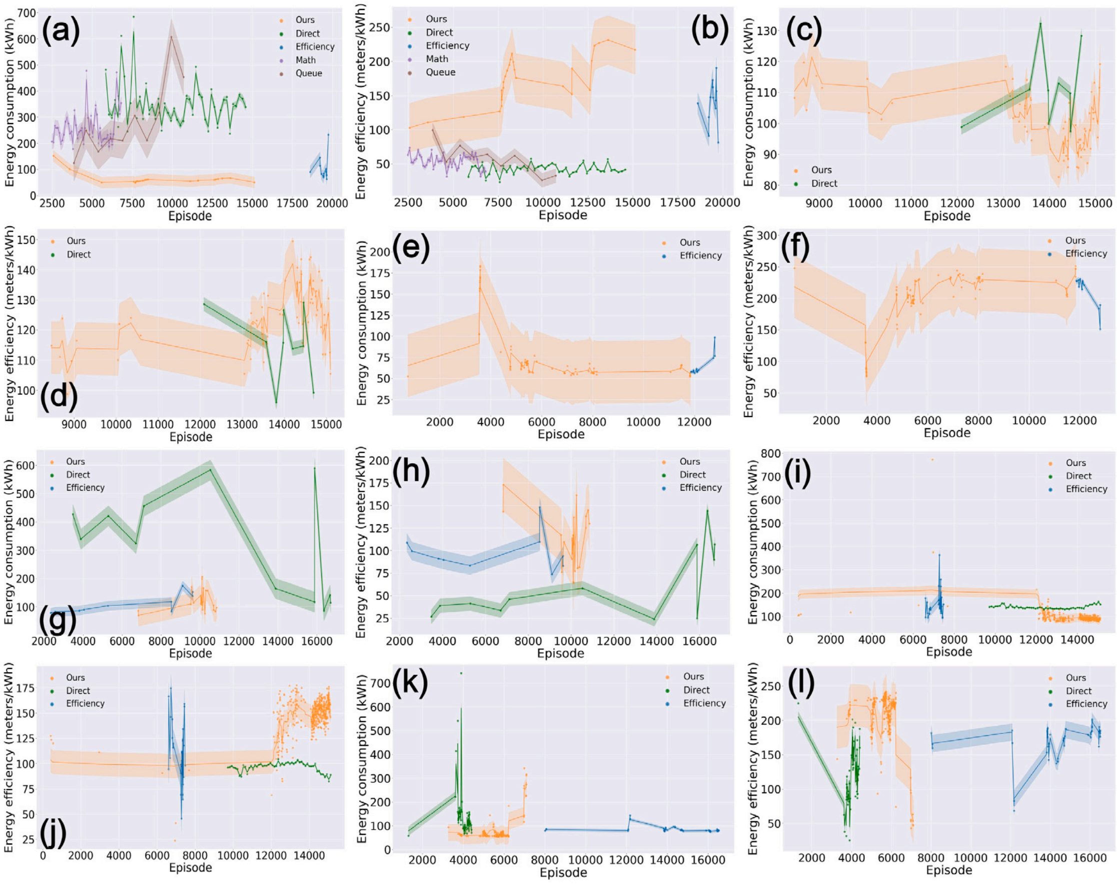


Fig. 9. Comparison of energy consumption and energy efficiency between our method and the two other common methods. The scattered data points are shown alongside smoothed curves to highlight trends across RL training episodes. (a, c, e, g, i, k) illustrate flight energy consumption in kWh, and (b, d, f, h, j, l) depict energy efficiency in meters/kWh. The rows represent different algorithms and comparison scenarios: (a, b) show results from DDPG algorithm in D0-8 wind field, comparing our method with the Direct, Efficiency, Math and Queue methods; (c, d) present results from our modified PPO algorithm in D0-4 wind field, comparing our method with the Direct method; and (e, f) display the TD3 algorithm's results in D0-4 wind field, comparing our method with the Efficiency method; (g,h) display the TD-MPC2 algorithm's results in D270-16 wind field, comparing our method with the Direct and Efficiency method; (i,j) display the Average TD3 algorithm's results in D270-20 wind field, comparing our method with the Direct and Efficiency method. (k,l) display the Average TD3 algorithm's results in D90-4 wind field, comparing our method with the Direct and Efficiency method.

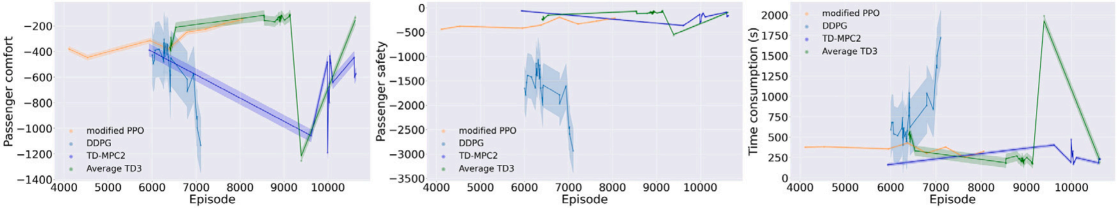


Fig. 10. Comparison of three passenger concerns between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. LEFT: The subplot displays the trend of total passenger comfort in one task period across training episodes. It shows our modified PPO can achieve high passenger comfort comparable to the performance of recent model-based and model-free RL algorithms, while DDPG algorithm's performance going down. MIDDLE: This subplot shows our modified PPO can achieve high passenger safety comparable to the performance of recent model-based and model-free RL algorithms. RIGHT: The subplot shows our modified PPO can achieve low time consumption comparable to the performance of recent model-based and model-free RL algorithms.

Table 6
Comparison of the performance for all three perspectives between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. It compares the four methods across six metrics: energy consumption, energy efficiency, passenger comfort, passenger safety, time consumption, and noise impact. The last row shows the percentage difference between our modified PPO and DDPG algorithm for each metric. Our modified PPO has comparable performance with the latest model-based and model-free RL algorithms..

Method	Energy consumption (kWh)	Energy efficiency (m/kWh)	Comfort	Safety	Time (s)	Noise (dB)
Ours-all	57.83	206.59	−401.59	−477.55	162	20.55
DDPG-all	123.54	99.72	−778.11	−917.69	348	25.01
TD-MPC2-all	76.58	156.56	−435.38	−382.53	160	21.59
Average TD3-all	67.52	172.35	−612.21	−682.55	270	21.39
Diff between ours and DDPG (%)	53.19	51.73	48.39	47.96	53.45	17.83

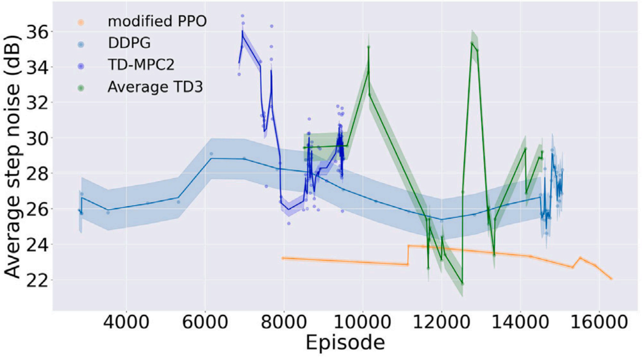


Fig. 11. Comparison of average step noise produced by eVTOL flight between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. Our modified PPO algorithm achieves lower noise level produced by eVTOL flight compared to DDPG and TD-MPC2 algorithm. Also, our performance is comparable to the recent model-free RL algorithm, Average TD3.

4.4.4. *Considering all objectives in one training process*

Fig. 12 shows the results from a single RL training process where all relevant factors (energy consumption, passenger concerns, and noise impact) are simultaneously considered, between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. Each subfigure represents a different objective. The results demonstrate that our RL agent could learn better policy that improves all the objectives concurrently.

For quantitative results, Table 6 shows that our method using the modified PPO, "Ours-all", consistently outperforms our method using DDPG, "DDPG-all", across all six metrics, with notable percentage differences highlighting the advantages. Also, our performance is comparable to the recent model-based and model-free RL algorithm, TD-MPC2 and Average TD3.

Based on Fig. 12 and Table 6, our results demonstrate that all RL algorithms involved in the experiment can learn to optimize all objectives simultaneously. There is performance gap, with "Ours-all" exhibiting superior learning outcomes across the board, compared to DDPG algorithm. Despite these differences, both methods show the ability to balance and improve upon the multiple metrics considered in the training process.

4.5. *Ablation study*

In this section, we perform a reward function ablation study to evaluate its contribution to task performance. We systematically remove certain reward terms and observe the results. The results demonstrate how each component contributes to balancing multiple objectives and improving the eVTOL flight performance.

4.5.1. *Experimental design*

We employ a single component removal approach, where we eliminate each key component of the reward function individually and observe the performance. The complete reward function serves as the baseline model for comparison.

- **Baseline**: uses the full reward function.
- **No Out-of-Bound Punish**: removes the out-of-bound punishment to analyze its importance in boundary control.
- **No Energy Optimization**: removes the energy optimization component to observe changes in energy consumption.
- **No Time Optimization**: removes the time optimization component to evaluate its impact on time efficiency.
- **No Passenger Comfort Optimization**: removes the passenger comfort component to assess its contribution to safety.
- **No Passenger Safety Optimization**: removes the passenger safety component to detect its effect on comfort optimization.
- **No Noise Optimization**: removes the noise control component to evaluate noise reduction.

4.5.2. *Results and analysis*

The experimental results are presented in Table 7 and Fig. 13. "Success count" means the number of successfully reaching destination in 50 k training episode. "Out-of-bound termination count" means the number of episode terminating because of eVTOL flying out of the boundary of the simulation environment. We analyze the results in the following items.

- **No Out-of-Bound Punish:** significantly impacts eVTOL flying out of the simulation environment boundary termination reduction, as removing it leads to an increase in Out-of-Bound termination count in Table 7 and no successfully reaching destination in 50k training episode.

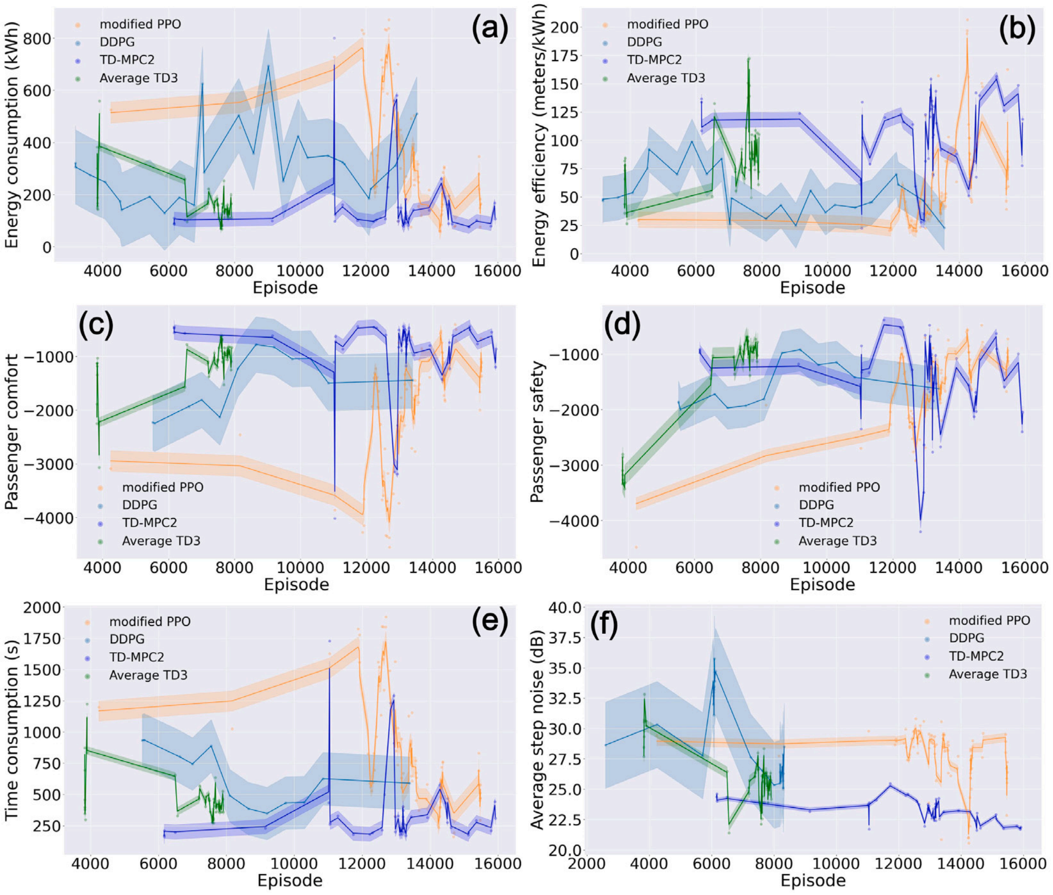


Fig. 12. Comparison of the performance for all three perspectives between our modified PPO, DDPG, TD-MPC2 and Average TD3 algorithm. Each subfigure represents a specific objective: (a) energy consumption, (b) energy efficiency, (c) passenger comfort, (d) passenger safety, (e) time consumption, and (f) average step noise.

Table 7
Results of reward function ablation study. "Success count" means the number of successfully reaching destination in 50 k training episode. "Out-of-bound termination count" means the number of episode terminating because of eVTOL flying out of the boundary of the simulation environment.

Experiment	Success count	Out-of-bound termination count	Timeout termination count	Energy consumption (kWh)	Passenger comfort	Passenger safety	Noise level (dB)
Baseline	3455	46 500	45	48.28	−416.45	−257.55	17.99
No out-of-bound punish	0	49 936	64	N/A	N/A	N/A	N/A
No energy optimization	3799	46 170	31	61.64	−436.75	−303.76	18.34
No time optimization	339	49 422	239	52.78	−420.06	−343.1	17.52
No passenger comfort optimization	2987	46 957	56	57.99	−452.93	−298.15	18.76
No passenger safety optimization	3782	46 175	43	49.73	−404.26	−424.70	16.3
No noise optimization	3617	46 305	78	50.18	−427.6	−417.9	20.70

- **No Energy Optimization:** increases eVTOL energy consumption from the baseline's 48.28 kWh to 61.64 kWh. It also leads to more success counts because the energy punishment during training is lessened compared to the baseline. eVTOL finds more successful flying paths with higher energy consumption.
- **No Time Optimization:** achieves 339 successes, much lower than the baseline with more timeout terminations.
- **No Passenger Comfort Optimization:** decreases the eVTOL passenger comfort during flight from −416.45 to −452.93.
- **No Passenger Safety Optimization:** decreases the eVTOL passenger safety during flight from −257.55 to −424.70.
- **No Noise Optimization:** increases average noise levels generated by eVTOL each training step from 17.99 dB to 20.70 dB.

5. Discussion and limitations

One focus of our work is on energy optimization, where we introduce a RL reward function design structure for energy consumption and energy efficiency. Compared to commonly used methods, our approach demonstrates clear advantages. To validate our method, we conduct comparative experiments in identical simulation environments against other reward functions, with results showing that our method can learn better RL policies for reducing energy consumption and enhancing energy efficiency. For the other two perspectives, our contribution lies in the integration of passenger concerns and environmental noise impact as evaluation metrics within the RL reward function. The results show that our RL method can improve these objectives, showing its

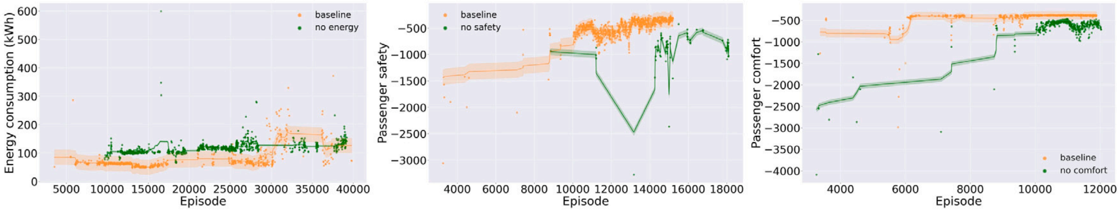


Fig. 13. Experimental results for reward function ablation study. We systematically remove certain reward terms and observe the results. The results demonstrate how each component contributes to balancing multiple objectives and improving the eVTOL flight performance. The left figure shows the comparison between baseline and no consumed energy punishment experiment. The middle figure shows the comparison between baseline and no passenger safety punishment experiment. The left figure shows the comparison between baseline and no passenger comfort punishment experiment.

potential in dealing with multi-objective problem. Among reward functions commonly used by others, the direct design generally outperforms the efficiency design in learning better RL policies.

There are limitations associated with our method. First, one notable drawback is the slower learning process resulting from a more complex, multi-objective reward function. Incorporating additional factors – such as energy consumption, timing, safety, comfort, and noise levels – broadens the agent's optimization goals. However, this complexity also necessitates more training episodes, longer training times, and the use of larger datasets, which can be problematic in real-time RL scenarios where time efficiency is crucial. To address these challenges, we have employed strategies such as an experience replay buffer, and early stopping criteria, to reduce computational overhead. We will investigate more efficient training techniques – such as transfer learning and curriculum learning – to improve sample efficiency and shorten training durations, all while preserving the comprehensiveness and robustness of the learned policies.

Second, the performance limitations may also arise from strict penalty rewards that curb the agent's exploration. These penalties can lead to slower routes with higher time costs. Although the agent eventually reduces energy consumption, it fails to match methods that initially explore faster and more efficient paths. To improve this, we could expand the simulation environment, allowing higher velocities without causing out-of-bound errors. Curriculum learning can also help by starting with simpler tasks and gradually increasing complexity. Additionally, easing penalties during early training might encourage the agent to explore more diverse options. We plan to investigate more flexible reward structures – such as hierarchical RL or human-in-the-loop feedback – to strike a better balance between effective exploration and strict safety requirements.

Third, our method currently rely on training under a single wind field condition. Although the results presented in this paper demonstrate its effectiveness in controlled settings, this approach may limit the policy's generalization capability when faced with the diverse and unpredictable wind scenarios encountered in real-world applications. To address this, we plan to adopt curriculum learning, starting with a single wind field condition and progressively introducing additional wind fields with increasing complexity during training. This multi wind field training strategy has the advantage of allowing the agent to develop more generalized and robust policies, making it better equipped to handle varying urban wind conditions. However, it also introduces challenges, such as increased training complexity, longer training times, and greater difficulty in achieving convergence to an optimal policy compared to single wind field training. Despite these challenges, we believe that curriculum learning is essential to enhance the robustness and adaptability of our method, ensuring reliable eVTOL performance in diverse urban environments.

Fourth, although model-free RL works well in complex, changing environments, it comes with drawbacks. It often needs many samples, making training expensive and time-consuming—especially in eVTOL scenarios where each simulation or test is costly. Model-free RL also lacks foresight, relying on immediate feedback rather than planning multiple steps ahead. This short-term focus may hinder its ability to achieve long-term goals, such as reducing energy use over long flights. A potential solution is to combine model-free and model-based RL into a hybrid approach. In predictable conditions, using a model-based method can improve efficiency and speed up learning. In less predictable conditions, such as fast-changing wind, the system can switch back to model-free RL to avoid errors from poor models. By blending these methods, we leverage their strengths, improving adaptability, efficiency, and overall performance in eVTOL flight control.

Our method faces challenges in real-world eVTOL scenarios, such as navigating through urban traffic congestion and performing emergency rescue missions. Training the multi-objective policy in simulation requires significant computational resources due to complex rewards and extensive exploration. However, once trained, the policy is lightweight and practical for deployment on embedded hardware. In urban congestion scenarios, the policy must efficiently optimize flight paths to minimize delays while adhering to noise and safety constraints. In emergency rescues, it must prioritize rapid response and safe navigation through dynamic environments. To meet such real-time demands, we refine the policy using techniques like pruning, quantization, and compression, reducing latency and enabling smooth operation on onboard processors. Hardware accelerators, such as low-power GPUs or inference units, further enhance speed. We also explore strategies like hierarchical policies for task prioritization and knowledge distillation to streamline optimization. By tailoring the method to these specific applications, we ensure adaptability and robustness across diverse eVTOL scenarios, addressing real-world constraints effectively.

We claim that our RL task is a long-horizon problem due to the flight distance (10 kilometers) involved. Unlike larger aircraft, current commercial manned eVTOL vehicles have a lower velocity range (60 m/s), requiring a large number of action steps to complete the task successfully. This increases the learning difficulty and reduces the exploration success rate, which is reflected in the sparsity of data points in the results.

Our results for the "ours-all" method demonstrate that traditional approaches, such as linear combination or scalarization of reward functions, can be sufficiently powerful to optimize multiple objectives simultaneously. This insight aligns with existing aircraft-related work and underscores the robustness of our approach. Despite managing six different objectives within our mixed code, the outcomes remain promising, validating the potential of our method in complex, multi-objective optimization scenarios.

6. Conclusion and future work

We introduce a DRL-based method for eVTOL flight 3D motion planning in urban environments with wind fields, emphasizing the optimization of energy conservation, passenger concerns, and noise impact on the urban environment. Our approach significantly tailors the PPO algorithm (Schulman et al., 2017) to learn better policies that balance these critical factors. We design a reward structure specifically aimed at reducing energy consumption and enhancing energy efficiency, which has proven to outperform commonly used design methods by conducting comparative experiments. Additionally, we integrate passenger comfort and safety, as well as noise impact, into the RL reward function. The results demonstrate that our method can learn better policy to improve these objectives simultaneously, offering a comprehensive solution for the multi-objective problem in eVTOL flight planning. Overall, our approach has shown to be effective and efficient in improving energy conservation, passenger experience, and minimizing noise impact on the urban environment during eVTOL flight motion planning.

There are many future research directions. First, we plan to extend our technique to multi-eVTOL systems to study collaborative flight optimization in urban wind fields. Second, various origin–destination demand patterns (as a result of different flight tasks) of eVTOL will be investigated. Lastly, we aim to explore the joint study of air mobility and ground mobility in large-scale, mixed traffic settings (Li et al., 2017; Wang et al., 2023b, 2024; Villarreal et al., 2024).

CRediT authorship contribution statement

Songyang Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Weizi Li:** Conceptualization. **Haochen Li:** Data curation. **Shuai Li:** Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the National Science Foundation (NSF), USA via Grant 2129003 and 2231710. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF and University of Florida.

Data availability

Data will be made available on request.

References

Abdolmaleki, A., Huang, S., Hasenclever, L., Neunert, M., Song, F., Zambelli, M., Martins, M., Heess, N., Hadsell, R., Riedmiller, M., 2020. A distributional view on multi-objective policy optimization. In: International Conference on Machine Learning. PMLR, pp. 11–22.

Abedin, S.F., Munir, M.S., Tran, N.H., Han, Z., Hong, C.S., 2020. Data freshness and energy-efficient UAV navigation optimization: A deep reinforcement learning approach. IEEE Trans. Intell. Transp. Syst. 22 (9), 5994–6006.

Akbari, M., Asadi, P., Besharati Givi, M., Khodabandehlouie, G., 2014. Artificial neural network and optimization. Adv. Friction- Stir Weld. Process. 2, 543–599.

Al-Habob, A.A., Dobre, O.A., Muhaidat, S., Poor, H.V., 2021. Energy-efficient data dissemination using a UAV: An ant colony approach. IEEE Wirel. Commun. Lett. 10, 16–20. http://dx.doi.org/10.1109/LWC.2020.3019001.

Alfonsi, G., 2009. Reynolds-averaged Navier–Stokes equations for turbulence modeling.

Alqahtani, H., Kumar, G., 2024. Machine learning for enhancing transportation security: A comprehensive analysis of electric and flying vehicle systems. Eng. Appl. Artif. Intell. 129, 107667.

An, G., Wu, Z., Shen, Z., Shang, K., Ishibuchi, H., 2023. Evolutionary multi-objective deep reinforcement learning for autonomous UAV navigation in large-scale complex environments. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 633–641.

Anon, 2020a. Openfoam. https://www.openfoam.com/, Accessed: 2023-08-05.

Anon, 2020b. Paraview. https://www.paraview.org/, Accessed: 2023-09-05.

Anon, 2023. Guardian agriculture's aircraft becomes first eVTOL authorized to operate in the U.S.. https://verticalmag.com/press-releases/guardian-agricultures-aircraft-becomes-first-evtol-authorized-to-operate-in-the-u-s/, Accessed: 2023-09-30.

Anon, 2024. Joby completes third stage of FAA certification process. https://verticalmag.com/press-releases/joby-completes-third-stage-of-faa-certification-process/, Accessed: 2024-05-30.

Arani, A.H., Azari, M.M., Hu, P., Zhu, Y., Yanikomeroglu, H., Safavi-Naeini, S., 2021. Reinforcement learning for energy-efficient trajectory design of UAVs. IEEE Int. Things J. 9 (11), 9060–9070.

Babu, N., Donevski, I., Valcarce, A., Popovski, P., Nielsen, J.J., Papadias, C.B., 2022. Fairness-based energy-efficient 3-D path planning of a portable access point: A deep reinforcement learning approach. IEEE Open J. Commun. Soc. 3, 1487–1500. http://dx.doi.org/10.1109/OJCOMS.2022.3201292.

Bahabry, A., Wan, X., Ghazzai, H., Menouar, H., Vesonder, G., Massoud, Y., 2019. Low-altitude navigation for multi-rotor drones in urban areas. IEEE Access 7, 87716–87731. http://dx.doi.org/10.1109/ACCESS.2019.2925531.

Bai, X., Jiang, H., Cui, J., Lu, K., Chen, P., Zhang, M., 2021. UAV path planning based on improved a <math id="m1"> <mo>*</mo> </math> and DWA algorithms. Int. J. Aerosp. Eng. 2021, 1–12. http://dx.doi.org/10.1155/2021/4511252.

Bhalla, S., Kim, D., Choi, D., 2024. Enhancing human comfort in eVTOL aircraft assisted by control moment gyroscopes. Int. J. Aeronaut. Space Sci. 1–21.

Chen, Q., He, Q., Zhang, D., 2023. UAV path planning based on an improved chimp optimization algorithm. Axioms 12 (7), http://dx.doi.org/10.3390/axioms12070702, URL https://www.mdpi.com/2075-1680/12/7/702.

Chen, D., Qi, Q., Zhuang, Z., Wang, J., Liao, J., Han, Z., 2020. Mean field deep reinforcement learning for fair and efficient UAV control. IEEE Int. Things J. 8 (2), 813–828.

Chodnicki, M., Siemiatkowska, B., Stecz, W., Stępień, S., 2022. Energy efficient UAV flight control method in an environment with obstacles and gusts of wind. Energies 15, 3730. http://dx.doi.org/10.3390/en15103730.

Dai, Z., Liu, C.H., Han, R., Wang, G., Leung, K.K., Tang, J., 2021. Delay-sensitive energy-efficient UAV crowdsensing by deep reinforcement learning. IEEE Trans. Mob. Comput. 22 (4), 2038–2052.

Do, Q.V., Pham, Q.-V., Hwang, W.-J., 2021. Deep reinforcement learning for energy-efficient federated learning in UAV-enabled wireless powered networks. IEEE Commun. Lett. 26 (1), 99–103.

Edwards, T., Price, G., 2020. eVTOL passenger acceptance. Tech. rep..

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., Madry, A., 2020. Implementation matters in deep policy gradients: A case study on PPO and TRPO. arXiv:2005.12729.

Forkan, M., Rizvi, M.M., Chowdhury, M.A.M., 2022. Optimal path planning of unmanned aerial vehicles (UAVs) for targets touring: Geometric and arc parameterization approaches. PLOS ONE 17 (10), 1–20. http://dx.doi.org/10.1371/journal.pone.0276105.

Fu, S., Tang, Y., Wu, Y., Zhang, N., Gu, H., Chen, C., Liu, M., 2021. Energy-efficient UAV-enabled data collection via wireless charging: A reinforcement learning approach. IEEE Int. Things J. 8 (12), 10209–10219.

Gunantara, N., 2018. A review of multi-objective optimization: Methods and its applications. Cogent Eng. 5 (1), 1502242.

Guo, K., Wu, M., Li, X., Song, H., Kumar, N., 2023. Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: Trajectory optimization and beamforming design. IEEE Trans. Intell. Transp. Syst. 24 (9), 10197–10210.

Hansen, N., Su, H., Wang, X., 2023. Td-mpc2: Scalable, robust world models for continuous control. arXiv preprint arXiv:2310.16828.

Hayes, C.F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., et al., 2022. A practical guide to multi-objective reinforcement learning and planning. Auton. Agents Multi- Agent Syst. 36 (1), 26.

Holden, J., Goel, N., 2016. Fast-forwarding to a future of on-demand urban air transportation.

Holmes, B., Parker, R., Stanley, D., McHugh, P., Garrow, L., Masson, P., Olcott, J., 2017. NASA strategic framework for on-demand air mobility. NASA Contractor Report NNL13AA08B, National Institute of Aerospace, Hampton, VA.

Hong, D., Lee, S., Cho, Y.H., Baek, D., Kim, J., Chang, N., 2021a. Energy-efficient online path planning of multiple drones using reinforcement learning. IEEE Trans. Veh. Technol. 70, 9725–9740. http://dx.doi.org/10.1109/TVT.2021.3102589.

Hong, D., Lee, S., Cho, Y.H., Baek, D., Kim, J., Chang, N., 2021b. Energy-efficient online path planning of multiple drones using reinforcement learning. IEEE Trans. Veh. Technol. 70 (10), 9725–9740.

Hu, S., Yuan, X., Ni, W., Wang, X., Jamalipour, A., 2023. RIS-assisted jamming rejection and path planning for UAV-Borne IoT platform: A new deep reinforcement learning framework. arXiv:2302.04994.

Huan, L., Ning, Z., Qiang, L., 2021. UAV path planning based on an improved ant colony algorithm. In: 2021 4th International Conference on Intelligent Autonomous Systems (ICoIAS). pp. 357–360. http://dx.doi.org/10.1109/ICoIAS53694.2021.00070.

International Organization for Standardization, 1996. Acoustics-attenuation of sound during propagation outdoors: Part 2: General method of calculation.

Jackson, P., Bardell, N., 2023. Some noise considerations for eVTOL traffic in Auckland City, New Zealand. In: Australasian Transport Research Forum (ATRF), 44th, 2023, Perth, Western Australia, Australia.

Kaufmann, E., Bauersfeld, L., Loquercio, A., Müller, M., Koltun, V., Scaramuzza, D., 2023. Champion-level drone racing using deep reinforcement learning. Nature 620 (7976), 982–987.

Kelsey, E., 2023. The future of mobility: Urban air. https://www.arup.com/perspectives/the-future-of-mobility-urban-air#, Accessed: 2023-08-30.

Kleinbekman, I.C., Mitici, M.A., Wei, P., 2018. eVTOL arrival sequencing and scheduling for on-demand urban air mobility. In: 2018 IEEE/AIAA 37th Digital Avionics Systems Conference. DASC, pp. 1–7. http://dx.doi.org/10.1109/DASC.2018.8569645.

Launder, B., Spalding, D., 1974. The numerical computation of turbulent flows. Comput. Methods Appl. Mech. Engrg. 3 (2), 269–289. http://dx.doi.org/10.1016/0045-7825(74)90029-2, arXiv:1204.1280v1.

Li, X., Li, J., Liu, D., 2021. Energy-efficient UAV trajectory design with information freshness constraint via deep reinforcement learning. Mob. Inf. Syst. 2021 (1), 1430512.

Li, Y., Liu, M., 2022. Path planning of electric VTOL UAV considering minimum energy consumption in urban areas. Sustainability 14, 13421. http://dx.doi.org/10.3390/su142013421.

Li, H., Sansalone, J., 2021. Benchmarking Reynolds-averaged Navier–Stokes turbulence models for water clarification systems. J. Environ. Eng. 147 (9), 04021031. http://dx.doi.org/10.1061/(asce)ee.1943-7870.0001889, URL https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29EE.1943-7870.0001889.

Li, W., Wolinski, D., Lin, M.C., 2017. City-scale traffic animation using statistical learning and metamodel-based optimization. ACM Trans. Graph. 36 (6), 200:1–200:12. http://dx.doi.org/10.1145/3130800.3130847.

Liu, X., Chai, Z.-Y., Li, Y.-L., Cheng, Y.-Y., Zeng, Y., 2023. Multi-objective deep reinforcement learning for computation offloading in UAV-assisted multi-access edge computing. Inform. Sci. 642, 119154.

Liu, C.H., Chen, Z., Tang, J., Xu, J., Piao, C., 2018. Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach. IEEE J. Sel. Areas Commun. 36 (9), 2059–2070.

Liu, X., Li, Y., Xie, Z., 2021. Path planning of UAV based on error correction. In: Proceedings of the 2021 13th International Conference on Machine Learning and Computing. ICMLC '21, Association for Computing Machinery, New York, NY, USA, pp. 392–396. http://dx.doi.org/10.1145/3457682.3457742.

Liu, C.H., Ma, X., Gao, X., Tang, J., 2019. Distributed energy-efficient multi-UAV navigation for long-term communication coverage by deep reinforcement learning. IEEE Trans. Mob. Comput. 19 (6), 1274–1285.

Liu, Z., Sengupta, R., Kurzhanskiy, A., 2017. A power consumption model for multi-rotor small unmanned aircraft systems. In: 2017 International Conference on Unmanned Aircraft Systems. ICUAS, IEEE, pp. 310–315.

Liu, X., Zhou, L., Zhang, X., Tan, X., Wei, J., 2022. A 3D REM-guided UAV path planning method under communication connectivity constraints. Wirel. Commun. Mob. Comput. 2022, 1–11. http://dx.doi.org/10.1155/2022/7410708.

Luna, M.A., Isaac, M.S.A., Ragab, A.R., Campoy, P., Peña, P.F., Molina, M., 2022. Fast multi-UAV path planning for optimal area coverage in aerial sensing applications. Sensors 22, 2297. http://dx.doi.org/10.3390/s22062297.

Luo, X., Wang, Q., Gong, H., Tang, C., 2024. UAV path planning based on the average TD3 algorithm with prioritized experience replay. IEEE Access.

Maciel-Pearson, B.G., Marchegiani, L., Akcay, S., Atapour-Abarghouei, A., Garforth, J., Breckon, T.P., 2019. Online deep reinforcement learning for autonomous UAV navigation and exploration of outdoor environments. arXiv preprint arXiv:1912.05684.

Murata, T., Ishibuchi, H., Tanaka, H., 1996. Multi-objective genetic algorithm and its applications to flowshop scheduling. Comput. Ind. Eng. 30 (4), 957–968.

Nagashima, T., Ding, M., Fujii, K., Takeda, K., 2022. Optimization of aircraft flight paths considering the conflicting parameters of economy and safety.

National Academies of Sciences and Division on Engineering and Physical Sciences and Aeronautics and Space Engineering Board and Committee on Enhancing Air Mobility A National Blueprint, 2020. Advancing Aerial Mobility: A National Blueprint. National Academies Press.

Nie, Y., Zhao, J., Liu, J., Jiang, J., Ding, R., 2020. Energy-efficient UAV trajectory design for backscatter communication: A deep reinforcement learning approach. China Commun. 17 (10), 129–141.

Olivares, D., Fournier, P., Vasishta, P., Marzat, J., 2024. Model-free versus model-based reinforcement learning for fixed-wing uav attitude control under varying wind conditions. arXiv preprint arXiv:2409.17896.

Omoniwa, B., Galkin, B., Dusparic, I., 2022. Optimizing energy efficiency in UAV-assisted networks using deep reinforcement learning. IEEE Wirel. Commun. Lett. 11 (8), 1590–1594.

Pradeep, P., Wei, P., 2018. Energy efficient arrival with RTA constraint for urban eVTOL operations. In: 2018 AIAA Aerospace Sciences Meeting. p. 2008.

Qi, H., Hu, Z., Huang, H., Wen, X., Lu, Z., 2020. Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach. IEEE Access 8, 53172–53184.

Qi, X., Luo, Y., Wu, G., Boriboonsomsin, K., Barth, M., 2019. Deep reinforcement learning enabled self-learning control for energy efficient driving. Transp. Res. Part C: Emerg. Technol. 99, 67–81.

Qiu, X., Gao, C., Wang, K., Jing, W., 2022. Attitude control of a moving mass–actuated UAV based on deep reinforcement learning. J. Aerosp. Eng. 35, http://dx.doi.org/10.1061/(ASCE)AS.1943-5525.0001381.

Ramezani, M., Habibi, H., luis Sanchez Lopez, J., Voos, H., 2023. UAV path planning employing MPC- reinforcement learning method considering collision avoidance. arXiv:2302.10669.

Rizzi, S.A., Huff, D.L., Boyd, D.D., Bent, P., Henderson, B.S., Pascioni, K.A., Sargent, D.C., Josephson, D.L., Marsan, M., He, H.B., et al., 2020. Urban air mobility noise: Current practice, gaps, and recommendations. Tech. rep..

Sandino, J., Galvez-Serna, J., Mandel, N., Vanegas, F., Gonzalez, F., 2022. Autonomous mapping of desiccation cracks via a probabilistic-based motion planner onboard UAVs. In: 2022 IEEE Aerospace Conference. AERO, pp. 1–14. http://dx.doi.org/10.1109/AERO53065.2022.9843299.

Schäffer, B., Pieren, R., Heutschi, K., Wunderli, J.M., Becker, S., 2021. Drone noise emission characteristics and noise effects on humans—a systematic review. Int. J. Environ. Res. Public Heal. 18 (11), 5940.

Schmähl, M., Nagy, B., Hornung, M., 2021. Comparison of a fully empiric and a semi-empiric noise modeling approach for a cargo-UAV. In: Deutscher Luft-Und Raumfahrtkongress 2021.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv:1707.06347.

Song, F., Deng, M., Xing, H., Liu, Y., Ye, F., Xiao, Z., 2024. Energy-efficient trajectory optimization with wireless charging in UAV-assisted MEC based on multi-objective reinforcement learning. IEEE Trans. Mob. Comput..

Song, F., Xing, H., Wang, X., Luo, S., Dai, P., Xiao, Z., Zhao, B., 2022. Evolutionary multi-objective reinforcement learning based trajectory control and task offloading in UAV-assisted mobile edge computing. IEEE Trans. Mob. Comput. 22 (12), 7387–7405.

Song, F., Xing, H., Wang, X., Luo, S., Dai, P., Xiao, Z., Zhao, B., 2023. Evolutionary multi-objective reinforcement learning based trajectory control and task offloading in UAV-assisted mobile edge computing. IEEE Trans. Mob. Comput. 22 (12), 7387–7405. http://dx.doi.org/10.1109/TMC.2022.3208457.

Su, G., Cheng, X., Feng, S., Liu, K., Song, J., Chen, J., Zhu, C., Lin, H., 2024. Flight path optimization with optimal control method. arXiv preprint arXiv:2405.08306.

Tu, G.-T., Juang, J.-G., 2023. UAV path planning and obstacle avoidance based on reinforcement learning in 3D environments. In: Actuators. Vol. 12, (2), MDPI, p. 57.

Villarreal, M., Wang, D., Pan, J., Li, W., 2024. Analyzing emissions and energy efficiency in mixed traffic control at unsignalized intersections. In: IEEE Forum for Innovative Sustainable Transportation Systems. FISTS.

Wan, Y., Zhong, Y., Ma, A., Zhang, L., 2023. An accurate UAV 3-D path planning method for disaster emergency response based on an improved multiobjective swarm intelligence algorithm. IEEE Trans. Cybern. 53, 2658–2671. http://dx.doi.org/10.1109/TCYB.2022.3170580.

Wang, Y., Chu, Z., Hu, Y., 2023a. Path planning of unmanned underwater vehicles based on deep reinforcement learning algorithm. In: 2023 International Conference on Advanced Robotics and Mechatronics. ICARM, IEEE, pp. 250–254.

Wang, X., Gursoy, M.C., Erpek, T., Sagduyu, Y.E., 2022. Learning-based UAV path planning for data collection with integrated collision avoidance. IEEE Int. Things J. 9, 16663–16676. http://dx.doi.org/10.1109/JIOT.2022.3153585.

Wang, D., Li, W., Pan, J., 2024. Large-scale mixed traffic control using dynamic vehicle routing and privacy-preserving crowdsourcing. IEEE Int. Things J. 11 (2), 1981–1989. http://dx.doi.org/10.1109/JIOT.2023.3335292.

Wang, D., Li, W., Zhu, L., Pan, J., 2023b. Learning to control and coordinate mixed traffic through robot vehicles at complex and unsignalized intersections.

Ware, J., Roy, N., 2016. An analysis of wind field estimation and exploitation for quadrotor flight in the urban canopy layer. In: 2016 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 1507–1514.

Weller, H.G., Tabor, G., Jasak, H., Fureby, C., 1998. A tensorial approach to computational continuum mechanics using object-oriented techniques. Comput. Phys. 12 (6), 620. http://dx.doi.org/10.1063/1.168744, URL http://scitation.aip.org/content/aip/journal/cip/12/6/10.1063/1.168744.

Wu, F., Yang, D., Xiao, L., Cuthbert, L., 2019. Energy consumption and completion time tradeoff in rotary-wing UAV enabled WPCN. IEEE Access 7, 79617–79635. http://dx.doi.org/10.1109/ACCESS.2019.2922651.

Xu, Y., Li, J., Wu, B., Wu, J., Deng, H., Hui, D., 2024. Cooperative landing on mobile platform for multiple unmanned aerial vehicles via reinforcement learning. J. Aerosp. Eng. 37, http://dx.doi.org/10.1061/JAEEEZ.ASENG-5053.

Xu, Z., Wang, Q., Kong, F., Yu, H., Gao, S., Pan, D., 2022. Ga-DQN: A gravity-aware DQN based UAV path planning algorithm. In: 2022 IEEE International Conference on Unmanned Systems. ICUS, IEEE, pp. 1215–1220.

Yang, Q., Liu, J., Li, L., 2020. Path planning of UAVs under dynamic environment based on a hierarchical recursive multiagent genetic algorithm. In: 2020 IEEE Congress on Evolutionary Computation. CEC, IEEE, pp. 1–8.

Yao, P., Wang, H., Liu, C., 2014. 3-D dynamic path planning for UAV based on interfered fluid flow. In: Proceedings of 2014 IEEE Chinese Guidance, Navigation and Control Conference. IEEE, pp. 997–1002.

Ye, H.-T., Kang, X., Joung, J., Liang, Y.-C., 2020. Optimization for full-duplex rotary-wing UAV-enabled wireless-powered IoT networks. IEEE Trans. Wirel. Commun. 19 (7), 5057–5072. http://dx.doi.org/10.1109/TWC.2020.2989302.

Yu, Y., Tang, J., Huang, J., Zhang, X., So, D.K.C., Wong, K.-K., 2021. Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm. IEEE Trans. Commun. 69 (9), 6361–6374. http://dx.doi.org/10.1109/TCOMM.2021.3089476.

Zhang, S., Cao, R., 2022. Multi-objective optimization for UAV-enabled wireless powered IoT networks: an LSTM-based deep reinforcement learning approach. IEEE Commun. Lett. 26 (12), 3019–3023.

Zhang, D., Li, X., Ren, G., Yao, J., Chen, K., Li, X., 2023a. Three-dimensional path planning of UAVs in a complex dynamic environment based on environment exploration twin delayed deep deterministic policy gradient. Symmetry 15, 1371. http://dx.doi.org/10.3390/sym15071371.

Zhang, Z., Wu, J., Dai, J., He, C., 2020. A novel real-time penetration path planning algorithm for stealth UAV in 3D complex dynamic environment. IEEE Access 8, 122757–122771. http://dx.doi.org/10.1109/ACCESS.2020.3007496.

Zhang, X., Xia, S., Li, X., Zhang, T., 2022. Multi-objective particle swarm optimization with multi-mode collaboration based on reinforcement learning for path planning of unmanned air vehicles. Knowl.-Based Syst. 250, 109075.

Zhang, D., Xuan, Z., Zhang, Y., Yao, J., Li, X., Li, X., 2023b. Path planning of unmanned aerial vehicle in complex environments based on state-detection twin delayed deep deterministic policy gradient. Machines 11, 108. http://dx.doi.org/10.3390/machines11010108.

Zhao, J., Wang, W., Yang, C., Li, Y., Yang, L., Cheng, J., 2024. A new efficient algorithm for short path planning of the vertical take-off and landing air-ground integrated vehicle. Eng. Appl. Artif. Intell. 127, 107386.

Zhou, Y., Long, L., Lin, Y., 2023. Application research of "vehicle+ UAV" mode based on floyd and genetic algorithm in 5G era. In: 2023 IEEE International Conference on Image Processing and Computer Applications. ICIPCA, IEEE, pp. 1705–1709.

Zhu, B., Bedeer, E., Nguyen, H.H., Barton, R., Henry, J., 2021. UAV trajectory planning in wireless sensor networks for energy consumption minimization by deep reinforcement learning. IEEE Trans. Veh. Technol. 70, 9540–9554. http://dx.doi.org/10.1109/TVT.2021.3102161.