Benchmarking Dynamic Path Planning Algorithms for Urban Low-Altitude Logistics

Xuanzheng Lin*[1], Runquan Zuo[1], Wenzhuo Li[1], Shuo Xu[1], Dawei Yan[1], Zihao Han[2]

[1]School of Computer Science and Engineering
[2]School of Physics
Southeast University, Nanjing, China

Abstract

As a critical component of smart cities and on-demand delivery systems, urban low-altitude unmanned aerial vehicle (UAV) logistics relies heavily on dynamic path planning to balance delivery efficiency, energy consumption, and operational safety. While existing approaches primarily utilize either heuristic optimization algorithms or data-driven reinforcement learning (RL), a systematic understanding of their performance trade-offs under dynamic order insertion and online replanning remains limited. To address the unique constraints of 3D urban airspace and dynamic order flows, this paper systematically evaluates these two paradigms—comparing the global search capabilities of heuristics against the policy-based inference of RL. Our analysis reveals a critical trade-off: heuristic algorithms excel in global path optimality and energy efficiency but struggle with convergence bottlenecks and high replanning latency during dynamic disturbances. Conversely, RL-based methods offer superior online responsiveness and adaptability, yet face generalization and stability risks in congested, heterogeneous environments. Motivated by these findings, we propose a novel hybrid path-planning framework. This framework leverages heuristic optimization for energy-aware global route initialization, while employing RL for localized, real-time dynamic replanning and obstacle avoidance. Experimental results across multi-dimensional criteria demonstrate that the proposed hybrid approach effectively bridges the gap, offering a reliable, efficient, and energy-sustainable solution for dynamic urban low-altitude logistics.

(Written by: Shuo Xu)

1 Introduction

Urban low-altitude unmanned aerial vehicle (UAV) delivery has received growing attention as a potential complement to last-mile logistics, especially in applications where rapid service, flexible access, and reduced ground-vehicle dependence are operationally valuable [16, 6]. Early routing studies, such as the flying sidekick traveling salesman problem and the traveling salesman problem with drone, formalized UAV-assisted delivery as combinatorial optimization problems [15, 1]. However, these early formulations were largely developed for truck–drone coordination or relatively static delivery settings and inherited the two-dimensional routing assumptions of classical vehicle routing problems. Pure low-altitude UAV logistics in dense urban areas raises a fundamentally different planning challenge: vehicles must operate in three-dimensional (3D) airspace subject to altitude restrictions, obstacle clearance, time-varying no-fly zones, and inter-UAV separation requirements, while simultaneously reacting to uncertain order arrivals and managing limited battery resources [10, 5].

*E-mails: {xuanzheng.lin, runquan.zuo, wenzhuo.li, shuo.xu, dawei.yan, zihao.han}@seu.edu.cn

These characteristics distinguish UAV path planning from conventional ground-vehicle routing in several important respects. Ground delivery routes are constrained by a fixed road network, whereas low-altitude UAV routes must navigate a continuous 3D airspace with dynamically changing feasible regions [7, 11]. Moreover, multirotor UAV energy consumption follows a nonlinear, U-shaped power–speed relationship, meaning the shortest geometric path is not necessarily the least costly or most feasible trajectory [4, 23]. Dynamic order arrivals further transform the planning problem from a static routing task into an online decision problem requiring continuous replanning throughout the operational horizon [17, 18].

Existing algorithmic approaches to UAV path planning can be broadly divided into two paradigms: metaheuristic optimization and learning-based decision making. Metaheuristics such as ant colony optimization (ACO), genetic algorithms (GA), and particle swarm optimization (PSO) provide flexible global search capabilities and have been widely applied to routing and 3D path planning [3, 19, 13]. Their principal limitation in dynamic environments is that route changes typically require repeated search, neighborhood reconstruction, or solution repair, introducing nontrivial computational overhead. Reinforcement learning (RL), by contrast, can learn a generalized policy during offline training that supports near-instantaneous online decision making without per-instance re-optimization [14, 20]. Recent advances in attention-based architectures and multi-agent RL frameworks have demonstrated that learned policies can handle variable-size routing states and coordinated fleet operation [9, 12, 2, 22]. Nevertheless, RL-based methods may suffer from limited interpretability, unstable generalization under out-of-distribution conditions, and safety risks when the airspace becomes highly constrained.

Despite this growing body of work, three critical gaps remain. First, most UAV routing studies rely on simplified two-dimensional or static airspace representations, which fail to capture the spatiotemporal complexity of dynamic urban environments with time-varying no-fly zones and congestion. Second, direct empirical comparisons between heuristic and RL-based replanning methods under standardized dynamic conditions are scarce, leaving their relative operational strengths in dynamic order insertion and online route repair largely unsubstantiated. Third, many evaluations emphasize route length or single-objective solution quality, while practical UAV logistics demands simultaneous consideration of replanning latency, energy consumption, service reliability, and safety. The multi-dimensional trade-offs among these criteria have yet to be systematically quantified.

To address these gaps, this paper develops a unified benchmark for dynamic path planning in urban low-altitude UAV logistics and proposes a hybrid planning framework that combines global heuristic search with local learned adaptation. The main contributions are as follows:

1. A 3D dynamic benchmark formulation is constructed that integrates airspace feasibility, time-varying no-fly zones, nonlinear energy consumption, and stochastic order arrivals, enabling fair and reproducible comparison across planning algorithms.
2. Representative heuristic (ACO, GA, PSO), reinforcement learning (MAPPO), and hybrid planning strategies are systematically compared under identical environmental assumptions, providing a direct empirical comparison of these paradigms in dynamic urban low-altitude logistics.
3. A hybrid global–local planning framework is proposed that leverages ACO-based energy-aware route initialization for global path quality while employing MAPPO-based local dynamic replanning with safety filtering for real-time adaptation to order insertion and airspace updates.
4. A multi-scenario evaluation is conducted across stable, moderately dynamic, and highly dynamic conditions to quantify trade-offs among route quality, energy consumption, replanning latency, order success rate, and safety violations, complemented by an ablation study that isolates the contribution of each component in the hybrid framework.

The results reveal that no single planning paradigm dominates across all operating conditions. Heuristic optimization excels in global path quality and energy efficiency under stable environments but suffers from convergence bottlenecks and elevated replanning latency under high-frequency disturbances. RL methods offer superior online responsiveness but face generalization and stability risks in congested, heterogeneous airspace. The proposed hybrid framework effectively bridges these complementary strengths, achieving the best overall balance of efficiency, reliability, and safety under highly dynamic urban low-altitude logistics scenarios.

(Written by: Dawei Yan)

2 Literature Review

2.1 UAV-Based Logistics and Its Divergence from Ground Delivery

Urban UAV logistics has emerged as a prominent research domain over the past decade, propelled by the rapid growth of e-commerce and the inherent limitations of ground-based last-mile delivery. Foundational work by Murray and Chu [15] and Agatz et al. [1] formalized UAV-assisted delivery as combinatorial optimization problems; however, these early formulations treated UAVs largely as supplements to ground vehicles, inheriting the two-dimensional routing assumptions of classical Vehicle Routing Problems (VRP). Pure UAV delivery networks, by contrast, require altitude to be treated as a first-class decision variable, and operational frameworks such as the FAA's UAS Traffic Management (UTM) initiative [10] and the EU's U-Space regulatory framework [5] have since established the institutional context within which technical research is situated.

2.2 Modeling of Airspace, Energy, and Dynamic Demand

Accurate spatial modeling is a prerequisite for any computationally tractable path planning framework. In the UAV literature, airspace is most commonly represented through one of three paradigms: continuous potential field models [8], volumetric occupancy grid maps [7], and graph-based decompositions such as roadmaps and visibility graphs [11]. Each offers a different trade-off between spatial fidelity and computational overhead. In urban low-altitude applications, dynamic no-fly zones (NFZs) arising from temporary restrictions around hospitals, airports, and public events impose non-stationary constraints that challenge all three paradigms; operational frameworks such as UTM [10] explicitly identify real-time constraint management in shared low-altitude airspace as a central unsolved engineering challenge.

Energy consumption adds a further layer of complexity: unlike ground vehicles, multirotor UAVs exhibit a non-linear, U-shaped power-velocity relationship [23], meaning the energy-optimal trajectory need not coincide with the shortest path, a distinction that ground delivery models routinely ignore. On the demand side, real deployments involve stochastic order arrivals typically modeled as Poisson processes [17], transforming the static VRP into a Dynamic VRP (DVRP) in which continuous replanning is required throughout the operational horizon [18].

2.3 Heuristic Optimization Methods

Heuristic and metaheuristic algorithms represent the dominant approach to large-scale routing optimization. Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) have both been applied to 3D UAV path planning, with comparative experiments demonstrating PSO's superior convergence speed on single-objective terrain-avoidance tasks [19]. Ant Colony Optimization (ACO) has shown particular effectiveness on graph-structured routing tasks through its pheromone-guided search mechanism [3]. Their core limitation in dynamic settings is a batch-processing nature: when the problem changes, the algorithm must be re-invoked, introducing reinitialization overhead that can compromise operational responsiveness. Although adaptive ACO variants employing immigrant schemes have been developed to incrementally repair incumbent solutions under dynamically changing customer demands [13], their performance under high-frequency disturbances in the UAV logistics context remains an open question.

2.4 Reinforcement Learning for Dynamic Replanning

Reinforcement learning (RL) offers a structurally different response to dynamic replanning. By learning a generalized policy during an offline training phase, an RL agent can produce near-instantaneous decisions at deployment without per-instance re-optimization. Deep Q-Networks [14] established the feasibility of learning navigation policies for high-dimensional state spaces, and subsequent architectures, including Proximal Policy Optimization [20] and attention-based Pointer Networks [9], extended this capability to continuous action spaces and variable-length order sets respectively. Multi-agent RL frameworks further enable coordinated fleet operation: Lowe et al. [12] established centralized training with decentralized execution (CTDE) as a principled approach to cooperative multi-agent learning, and Brittain and Wei [2] demonstrated its effectiveness for UAV separation assurance in high-density stochastic airspace, proposing a scalable attention-based PPO framework capable of resolving inter-vehicle conflicts among variable numbers of aircraft. Nevertheless, the generalization of learned policies to out-of-distribution (OOD) conditions remains a recognized limitation.

2.5 Research Gaps and Motivation

Despite the advancements in UAV logistics, a synthesis of the existing literature reveals three critical gaps that hinder the transition from theoretical models to real-world deployment. Primarily, most routing studies rely on simplified two-dimensional or static problem formulations, which fail to capture the spatial-temporal complexity and the high-dimensional constraints inherent in dynamic urban environments. Furthermore, while reinforcement learning has emerged as a promising alternative to traditional heuristics, there is a notable scarcity of direct empirical comparisons between these methodologies under standardized, controlled dynamic conditions, leaving the relative operational merits of each approach largely unsubstantiated. Finally, the multi-dimensional trade-offs between response latency, energy efficiency, and computational overhead have yet to be systematically quantified.

The present study is therefore designed to bridge these divides by introducing a unified modeling framework and a structured comparative evaluation of dynamic path planning algorithms for urban UAV logistics. Specifically, it aims to examine how heuristic optimization and reinforcement learning methods perform under the same dynamic airspace, energy, and demand conditions. The central question guiding this study is how different dynamic path planning algorithms balance route quality, energy efficiency, response latency, and computational overhead in urban low-altitude UAV logistics.

(Written by: Runquan Zuo)

3 Methodology

This section develops a unified benchmarking framework for dynamic path planning in urban low-altitude UAV logistics. The logic of the method is as follows. First, a common operational environment is defined so that all algorithms face the same 3D airspace, energy, demand, and safety constraints. Second, heuristic, reinforcement learning, and hybrid replanning strategies are specified on top of this shared environment. Third, the same dynamic triggers and performance metrics are used to connect the methodology to the experimental evaluation. This structure directly follows the research gap identified above: the comparison is not a test of isolated algorithms, but a test of how different planning mechanisms respond to the same dynamic logistics system.

3.1 System Model and Problem Formulation

This subsection defines the shared environment used by all algorithms. Rather than describing a complete UAV simulator, it keeps only the state variables and constraints needed for a fair comparison: feasible airspace, fleet energy, dynamic demand, and the common objective.

3.1.1 Airspace Feasibility

Let $\mathcal{W} \subset \mathbb{R}^3$ denote the urban operating region and $[0, T]$ the planning horizon. At time t, the feasible airspace is

$$\mathcal{F}(t) = \mathcal{W} \setminus \left(\mathcal{O}_{\text{static}} \cup \mathcal{O}_{\text{dynamic}}(t) \cup \mathcal{Z}(t)\right),$$

where $\mathcal{O}_{\text{static}}$ is the set of fixed obstacles, $\mathcal{O}_{\text{dynamic}}(t)$ is the set of moving or temporary obstacles, and $\mathcal{Z}(t) = \bigcup_m \mathcal{Z}_m(t)$ denotes time-varying no-fly zones. A buffered region $\mathcal{F}_\delta(t)$ is used to account for localization error and minimum separation requirements, consistent with low-altitude traffic management concepts [10, 5]. An edge e_{ij} with trajectory $\gamma_{ij}(\tau)$ is feasible if $\gamma_{ij}(\tau) \in \mathcal{F}_\delta(\tau)$ for all $\tau \in [t_i, t_j]$. The same physical airspace is represented as an occupancy grid for reinforcement learning and as a graph $G_t = (V_t, E_t)$ for heuristic search [7, 11].

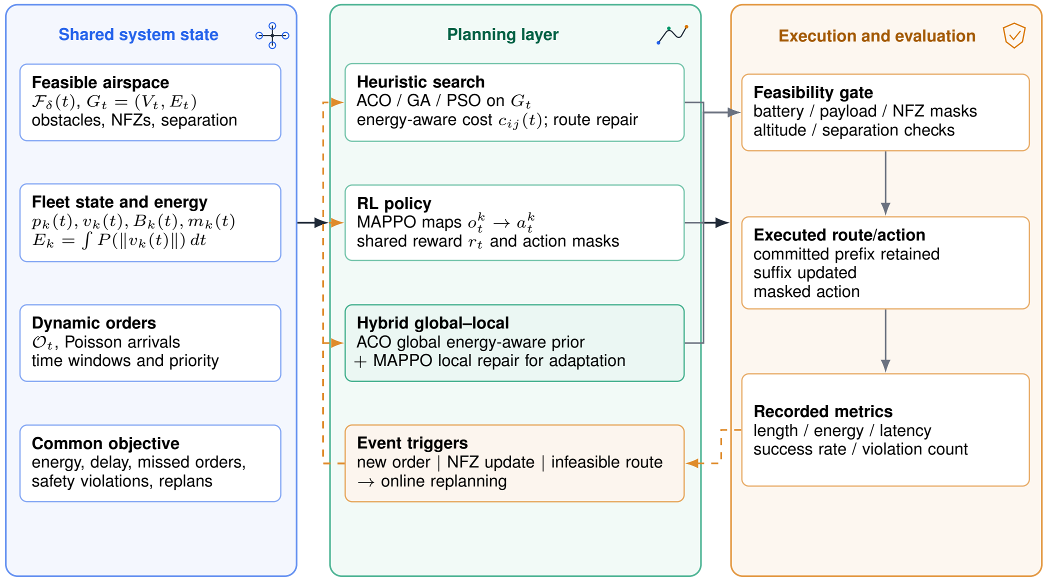

Figure 1: System architecture of the unified benchmark for dynamic path planning in urban low-altitude UAV logistics. Shared airspace, fleet, demand, and objective information feed the three planning strategies. Dynamic events trigger replanning; feasibility checks and common metrics close the benchmark loop.

3.1.2 Fleet State and Energy

For each UAV $k \in \mathcal{K}$, the state at time t is summarized by position $p_k(t)$, velocity $v_k(t)$, battery level $B_k(t)$, and payload $m_k(t)$. The state must satisfy altitude, speed, payload, battery, and inter-UAV separation constraints. For example, $\|p_k(t) - p_l(t)\|_2 \geq d_{\text{sep}}$ for $k \neq l$.

Because rotary-wing UAV energy is not proportional to distance, the propulsion power is modeled as a nonlinear function of speed [4, 23]:

$$P(v) = P_0\left(1 + \frac{3v^2}{U_{\text{tip}}^2}\right) + P_i\left(\sqrt{1 + \frac{v^4}{4v_0^4}} - \frac{v^2}{2v_0^2}\right)^{1/2} + \frac{1}{2}d_0\rho s A v^3.$$

The energy consumed over a trajectory segment is

$$E_k = \int_{t_s}^{t_e} P(\|v_k(t)\|)\,dt,$$

with optional extensions for vertical motion, acceleration, and payload effects. This keeps the benchmark energy-aware without making route length the only measure of cost.

3.1.3 Dynamic Demand and Common Objective

Orders are revealed during the planning horizon. The number of revealed orders $N(t)$ is modeled as a Poisson process, a standard abstraction for stochastic demand in dynamic routing [17, 18]:

$$\Pr\{N(t+\Delta) - N(t) = k\} = \frac{(\lambda\Delta)^k e^{-\lambda\Delta}}{k!}.$$

Each order is denoted by $o_i = (r_i, \ell_i, q_i, [a_i, b_i], h_i, w_i)$, representing release time, delivery location, payload demand, time window, service duration, and priority. At time t, the active order set is $\mathcal{O}_t = \{i : r_i \leq t,\ i \text{ is not yet served}\}$.

The shared planning objective combines the operational costs that later become the evaluation metrics:

$$\min_\pi \mathbb{E}\left[\sum_{k\in\mathcal{K}} E_k + \alpha_T \sum_{i\in\mathcal{S}} \max(T_i^{\text{start}} - b_i, 0) + \alpha_M N_{\text{missed}} + \alpha_C N_{\text{viol}} + \alpha_R N_{\text{replan}}\right].$$

Here, $\mathcal{S}$ is the served-order set, N_{missed} counts missed or rejected orders, N_{viol} counts safety violations, and N_{replan} counts replanning operations. The objective is therefore a compact definition of the shared benchmark: algorithms are compared by the same energy, delay, reliability, safety, and responsiveness criteria.

3.2 Heuristic-Based Replanning Strategy

An adaptive Ant Colony Optimization algorithm with immigrant schemes is used as the representative heuristic replanning strategy. ACO is suitable for graph-based routing because route construction can be guided by pheromone values and heuristic edge costs [3]. Immigrant schemes are included because they allow previous solutions to be partially reused when a dynamic routing instance changes [13]. At replanning epoch t, each edge e_{ij} is assigned a composite cost,

$$c_{ij}(t) = \omega_E E_{ij}(t) + \omega_T t_{ij}(t) + \omega_R R_{ij}(t) + \omega_W W_{ij}(t),$$

where $E_{ij}(t)$ is the energy cost, $t_{ij}(t)$ is the travel time, $R_{ij}(t)$ is the airspace risk, and $W_{ij}(t)$ is the time-window penalty. The heuristic desirability is $\eta_{ij}(t) = 1/\left(c_{ij}(t) + \epsilon\right)$.

For ant a, the probability of moving from node i to node j is

$$p_{ij}^{(a)}(t) = \frac{\left[\tau_{ij}(t)\right]^{\alpha}\left[\eta_{ij}(t)\right]^{\beta}}{\sum_{\ell \in \mathcal{N}_i^{(a)}(t)} \left[\tau_{i\ell}(t)\right]^{\alpha}\left[\eta_{i\ell}(t)\right]^{\beta}},$$

where $\tau_{ij}(t)$ is the pheromone value, α and β control the relative importance of pheromone and heuristic information, and $\mathcal{N}_i^{(a)}(t)$ is the feasible neighborhood.

Dynamic replanning is triggered by new order arrivals, no-fly-zone updates, or infeasibility of an existing route. To reduce route instability, only the unexecuted suffix of a route is replanned, while the committed prefix is preserved. Immigrant schemes are introduced to improve adaptability. Random immigrants are used under abrupt changes, elite immigrants are generated from previous high-quality solutions, and memory-based immigrants are used when environmental patterns are recurrent.

The online computational cost of the heuristic method is decomposed as $T_{\text{heur}} = T_{\text{neighbor}} + T_{\text{search}} + T_{\text{repair}} + T_{\text{memory}}$. If n_t is the number of active routing nodes, μ is the number of ants, and I is the number of iterations, the dominant search cost is approximately $O\left(I\mu n_t^2\right)$. Under extremely frequent order insertion or no-fly-zone updates, the algorithm may become limited by repeated neighborhood reconstruction rather than pheromone updating. This is treated as an important stress condition in the evaluation.

3.3 Reinforcement Learning Framework

The reinforcement learning method is formulated under the centralized training and decentralized execution paradigm. During training, the critic can access a global state summary, while during execution each UAV acts only on its local observation. The multi-agent system is modeled as a cooperative partially observable Markov decision process with UAV agents $\mathcal{K}$, global state s_t, local observation o_t^k, action a_t^k, shared reward r_t, and discount factor γ. This formulation follows the standard reinforcement learning view of sequential decision making and extends it to multi-agent centralized training with decentralized execution [21, 12].

For UAV k, the local observation is defined as

$$o_t^k = \left[p_t^k, v_t^k, B_t^k, m_t^k, \mathcal{N}_{\text{order}}^k(t), \mathcal{N}_{\text{uav}}^k(t), \mathcal{B}_t^k, \phi_t\right],$$

where p_t^k is the 3D position, v_t^k is velocity, B_t^k is remaining battery, m_t^k is payload, $\mathcal{N}_{\text{order}}^k(t)$ is the local visible order set, $\mathcal{N}_{\text{uav}}^k(t)$ is the nearby UAV set, $\mathcal{B}_t^k$ is the local airspace mask, and ϕ_t contains temporal demand features. The centralized critic uses an aggregated global state $s_t = [\psi_{\text{fleet}}, \psi_{\text{order}}, \psi_{\text{airspace}}, \psi_{\text{time}}]$, rather than directly concatenating all raw observations. This reduces dimensional growth when the fleet or order set becomes large.

Two action designs are considered. In the continuous-control setting, the action is $a_t^k = \left(\Delta x_t^k, \Delta y_t^k, \Delta z_t^k, \Delta v_t^k\right)$, sampled from a Gaussian policy $\pi_\theta\left(a_t^k | o_t^k\right)$. In the routing-oriented setting, a hybrid action $a_t^k = \left(i_t^k, \Delta v_t^k\right)$ is used, where i_t^k is the selected order or waypoint. An action mask prevents the agent from selecting unrevealed, unreachable, battery-infeasible, or no-fly-zone-violating targets. The shared reward is defined as

$$r_t = \lambda_S r_t^{\text{serve}} - \lambda_E \sum_{k \in \mathcal{K}} \Delta E_t^k - \lambda_D D_t - \lambda_C C_t - \lambda_R R_t - \lambda_U U_t.$$

Here, r_t^{serve} rewards completed deliveries, ΔE_t^k is the energy consumed by UAV k, D_t is the delay penalty, C_t is the collision or no-fly-zone violation penalty, R_t penalizes excessive replanning, and $U_t = \sum_k \left\| a_t^k - a_{t-1}^k \right\|_2^2$ penalizes unstable control changes. The smoothness term is necessary because the U-shaped power model may otherwise encourage unrealistic speed oscillation around an energy-efficient speed.

The policy is trained offline using PPO or MAPPO-style optimization. PPO is selected because it provides a stable clipped policy update, and MAPPO-style implementations have been shown to be strong baselines in cooperative multi-agent settings [20, 22]. The clipped policy objective is

$$L^{\text{clip}}(\theta) = \mathbb{E}_t \left[\min\left(\rho_t(\theta) \widehat{A}_t, \operatorname{clip}\left(\rho_t(\theta), 1-\epsilon, 1+\epsilon\right) \widehat{A}_t \right) \right],$$

where $\rho_t(\theta) = \pi_\theta\left(a_t | o_t\right) / \pi_{\theta_{\text{old}}}\left(a_t | o_t\right)$ and $\widehat{A}_t$ is the estimated advantage. During online deployment, the decision time is $T_{\text{RL}} = T_{\text{forward}} + T_{\text{mask}} + T_{\text{safety}}$, which is expected to be lower than repeated heuristic search. However, when the feasible airspace becomes highly fragmented, the RL agent may suffer from sparse positive rewards and excessive action masking. Therefore, stress tests with dense no-fly zones are included to evaluate the robustness of the learned policy.

3.4 Hybrid Global–Local Planning Strategy

The hybrid strategy is introduced to make the comparison consistent with the paper's central argument: global route quality and real-time adaptability are complementary rather than mutually exclusive. In this strategy, ACO first produces an energy-aware global route prior over the graph G_t. When new orders or airspace updates arrive, the committed prefix of each route is preserved, the unexecuted suffix is checked against the updated feasible region $\mathcal{F}_\delta(t)$, and the RL policy performs local adjustment for waypoint selection, speed control, and short-horizon avoidance.

The hybrid method therefore uses the same state variables, constraints, and objective terms defined above, but assigns different roles to the two algorithmic components. The heuristic layer supplies interpretable global structure and energy-aware initialization; the learning layer supplies fast local adaptation and continuous replanning. Safety masks are applied before execution so that battery-infeasible, unrevealed, or no-fly-zone-violating actions are removed from the available action set. This global–local design is the methodological bridge between the problem formulation and the later ablation study, where the energy-aware edge cost, action mask, and immigrant-based repair mechanism are tested separately.

3.5 Benchmarking Workflow

All algorithms are evaluated through the same event-driven workflow as shown in Figure 1. At the beginning of each simulation, the environment initializes the airspace, fleet state, depots, charging points, and initial order set. During the planning horizon, new orders and no-fly-zone updates trigger replanning. Each algorithm receives the same updated feasible region, active order set, battery state, and safety constraints. The resulting route or action is then checked for feasibility before execution, and the system records energy use, route length, latency, order completion, and safety violations.

This workflow ensures that the experimental comparison is aligned with the methodology. The algorithms differ in their decision mechanism, but not in their environment, demand process, constraint set, or evaluation metrics. As a result, the later performance differences can be interpreted as differences in planning behavior rather than differences in problem definition.

(Written by: Xuanzheng Lin)

4 Experiments & Discussion

This section reports and interprets the experimental findings in an integrated form. Following the logic of the methodology, the section first clarifies the experimental setting and reporting metrics, then presents the quantitative results, and finally discusses what the results imply for dynamic path planning in urban low-altitude logistics. This organization avoids separating numerical results from their interpretation: each set of results is discussed in relation to the research problem introduced earlier.

4.1 Experimental Setup

The simulation experiment was constructed according to the modeling assumptions defined in the Methodology section. The low-altitude logistics network contains depots, delivery nodes, charging points, feasible waypoints, fixed urban obstacles, moving obstacles, and time-varying no-fly zones. All algorithms are evaluated under the same 3D airspace constraints, battery constraints, payload constraints, and dynamic order arrival process. Each scenario is repeated with 30 random seeds, and the reported values are averaged over all runs.

The operating area is a 5 km x 5 km urban region with an altitude range of 30-120 m. The planning horizon is 120 minutes, and the UAV fleet size is fixed at 10. Orders are generated according to a Poisson arrival process, while temporary no-fly zones are updated at different frequencies to represent increasing levels of airspace disturbance. Table 1 summarizes the three scenario types: stable, moderately dynamic, and highly dynamic.

Table 1: Simulation scenario settings

Scenario	Order intensity	NFZ update interval	Congestion ratio	Planning horizon	Fleet size
Stable	0.05 orders/min	30 min	5%	120 min	10 UAVs
Moderately dynamic	0.10 orders/min	15 min	15%	120 min	10 UAVs
Highly dynamic	0.20 orders/min	5 min	25%	120 min	10 UAVs

4.2 Baseline and Evaluation Metrics

Five algorithms are compared in the experiment. The first three are representative heuristic optimization methods, including ACO, GA, and PSO. The reinforcement-learning baseline is a MAPPO-style policy trained under centralized training and decentralized execution. In addition, a hybrid method is included, where ACO is used to generate global energy-aware route initialization and the learned RL policy is used for local replanning and obstacle avoidance.

Table 2: Baseline algorithms and main implementation settings

Algorithm	Main setting	Planning role
ACO	40 ants, 80 iterations, evaporation rate 0.2, 20% immigrant solutions	Graph-based global replanning
GA	Population size 80, crossover rate 0.8, mutation rate 0.1, 80 generations	Population-based route optimization
PSO	Swarm size 60, inertia 0.7, cognitive/social coefficients 1.5	Continuous waypoint search
MAPPO	Two-layer actor-critic networks, hidden size 256, $\gamma = 0.99$, PPO clip $= 0.2$	Policy-based online decision making
Hybrid	ACO global initialization plus MAPPO local replanning with safety masks	Global-local collaborative planning

The algorithms are evaluated from five perspectives: route quality, energy sustainability, online responsiveness, service reliability, and safety. The average route length measures the geometric efficiency of delivered orders. Average energy consumption measures the propulsion energy required per successfully served order. Replanning latency records the average online decision time after order insertion or airspace updates. The order success rate is the percentage of orders completed within the allowed service window. Safety violations include collisions, minimum-separation violations, battery infeasibility, and no-fly-zone intrusions. For route length, energy consumption, latency, and safety violations, lower values are preferred; for order success rate, higher values are preferred.

4.3 Overall Quantitative Comparison

Table 3 reports the comparative performance of the five algorithms under the three dynamic scenarios. The table answers the central empirical question of the paper: whether heuristic, reinforcement-learning, and hybrid planners behave differently when they face the same airspace, demand process, and evaluation metrics.

Table 3: Comparative results under different dynamic scenarios

Scenario	Algorithm	Length (km/order)	Energy (Wh/order)	Latency (s)	Success (%)	Violations (/100 orders)
Stable	ACO	8.42	162.8	8.60	96.4	0.8
Stable	GA	8.61	168.3	10.20	94.8	1.1
Stable	PSO	8.55	166.7	7.90	95.2	1.0
Stable	MAPPO	8.94	174.5	0.38	94.9	1.4
Stable	Hybrid	8.49	164.1	1.26	96.9	0.6
Moderately dynamic	ACO	9.73	191.2	13.80	90.5	2.7
Moderately dynamic	GA	10.05	199.5	16.40	88.9	3.2
Moderately dynamic	PSO	9.96	196.8	12.70	89.7	3.0
Moderately dynamic	MAPPO	10.18	203.6	0.45	92.8	2.2
Moderately dynamic	Hybrid	9.81	193.0	1.48	94.1	1.5
Highly dynamic	ACO	12.56	248.7	22.90	80.6	6.8
Highly dynamic	GA	13.08	260.4	25.70	77.9	7.4
Highly dynamic	PSO	12.91	256.1	20.80	79.2	7.0
Highly dynamic	MAPPO	12.37	241.5	0.51	87.6	4.3
Highly dynamic	Hybrid	11.98	233.8	1.73	90.2	3.1

In the stable scenario, heuristic search remains competitive because the environment changes slowly enough for global optimization to be useful. ACO obtains the shortest average route and the lowest energy consumption, with 8.42 km/order and 162.8 Wh/order. The Hybrid method is only slightly worse on these two measures, with 8.49 km/order and 164.1 Wh/order, but it reduces replanning latency from 8.60 s to 1.26 s. It also achieves the highest order success rate and the fewest safety violations in the stable setting. This result shows that the Hybrid method preserves most of the global route-quality advantage of ACO while already improving responsiveness and safety.

As environmental disturbance increases, the advantage of the Hybrid method becomes more substantial. In the moderately dynamic scenario, ACO remains marginally better in route length and energy, but the Hybrid method reduces latency from 13.80 s to 1.48 s, raises success from 90.5% to 94.1%, and lowers safety violations from 2.7 to 1.5 per 100 orders. In the highly dynamic scenario, the Hybrid method achieves the best overall balance: 11.98 km/order, 233.8 Wh/order, 1.73 s latency, 90.2% success, and 3.1 violations per 100 orders. Compared with ACO, it reduces latency by 21.17 s and improves success by 9.6 percentage points. These differences provide direct quantitative support for the global-local planning perspective developed in Section 3.

4.4 Experimental Results

The experimental results reveal three main findings. First, heuristic optimization remains useful when the environment is stable and global route quality is the dominant objective. Second, reinforcement learning provides a clear advantage in low-latency replanning under frequent order insertion and airspace updates. Third, the Hybrid strategy is the most practical option in highly dynamic urban logistics because it combines energy-aware global planning, rapid local adaptation, and explicit safety filtering.

As shown in Table 3, ACO achieves the shortest route length and the lowest energy consumption in the stable scenario, with 8.42 km/order and 162.8 Wh/order. However, its replanning latency is 8.60 s. In comparison, the Hybrid method maintains a similar route length and energy consumption while reducing latency to 1.26 s. It also achieves the highest success rate of 96.9% and the lowest safety violation rate of 0.6 per 100 orders.

When the environment becomes more dynamic, the advantage of the Hybrid method becomes clearer. In the highly dynamic scenario, the Hybrid method achieves 11.98 km/order in route length, 233.8

Wh/order in energy consumption, 1.73 s in replanning latency, 90.2% in order success rate, and 3.1 safety violations per 100 orders. These results indicate that the Hybrid method performs better than using heuristic optimization or reinforcement learning alone when frequent order arrivals and no-fly-zone updates occur.

Table 4 further presents the ablation results of the Hybrid framework. Removing the energy-aware edge cost increases energy consumption from 233.8 to 245.4 Wh/order. Removing the action mask increases safety violations from 3.1 to 7.8 per 100 orders. Removing immigrant-based repair reduces the success rate from 90.2% to 87.9%. These results show that each component contributes to the overall performance of the Hybrid framework.

Overall, the results support the effectiveness of combining global heuristic planning with local reinforcement-learning-based replanning. The Hybrid method provides a better balance among route quality, energy consumption, response speed, service reliability, and safety. These findings provide the basis for the following discussion.

4.5 Replanning Responsiveness and Algorithmic Trade-offs

The latency results show a clear trade-off between global search quality and online responsiveness. Heuristic methods such as ACO, GA, and PSO perform well when the graph changes slowly, because they can repeatedly evaluate feasible routes and optimize path length or energy-related objectives over a relatively stable environment. However, the same mechanism becomes costly under frequent order insertion and no-fly-zone updates. From the stable to the highly dynamic scenario, ACO latency increases from 8.60 s to 22.90 s, GA latency increases from 10.20 s to 25.70 s, and PSO latency increases from 7.90 s to 20.80 s.

This increase is not only an implementation issue. It reflects the batch-processing nature of heuristic replanning, where each disturbance requires neighborhood reconstruction, route repair, and repeated solution evaluation. This interpretation is consistent with dynamic vehicle routing and immigrant-based ACO studies, which show that adaptive repair improves flexibility but does not remove the cost of repeated search [17, 18, 13].

MAPPO shows the opposite behavior. Its latency remains very low across all scenarios, increasing only from 0.38 s to 0.51 s. This confirms the advantage of policy inference for online decision making under centralized training and decentralized execution. Nevertheless, pure MAPPO does not provide the best overall solution. In the stable scenario, its route length and energy consumption are both higher than those of ACO and the Hybrid method. In the highly dynamic scenario, MAPPO is faster than the Hybrid method, but it consumes more energy, achieves a lower success rate, and produces more safety violations. Fast policy inference is therefore valuable, but it is insufficient when global route quality, battery feasibility, and safety constraints must be optimized together.

4.6 Route Quality and Energy Sustainability

The energy results show that shortest-path planning is not an adequate objective for low-altitude UAV logistics. Because multirotor energy consumption follows a nonlinear power-speed relationship, route length and energy cost do not always change in the same way. In stable conditions, ACO performs strongly because its composite edge cost directly includes energy, travel time, airspace risk, and time-window penalty. Under high disturbance, however, route invalidation, hovering, detours, and fragmented feasible airspace make energy performance depend more strongly on replanning behavior than on geometric distance alone.

The highly dynamic scenario illustrates this pattern. MAPPO produces a slightly shorter route than ACO and consumes less energy, suggesting that fast local adaptation can avoid some unnecessary detours. The Hybrid method further improves both measures, reaching 11.98 km/order and 233.8 Wh/order. Its advantage comes from combining an energy-aware global route prior with local policy adjustment, so that the UAV can avoid repeated full-route invalidation while still respecting battery, no-fly-zone, and separation constraints. In this sense, the Hybrid method is not merely a compromise between ACO and MAPPO; it assigns different planning functions to the two components.

4.7 Ablation Study of the Hybrid Framework

To further examine the contribution of each component, an ablation study is conducted under the highly dynamic scenario. Removing the energy-aware edge cost increases average energy consumption, while removing action masks greatly increases safety violations. Removing immigrant-based ACO repair reduces adaptation to recurrent no-fly-zone patterns and lowers the order success rate. The full hybrid framework achieves the best balance among energy efficiency, service reliability, and operational safety.

Table 4: Ablation results of the hybrid framework in the highly dynamic scenario

Variant	Energy (Wh/order)	Latency (s)	Success (%)	Violations (/100 orders)
w/o energy-aware edge cost	245.4	1.69	88.7	4.0
w/o action mask	236.9	1.55	86.5	7.8
w/o immigrant-based repair	239.6	2.14	87.9	4.9
hybrid framework	233.8	1.73	90.2	3.1

The ablation results connect the quantitative comparison to the mechanism of the proposed framework. Removing the energy-aware edge cost raises energy consumption from 233.8 Wh/order to 245.4 Wh/order, confirming that global route initialization remains important even when local policy adjustment is available. Removing the action mask increases safety violations from 3.1 to 7.8 per 100 orders, showing that learned policies require explicit feasibility filtering in constrained low-altitude airspace. Removing immigrant-based repair lowers the success rate and increases violations, indicating that memory and repair mechanisms still matter when no-fly-zone patterns recur.

Together, these results clarify the methodological advantage of the Hybrid framework. Heuristic search contributes structured global feasibility and energy-aware route priors, while reinforcement learning contributes fast local response under dynamic disturbances. The safety mask then acts as a final operational filter before execution. This division of labor explains why the Hybrid method improves service completion, energy use, and safety at the same time rather than simply trading one metric for another.

4.8 Interpretation in the Context of Literature

The experimental results should also be interpreted in relation to the literature gaps identified earlier. Previous UAV routing studies often simplify the airspace into static or two-dimensional models, evaluate heuristic and learning-based methods under different assumptions, or focus mainly on route length as the dominant objective. The present experiment responds to these limitations by using a dynamic 3D benchmark, comparing heuristic optimization, reinforcement learning, and Hybrid planning under the same environment and metrics, and evaluating latency, energy, reliability, and safety together.

In this context, the Hybrid result is meaningful because it shows that global search and policy inference are complementary rather than mutually exclusive. Heuristic optimization provides strong global route quality in stable settings, while reinforcement learning provides fast local decisions when constraints change frequently. The Hybrid framework integrates these strengths and therefore offers stronger operational balance under dynamic urban constraints. The contribution of the experiment is therefore not only to compare algorithms, but also to show how different planning paradigms can be combined for realistic low-altitude logistics.

4.9 Synthesis and Practical Implications

The integrated results lead to three main findings. First, heuristic optimization remains useful when the environment is stable and global route quality is the dominant objective. Second, reinforcement learning provides a major advantage in response latency and continuous replanning under frequent order insertion and airspace updates. Third, the Hybrid strategy is the most practical option in highly dynamic urban low-altitude logistics because it combines global energy-aware optimization with rapid local adaptation and explicit safety filtering.

These findings have implications for urban UAS traffic management and logistics deployment. Static 3D maps are insufficient when orders, temporary no-fly zones, and congestion ratios change over time; planning systems need to update feasible airspace as a spatio-temporal object. This supports the motivation of UTM and U-Space frameworks, which emphasize real-time constraint management in shared low-altitude airspace [10, 5]. The results also suggest that operators should not optimize only for the shortest route. In dense urban delivery, a slightly longer route may be preferable if it reduces replanning delay, avoids hovering, improves service reliability, and lowers safety risk.

(Written by: Wenzhuo Li, Zihao Han)

5 Conclusion

This paper developed a unified benchmark for dynamic path planning in urban low-altitude UAV logistics. The study modeled a shared 3D operating environment with dynamic orders, time-varying no-fly zones, energy-aware flight costs, and safety constraints, and then compared heuristic, reinforcement-learning, and hybrid planning strategies under the same assumptions. The results show that no single planning mechanism is universally superior: heuristic search provides strong route quality in stable environments, reinforcement learning provides fast online response, and the Hybrid framework offers the best overall balance under highly dynamic conditions.

The main conclusion is that practical UAV logistics requires a global-local planning architecture rather than a purely static or single-objective routing method. By combining ACO-based global energy-aware initialization, MAPPO-based local replanning, and explicit safety masks, the Hybrid method reduces latency while maintaining strong energy, reliability, and safety performance. This supports the broader argument that urban UAV routing should be evaluated by multiple operational metrics, not only by geometric path length.

Several limitations remain. The evaluation is simulation-based and does not include wind disturbance, communication delay, localization error, sensor uncertainty, heterogeneous UAV platforms, or real traffic-control interactions. The learning-based results also depend on training coverage, reward design, and safety-filter implementation. Future research should test the framework in high-fidelity simulators and field trials, evaluate larger and more heterogeneous fleets, and examine charging-station scheduling, battery aging, and coordination with ground-based logistics systems.

(Written by: Shuo Xu)

References

[1] Niels Agatz, Paul Bouman, and Marie Schmidt. Optimization approaches for the traveling salesman problem with drone. *Transportation Science*, 52(4):965–981, 2018.

[2] Marc Brittain and Peng Wei. Autonomous separation assurance in an high-density en route sector: A deep multi-agent reinforcement learning approach. In *Proceeding of IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 3256–3262, 2019.

[3] Marco Dorigo and Luca Maria Gambardella. Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, 1997.

[4] Kevin Dorling, Jordan Heinrichs, Geoffrey G. Messier, and Sebastian Magierowski. Vehicle routing problems for drone delivery. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(1):70–85, 2017.

[5] European Commission. Commission implementing regulation (eu) 2021/664 of 22 april 2021 on a regulatory framework for the u-space. Official Journal of the European Union, L 139, 161–183, 2021.

[6] Anne Goodchild and Jordan Toy. Delivery by drone: An evaluation of unmanned aerial vehicle technology in reducing CO2 emissions in the delivery service industry. *Transportation Research Part D: Transport and Environment*, 61:58–67, 2018.

[7] Armin Hornung, Kai M. Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. OctoMap: An efficient probabilistic 3D mapping framework based on octrees. *Autonomous Robots*, 34(3):189–206, 2013.

[8] Oussama Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research*, 5(1):90–98, 1986.

[9] Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems! In *Proceedings of International Conference on Learning Representations (ICLR)*, 2019.

[10] Parimal Kopardekar, Joseph Rios, Thomas Prevot, Marcus Johnson, Jaewoo Jung, and John E. Robinson. Unmanned aircraft system traffic management (UTM) concept of operations to safely enable low altitude flight operations. In *Proceeding of AIAA Aviation Technology, Integration, and Operations Conference*, 2016.

[11] Steven M. LaValle. *Planning Algorithms*. Cambridge University Press, 2006. ISBN 9780521862059.

[12] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 6379–6390, 2017.

[13] Michalis Mavrovouniotis and Shengxiang Yang. Ant algorithms with immigrants schemes for the dynamic vehicle routing problem. *Information Sciences*, 294:456–477, 2015.

[14] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

[15] Chase C. Murray and Amanda G. Chu. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. *Transportation Research Part C: Emerging Technologies*, 54:86–109, 2015.

[16] Alena Otto, Niels Agatz, James Campbell, Bruce Golden, and Erwin Pesch. Optimization approaches for civil applications of unmanned aerial vehicles (UAVs) or aerial drones: A survey. *Networks*, 72(4):411–458, 2018.

[17] Victor Pillac, Michel Gendreau, Christelle Guéret, and Andrés L. Medaglia. A review of dynamic vehicle routing problems. *European Journal of Operational Research*, 225(1):1–11, 2013.

[18] Harilaos N. Psaraftis, Min Wen, and Christos A. Kontovas. Dynamic vehicle routing problems: Three decades and counting. *Networks*, 67(1):3–31, 2016.

[19] Vincent Roberge, Mohammed Tarbouchi, and Gilles Labonté. Comparison of parallel genetic algorithm and particle swarm optimization for real-time UAV path planning. *IEEE Transactions on Industrial Informatics*, 9(1):132–141, 2013.

[20] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[21] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018. ISBN 9780262039246.

[22] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of PPO in cooperative, multi-agent games. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 24611–24624, 2022.

[23] Yong Zeng, Jie Xu, and Rui Zhang. Energy minimization for wireless communication with rotary-wing UAV. *IEEE Transactions on Wireless Communications*, 18(4):2329–2345, 2019.