



Research Article

Cross-Layer Co-Design and Physics-Grounded Evaluation of Hybrid Reinforcement Learning and Federated ANN-PID Protocol Suites for Dense Urban UAV Delivery Networks

Randa Shaker Abd-Alhussain ^{1,*}, Qasim Radam Mahmood ², Hasan Mohammed Taqi Salih Albdaire ³, Mohammad Khaleel Sallam Ma'aitah ⁴, Sanad Abdullah Alafghani ⁵, Ashraf Al-Shaikh Khalil ⁶, Ashraf Zaher ⁷

¹ Department of Scholarships and Cultural Relations, University of Babylon, Babylon, Iraq.

² Computer Science and Artificial Intelligence, College of Education for Pure Sciences, University of Diyala, Diyala, Iraq.

³ Department of Computer Science, College of Computer Science and Information Technology, Wasit University, Wasit, Iraq.

⁴ Electrical Engineering / Robotics and Artificial Intelligence Engineering, Applied Science Private University, Amman, Jordan.

⁵ CCSIT, IS Dep, King Faisal University, 31982, Al-Ahsa, Saudi Arabia.

⁶ Deanship of E-learning and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia.

⁷ Vice-presidency for Postgraduate Studies and Scientific Research, King Faisal University, Al Ahsa 31982, Saudi Arabia.

ARTICLE INFO

Article History

Received 15 Jul 2026
Revised 18 Aug 2026
Accepted 21 Sep 2026
Published 24 Oct 2026

Keywords

UAV networks,
cross-layer optimization,
deep reinforcement
learning,
federated learning,
CMA-ES calibration,
AODV,
Q-learning routing,
ANN-PID control,
urban air mobility.



ABSTRACT

Urban unmanned aerial vehicle (UAV) fleets require coordinated optimization of communication, mobility, and energy management. This study evaluates conventional and artificial intelligence (AI)-enabled cross-layer protocol suites in a common five-hour simulation of 7,500 UAVs, 240 base stations, and 40 vertiports across 144 km². They include AODV-based conventional routing, reinforcement learning-based UAV-AI-EDRL, federated ANN-PID-based UAV-AI-FHAP, reduced learning baselines, and a composite. Methodological contributions comprise a mission-constrained cross-layer objective, CMA-ES calibration benchmarks, physics-grounded training, paired evaluation over six environmental realizations, and packet-level NS-3 checks. In fleet-scale runs, proactive routing reduced mean end-to-end delay from 66.0 to 25.3 ms (61.6%), an advantage NS-3 reproduced at every transmit power only with pre-resolved link-layer addresses. UAV-AI-EDRL reduced power by 8.7% and near-miss events by 18.1% on the principal training seed, although safety gains varied across seeds. UAV-AI-FHAP reduced power by 4.6% and near-miss events by 30.9%, but completed 15.9% fewer missions because a charging threshold calibrated on a reduced-area oracle interrupted long delivery legs; geometry-preserving recalibration reduced the deficit to 2.2%, and a leg-aware charging rule produced a 13.9% gain (2.3% if the conventional suite also uses it). Ablations with equivalence tests attributed the gains to proactive forwarding, throttling, spread control, and reactive avoidance, finding no measurable suite-level benefit from DQN, PPO, path planning, ANN-PID, or federated aggregation; suites without learned power control preserved superior raw radio metrics. These findings, limited to the simulated conditions, shifts, and fleet sizes, reveal trade-offs among latency, energy, safety, and mission completion, emphasizing operationally weighted selection and field validation.

1. INTRODUCTION

In urban delivery, unmanned aerial vehicles (UAVs) are currently being used for many applications, such as infrastructure inspection, traffic monitoring, and aerial sensing, posing challenges of communication reliability, trajectory coordinating, and energy management [1]. In the last few years, artificial intelligence (AI) with the specific focus of reinforcement learning (RL) and deep reinforcement learning (DRL) have facilitated the adaptive design of UAV communication and mobility optimization [2–4]. But, the management of wireless resources, routing, flight trajectory, and battery consumption need to be coordinated in a dense network of UAVs operating in dynamic environment. The interdependencies between these

*Corresponding author. Email: Randa.alshaibani@uobabylon.edu.iq

requirements render the optimization of the cross-layer aspects as a vital requirement for ensuring reliable and sustainable fleet operations.

Existing routing protocols like Ad Hoc On-Demand Distance Vector (AODV) do not work well in FANETs with frequent changes in the network topology and long route discovery delays [5]. In the context of learning-based routing, Q-learning algorithms and multi-agent DRL algorithms have been proven to achieve better packet delivery, communication latency, trajectory optimization, and resource allocation [6–11]. Meanwhile, federated learning (FL) has been introduced as an emerging solution to coordinate UAVs without sharing raw operational data, which can be applied for collaborative model training across different devices [12–15]. Implementing hybrid architectures combining learning-based adaptation with classic routing and control mechanisms provides further flexibility, and maintains the operational properties of the existing controllers [4, 16]. However, incorporation of these techniques in a single, scalable system for UAV management is still difficult.

In spite of these advances there are still key research questions that have not yet been answered. First, previous research mainly focuses on optimizing the individual layers in the network, such as the physical layer of communication, the medium access control layer, the routing layer, the trajectory planning layer, or the energy management layer, without adequately accounting for the coupling among physical-layer communication, medium access, routing, trajectory planning and energy management [2, 17]. Second, most hybrid UAV controllers rely on manual tuning of many parameters such as transmission-power limits, control thresholds and charging policies, and systematic EIL tuning is still not well understood [18, 19]. Third, learning-based solutions often use simplified versions of the simulated environments or even the reward functions themselves, which are not representative of the actual physical models used in evaluation. Moreover, extensive comparisons between integrated RL, FL and hybrid control architectures under the same operation conditions are still very limited [2, 4, 6, 12, 16]. These constraints emphasize the requirement for consistent evaluation framework that takes cross-layer interactions, reproducibility and mission-level performance into account.

In response to these, in this study, a physics-inspired cross-layer assessment framework for dense urban UAV delivery networks is proposed. Seven protocol suites are explored such as a traditional AODV-based architecture (UAV-Conventional), a federated artificial neural network–proportional-integral-derivative architecture (UAV-AI-FHAP), three reduced learning-based reference architectures, and a composite architecture (UAV-AI-EDRL). The assessment uses 7,500 UAVs in five simulated hours using six independent environmental realizations and common random numbers to make the comparisons. Four mission-validity constraints are used to formulate a unified optimization objective combining decisions for physical-layer, medium-access, routing, trajectory, and energy-management. Furthermore, ten policy parameters are learned with covariance matrix adaptation evolution strategy (CMA-ES) and the learning components are trained using the Markov decision processes whose state transitions are based on the physical models used for the evaluation.

The proposed framework will utilize complementary NS-3 packet-level testing of the routing abstraction and will also include fleet-scale Python/NumPy simulations. It is not in the development of algorithms, but in systematically combining and properly testing existing learning and control algorithms, that it makes its main contribution. Specifically, the study provides an objective that is reproducible for all UAV protocol suites, defines a mission-dependent cross-layer objective, studies automatic policy calibration compared with conventional parameter-selection approaches, and uses physics-grounded training using reference policies and controlled experiments to evaluate. Ablation studies at the component level, paired statistical analysis, practical equivalence testing and Pareto analysis are used to separate out the roles of individual mechanisms from the global architectures. Special focus is placed on how each of these improvements in communication latency, energy efficiency, and collision avoidance is achieved and an analysis of what are the limitations on charging policy and how effective is the use of leg-aware energy management. These analyses complement each other and offer a body of evidence for the trade-offs in operating selection of learning-based and hybrid control architectures for large-scale UAV networks.

2. METHODOLOGY

2.1 Scenario and Configuration

The scenario models a dense urban delivery deployment on a 12 km × 12 km square (144 km²) populated by 7,500 delivery UAVs of the 1.4 kg class cruising at 15 m·s⁻¹, 240 terrestrial base stations, and 40 vertiport charging stations; the infrastructure density was fixed from the coverage and traffic requirements of the served area, the criterion used when a wireless implementation is selected for an institutional environment [20]. The fleet flies in three cruise-altitude bands at 50, 80, and 110 m, and each aircraft follows a stochastic duty cycle with an in-flight fraction between 0.50 and 0.70, re-rolled every 300 s, with off-duty aircraft split between landed (60%) and hovering (40%) states. The simulation horizon is 18,000 s (five hours) advanced in 1,800 steps of $\Delta t=10$ s. Every suite runs on the same scenario realization: fleet layout, base-station and vertiport positions, weather trajectory, fading draws, duty rolls, and mission targets all follow one environment seed, and the learned components come from one training seed, so that suites are compared with common random numbers. The

principal comparison repeats this design over six environment seeds; Table 1 lists the frozen configuration, and Fig. 1 shows the spatial layout and the altitude-band assignment.

TABLE I. SCENARIO AND CONFIGURATION PARAMETERS FROZEN ACROSS ALL PROTOCOL SUITES, AND THE NS-3 PACKET-LEVEL SCENARIO.

Parameter	Value
Number of UAVs (N)	7,500
Number of base stations	240
Vertiport charging stations	40
Deployment area	12 km $\times$ 12 km (144 km ²)
UAV density / base-station density	52.1 km ⁻² / 1.67 km ⁻²
Simulation horizon T / time step Δt	18,000 s (5 h) / 10 s (1,800 steps)
Cruise-altitude bands	{50, 80, 110} m
Duty cycle (in-flight fraction)	0.50 to 0.70, re-rolled every 300 s
UAV mass / cruise speed / maximum speed	1.4 kg / 15 m·s ⁻¹ / 22 m·s ⁻¹
Battery capacity / charging rate	120 Wh / 12% of capacity per minute at a vertiport
Hover / propulsion / radio power	180 W / 140 W / 6 W
Carrier frequency / bandwidth / MAC channels	3.0 GHz / 20 MHz / 4 orthogonal channels
Nominal transmit power / transmit-power tiers	23 dBm / {17, 20, 23, 26} dBm
Noise floor / Rician K -factor / urban NLOS excess	-95 dBm / 6 dB / 8 dB
Neighbour-density radius / same-channel saturation count	200 m / 3 neighbours
Mission generator	next target uniform in the inner 85% of the area, arrival radius 150 m
Mission gates (PDR / SOC / near-miss / HC)	PDR _{30s} $\geq$ 0.40; SOC $\geq$ 0.10; no near miss within 60 s; HC composite $<$ 0.85
Environment seeds (principal comparison)	11, 12, 13, 14, 15, 16
Training seeds (learned components)	1 (principal), 2 and 3 (variance decomposition)
NS-3 packet-level scenario	
Simulator	NS-3 3.43 built from source
Nodes	60 UAV nodes; 9 fixed sinks on 25 m masts (jittered grid)
Area / altitude range	1,200 m $\times$ 1,200 m; 50 to 110 m
Mobility	3-D random waypoint, 15 m·s ⁻¹ , no pause
Link / MAC	IEEE 802.11a ad hoc, fixed OFDM 6 Mb/s rate
Propagation	log-distance (exponent 2.4, 46.7 dB reference loss at 1 m) with Nakagami fading ($m = 1.5$)
Traffic	64-byte UDP packets at 2 packets per second from every UAV to its nearest sink
Forwarding schemes	AODV (reactive); OLSR with default timers (HELLO 2 s, TC 5 s) and with fast timers (HELLO 0.5 s, TC 1 s); zero-overhead proactive forwarding (minimum-hop routes recomputed every 1 s from true positions over links with mean received power $\geq$ -84 dBm, no control packets)
Address resolution	standard ARP, or ARP caches populated before the run

Parameter	Value
Delivery denominator	packets generated by the sources, counted before the socket
Transmit-power sweep	11, 14, 17, 20, 23, 26 dBm
Duration / runs	45 s per run; run numbers 1 to 3

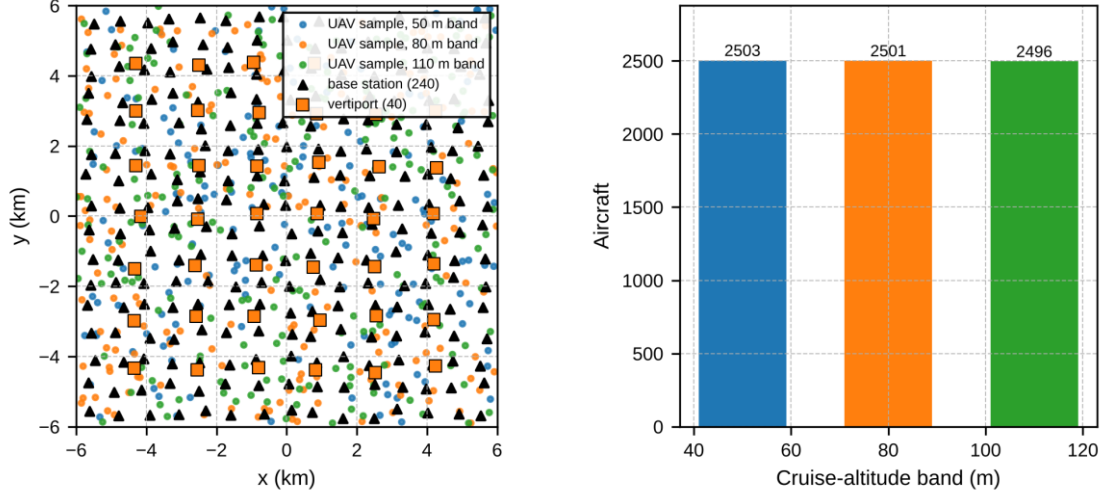


Fig. 1. Scenario layout: a 600-UAV sample of the fleet with base stations and vertiports (left) and the altitude-band assignment (right).

2.2 Wireless Channel and Link Model

Every UAV-to-base-station link uses an air-to-ground channel with a log-distance path-loss term, an elevation-dependent line-of-sight (LOS) probability following the 3GPP aerial-vehicle channel study [21], and Rician small-scale fading. The same functions serve the fleet-scale evaluation and the training environment which is the mechanism that grounds training rewards in evaluation physics. Path loss is computed from the ground-projected distance d and the aircraft height h above the base-station mast as

$$PL(d, h) = 20\log_{10}\sqrt{d^2 + h^2} + 20\log_{10}(f_c) - 147.55 + (1 - p_{\text{LOS}}(d, h))\eta_{\text{NLOS}} \quad (1)$$

where f_c is the carrier frequency in Hz (3×10^9 Hz here), η_{NLOS} is the additional urban non-LOS shadowing (8 dB), and the LOS probability follows the elevation angle θ_{elev} through

$$p_{\text{LOS}}(d, h) = \frac{1}{1 + 0.6\exp(-0.10(\theta_{\text{elev}} + 8))}, \quad \theta_{\text{elev}} = \arctan(h/d) \quad (2)$$

Here, h is the height of each aircraft, so altitude-band decisions change the link budget. Small-scale fading draws one Rician sample per link and step,

$$F_{\text{dB}} = 20\log_{10}|X + jY|, \quad X \sim \mathcal{N}(s, \sigma^2), Y \sim \mathcal{N}(0, \sigma^2) \quad (3)$$

with $s = \sqrt{K/(K+1)}$, $\sigma^2 = 1/(2(K+1))$, and K the linear-scale Rician factor ($10^{6/10} \approx 3.98$). The received signal strength combines the transmit power P_{TX} (nominal 23 dBm; the AI suites emit a per-UAV value), base-station sectoral gain $G_{\text{BS}} = 12$ dBi, and UAV omnidirectional gain $G_{\text{UAV}} = 2$ dBi:

$$\text{RSSI} = P_{\text{TX}} + G_{\text{BS}} + G_{\text{UAV}} - PL(d, h) + F_{\text{dB}} \quad (4)$$

The signal-to-interference-plus-noise ratio converts the RSSI to linear signal power S and divides by aggregate co-channel interference I plus thermal noise N (linear scale of -95 dBm), floored at $\varepsilon = 10^{-12}$ to avoid logarithmic singularities:

$$\text{SINR} = 10\log_{10} \frac{S}{\max(I + N, \varepsilon)} \quad (5)$$

The medium is divided into four orthogonal MAC channels; the conventional suite assigns each aircraft a static random channel, and the EDRL suite selects channels through its DQN. The interference term aggregates a density component driven by the number ρ_c of same-channel neighbours with active radios inside a 200 m radius, scaled by the mean relative transmit power $\bar{p}$ of those neighbours (so a neighbourhood that trims its transmit power lowers the interference it imposes), and the environmental interference, side-lobe, and multipath severities:

$$I_{\text{dBm}} = N_{\text{dBm}} + (I_{\text{sat}} - I_{\text{floor}}) \min(\rho_c / \rho_{\text{sat}}, 1) + 10 \log_{10} \bar{p} + 6\psi_{\text{interf}} + 4\psi_{\text{side}} + 3\psi_{\text{multi}} \quad (6)$$

With $I_{\text{floor}} = 0$ dB, $I_{\text{sat}} = 9$ dB, and $\rho_{\text{sat}} = 3$ same-channel neighbours, the saturation point lies at three same-channel neighbours, the value reached under uniform random channel assignment at the scenario density. The density term represents congestion that grows with the number of contending radios, as predicted by data-driven congestion models for 5G links based on the offered load [22]. Aircraft on a charging pad and landed aircraft do not contribute to ρ_c . Spectral efficiency follows the Shannon bound, capped at the modulation-and-coding ceiling of $8 \text{ bps} \cdot \text{Hz}^{-1}$:

$$\eta_{\text{SE}} = \min(\log_2(1 + 10^{\text{SINR}/10}), 8) \quad (7)$$

2.3 Environmental Hard-Condition Model

Environmental stressors are modeled as eight time-varying severities $\psi(t) \in [0,1]^8$, covering icing, wind, rain, handover pressure, interference, sidelobe degradation, beyond-visual-line-of-sight (BVLoS) conditions, and multipath. Each severity follows a smoothed diurnal cycle with a stochastic gust term, so weather fronts traverse the deployment during the five-hour window:

$$\psi_k(t) = \text{clip}(a_k + b_k \sin(2\pi t/T + \varphi_k) + \xi_k(t), 0, 1) \quad (8)$$

where a_k and b_k are the type- k baseline and amplitude ($b_{\text{wind}} = 0.9$ for wind), φ_k is a type-specific phase offset (1.1 rad for handover), and $\xi_k(t)$ is a Gaussian gust term with standard deviation 0.1 (0.15 for the diurnal trend). The gust terms are drawn from the environment-seed stream, so every suite experiences the same weather trajectory. The composite severity used by the adaptive hand-off gate of averages the eight components:

$$S(t) = \frac{1}{8} \sum_{k=1}^8 \psi_k(t) \quad (9)$$

Wind and icing act more strongly at higher altitudes. A height factor

$$h_z(z) = 0.8 + 0.4 \frac{z - 50}{60}, \quad z \in [50, 110] \text{ m} \quad (10)$$

scales the wind and icing severities seen by an aircraft at height z , so that the 50 m band is 20% more sheltered and the 110 m band 20% more exposed than the 80 m reference; this gives the altitude-band actions of the path planner a physical trade-off between link quality (Eq. (2)) and energy (Eq. (11)).

2.4 Energy and Mobility Model

Instantaneous per-UAV power combines hover, propulsion, drag, ice-loading, and radio terms. Wind raises effective drag through $k_{\text{wind}} = 1 + 0.6 \cdot h_z \cdot \psi_{\text{wind}}$, icing degrades the propeller through $k_{\text{ice}} = 1 + 0.4 \cdot h_z \cdot \psi_{\text{icing}}$, and the propeller efficiency $\eta = 0.70 - 0.20 \cdot \psi_{\text{icing}} - 0.10 \cdot \psi_{\text{wind}}$ is a weather-driven input identical for all suites, further derated by $1 - 0.3 \cdot h_z \cdot \psi_{\text{icing}}$ with a floor of 0.40:

$$P_i(t) = \frac{(P_{\text{hover}} + P_{\text{prop}}(v_i/v_c)^3)k_{\text{wind}}k_{\text{ice}}}{\eta} + P_{\text{radio}} \quad (11)$$

where v_i is the instantaneous airspeed, $v_c = 15 \text{ m} \cdot \text{s}^{-1}$ the cruise reference, and P_{hover} , P_{prop} , and P_{radio} equal 180, 140, and 6 W for the simulated 1.4 kg delivery airframe. No term of Eq. (11) depends on the identity or training quality of a policy, so all energy differences between suites arise from commanded speed (through the cubic term), velocity jitter, altitude, and flight patterns. Battery state of charge integrates the draw at each step:

$$E_i(t + \Delta t) = E_i(t) - P_i(t) \frac{\Delta t}{3600} \text{ [Wh]} \quad (12)$$

Mobility follows a three-dimensional Gauss-Markov process (memory 0.40) that steers each aircraft toward its target at the commanded speed, with process noise of $0.8 \text{ m}\cdot\text{s}^{-1}$ per axis scaled by a per-aircraft noise factor, wind-driven drift, and reflection at the area boundary. Depletion of E_i drives the charging gate: an aircraft whose state of charge (SOC) falls below its charging threshold diverts to a vertiport, with the threshold fixed at 0.20 for the conventional and EDRL suites and set per zone by federated consensus in the FHAP suite. Mission targets are drawn by one generator shared by all suites (Table 1), so mission schedules differ between suites only through the aircraft's own progress. also evaluates a leg-aware variant of the charging gate. After each delivery, an aircraft starts its next leg only if its remaining energy covers the predicted energy of that leg plus the reach from the leg's target to the nearest vertiport, with a 15% margin and a reserve of 0.12 of capacity (above the 0.10 reserve of the mission gate); otherwise, it diverts to charge at once. The energy rate per kilometre is either the per-zone prediction of the federated model. The cruise power in Eq. (11) under the current weather conditions, and a safety trigger diverts an aircraft in flight only when its energy falls below the reserve plus the reach to the nearest vertiport. A fixed SOC threshold diverts aircraft wherever they are on their leg, whereas the leg-aware rule abandons a leg only through the safety trigger.

2.5 Routing protocols: AODV and IQMR

Two routing behaviours are represented, both using the same per-hop loss map, so that routing differences arise solely from the forwarding mechanism. Per-hop loss rises as the link SINR falls and with the side-lobe and multipath severities, $\ell(\text{SINR}) = \text{clip}\left(1/\left(1 + \exp((\text{SINR} - 6)/3)\right) + 0.15\psi_{\text{side}} + 0.10\psi_{\text{multi}}, 0, 0.95\right)$, delivery follows $(1 - \ell)^{\text{hops}}$, and the per-hop queueing cost is $12 + 8\psi_{\text{interf}} + 6\psi_{\text{handover}}$ ms. The hop count is 1, 2, or 3 according to the distance to the serving base station (thresholds 800 m and 3,000 m), the shortest-path structure that Boolean-algebra formulations compute for networks of this kind [23]; the SINR-ranked candidate set and the progress flag of IQMR (below) follow the topology-sensing rationale of graph-based routing for wireless ad hoc networks [24]. UAV-Conventional, UAV-FL, and UAV-AI-FHAP run the reactive AODV protocol [5], abstracted statistically: the loss map is evaluated at the aircraft's own link SINR, and delay adds a route-discovery term of $60 \cdot \exp(-\text{PDR}) + 25$ ms to the per-hop cost. UAV-QR, UAV-AI-EDRL, and the composite suite run IQMR, an improved Q-learning multi-hop routing protocol with $Q(\lambda)$ eligibility traces. The IQMR forwarding decision is learned: at every step, each aircraft ranks its eight nearest active neighbours by their link SINR, labels each candidate with a progress flag (candidate closer to its base station than the forwarder), and a tabular $Q(\lambda)$ policy over 64 states ($8 \text{ SINR bins} \times 4 \text{ same-channel-density bins} \times 2 \text{ severity bins}$) selects a rank. The loss map is then evaluated at the chosen hop's SINR, the effective hop count is reduced by the progress flag, and the delay replaces the discovery term by an 8 ms proactive table lookup. The learner uses Watkins's $Q(\lambda)$ with learning rate $\beta = 0.3$, discount $\gamma = 0.85$, trace decay $\lambda = 0.7$, and traces reset after exploratory actions:

$$\delta_t = r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \quad (13)$$

$$e(s, a) \leftarrow \gamma \lambda e(s, a), \quad e(s_t, a_t) \leftarrow e(s_t, a_t) + 1 \quad (14)$$

$$Q(s, a) \leftarrow Q(s, a) + \beta \delta_t e(s, a) \quad \forall (s, a) \quad (15)$$

2.6 Learned Components of the Suites

The deep block contains a deep Q-network (DQN) [25] selecting among 16 discrete actions ($4 \text{ channels} \times 4 \text{ transmit-power tiers}$), a proximal policy optimization (PPO) actor-critic [26] emitting a continuous per-UAV transmit-power refinement within the autotuned band, and a multi-agent actor-critic swarm controller trained with multi-agent PPO (MAPPO) [27], with parameter-shared actors acting on local observations and a centralized critic on the joint observation, emitting per-UAV spread magnitudes bounded by the autotuned cap. The classical block contains the IQMR router and a tabular Q-learning path planner [28] choosing, per aircraft, a base-station-ward heading blend among five levels and an altitude-band action among three. UAV-AI-FHAP retains AODV at the routing layer and learns fleet-level energy governance: federated averaging (FedAvg) [29] trains a 32-parameter linear model that predicts energy per kilometre and link quality from local features and is decoded into per-zone state-of-charge charging thresholds and cruise factors over 16 zones, in the architectural spirit of federated FANET learning [12], and a hybrid ANN-PID controller schedules per-UAV proportional, integral, and derivative gains through a $4 \times 8 \times 3$ multilayer perceptron trained by an evolution strategy on the closed loop.

Table 2 summarizes the seven learned components, and Table 3 gives their complete architecture and training specifications. The per-component inputs are compact sets of physics-derived attributes rather than exhaustive telemetry, consistent with comparative evidence from supervised and unsupervised learners that pruning weakly correlated input attributes improves predictive accuracy [30]. The three reduced reference suites reuse the same trained components: UAV-QR adds only learned IQMR routing to the conventional suite; UAV-DRL runs the deep block without the classical specialists; and UAV-FL runs the federated energy layer with a fixed-gain PID. The composite suite runs the EDRL and FHAP stacks together.

TABLE II. THE SEVEN LEARNED COMPONENTS, THEIR SUITES, INPUTS, DECISION SPACES, AND TRAINED QUANTITIES.

Component	Suite (block)	Input (per aircraft unless stated)	Decision space	Trained quantity (parameters)
DQN	EDRL (deep)	11-d: SINR, RSSI, SOC, density, severity, altitude, airspeed, occupancy of the 4 channels	16 discrete actions: 4 channels $\times$ 4 TX tiers	11-64-64-16 ReLU MLP (5,968)
PPO	EDRL (deep)	6-d: SINR, density, SOC, severity, airspeed, altitude	Continuous TX offset in $[-3, +3]$ dB, clipped to $[e2_floor, e2_ceil]$	Actor 6-64-64-1 tanh, critic 6-64-64-1, log-std (9,347)
MADRL (MAPPO)	EDRL (deep)	4-d local: SINR, density, SOC, severity; critic sees the 80-d joint observation of 20 cohorts	Spread magnitude $s \in (0,1)$ applied to the repulsion direction, cap $e1$	Shared actor 4-64-64-1, centralized critic 80-64-64-1 (13,955)
IQMR $Q(\lambda)$	EDRL (classical)	64 states: 8 SINR $\times$ 4 same-channel density $\times$ 2 severity bins	8 ranked next-hop candidates	64 $\times$ 8 Q-table (512)
RL path planner	EDRL (classical)	40 states: 10 severity $\times$ 4 SINR bins; 10 severity states for altitude	5 heading-blend levels; 3 altitude actions	40 $\times$ 5 and 10 $\times$ 3 Q-tables (230)
FedAvg	FHAP	16 features: density, SINR, wind, SOC, distance, altitude, speed, 8 severities, bias	Per-zone (SOC threshold, cruise factor) over 16 zones	Linear 16 $\times$ 2 global model (32)
ANN-PID	FHAP	4-d: wind, density, tracking error, SOC	Per-UAV (K_p, K_i, K_d)	4 $\times$ 8 $\times$ 3 tanh MLP (67)

TABLE III. ARCHITECTURE AND TRAINING HYPERPARAMETERS OF EVERY LEARNED COMPONENT (ALL VALUES AS IMPLEMENTED).

Component	Algorithm and update rule	Optimizer, rates, budgets	Exploration, buffers, schedules
DQN	Huber TD loss ($\delta = 1$) on $y = r + \gamma \max Q_{\text{target}}(s'; \cdot)$; hard target update every 250 gradient steps; 4 gradient steps per environment step	Adam, learning rate 5×10^{-4} , batch 64, $\gamma = 0.5$ (near-memoryless channel decision), gradient-norm clip 10	ϵ -greedy, $1.0 \rightarrow 0.05$ linearly over 60% of the 2,520 environment steps; replay 100,000 transitions; 42 episodes (6 passes over 7 training instances); a random 25% of aircraft re-decide each step
PPO	Gaussian policy, GAE(λ) advantages, clipped surrogate ($\epsilon = 0.2$), value loss (coefficient 0.5), entropy bonus (0.01); 4 epochs over minibatches of 512	Adam 3×10^{-4} , $\gamma = 0.95$, $\lambda_{\text{GAE}} = 0.95$, gradient-norm clip 0.5	Initial standard deviation 0.5 (floor 0.05); 40 iterations $\times$ 12,000 samples (60 steps $\times$ 200 aircraft); validation every 5 iterations
MADRL	parameter-shared Gaussian actors, centralized value on the joint observation, same clipped-surrogate update ($\epsilon = 0.2$, 4 epochs, minibatch 256, entropy 0.001)	Adam 3×10^{-4} , $\gamma = 0.95$, $\lambda_{\text{GAE}} = 0.95$, gradient-norm clip 0.5	60 iterations $\times$ 1,200 samples (60 steps $\times$ 20 cohorts); shared reward = mean $\eta_{\text{SE}}/8 - 0.15 \text{ var}(\eta_{\text{SE}})/8$ - near-miss fraction - 0.01 mean spread
IQMR	Watkins $Q(\lambda)$, Eqs. (13)-(15); reward = per-hop delivery of the chosen hop - 0.10(1 - progress)	$\beta = 0.3$, $\gamma = 0.85$, $\lambda = 0.7$	$\epsilon = 0.5 \times 0.85^{\text{episode}}$ (floor 0.05); 21 episodes (3 passes), all 200 aircraft per step
RL path planner	Tabular Q-learning; reward = 0.5 PDR - 0.5 P/400 W + 0.3 progress - 0.02 [band change]	$\alpha = 0.2$, $\gamma = 0.9$	$\epsilon = 0.5 \times 0.85^{\text{episode}}$ (floor 0.05); 21 episodes

Component	Algorithm and update rule	Optimizer, rates, budgets	Exploration, buffers, schedules
FedAvg	48 clients (3 per zone, non-IID by zone), 2 local SGD epochs, batch 32, sample-size-weighted averaging, 70% participation per round	SGD 0.05; 50 rounds; standardized targets (energy per km, delivery ratio)	Round selected on validation MSE; references: centralized SGD with the same epochs, local-only models per zone
ANN-PID	(8, 16) evolution strategy on the 67 network weights; cost = ISE + 0.05 overshoot + $10^{-3}\sum u^2$ over 300-step episodes with set-point changes every 50 steps and wind disturbance from the physics sampler	$\sigma = 0.10$, 100 generations, 3 training episodes per evaluation	References: fixed-gain PID tuned by the same strategy and budget over (K_p, K_i, K_d) ; default gains (1.5, 0.2, 0.1)

2.7 Cross-Layer Co-Design Objective

The first methodological element formalizes one multi-objective optimization spanning the five cooperating layers: PHY (transmit power), MAC (channel), routing (IQMR or AODV), trajectory (spread, detour, heading, damping), and energy (throttle, charging threshold, cruise factor, consensus). The objective is

$$J(\theta, \pi) = \alpha(\bar{E} - 1) + \beta(\bar{D} - 1) - \gamma(\bar{R}_{\text{PDR}} - 1) - \delta(\bar{M} - 1) \quad (16)$$

where $\theta \in \mathbb{R}^{10}$ is the decision vector of π identifies the policy class (a suite, a reference, or an ablation), and the four normalized terms divide mean power, mean end-to-end delay, packet delivery ratio, and gated missions per UAV by the conventional suite's values at the same scale, averaged over the same environment seeds ($E_{\text{ref}} = 407.4$ W, $D_{\text{ref}} = 66.0$ ms, $R_{\text{ref}} = 0.931$, $M_{\text{ref}} = 3.25$ at full scale; oracle-scale values. The weights $\alpha = \beta = 0.30$ and $\gamma = \delta = 0.20$ sum to one; because each normalized term is centred at its reference value, $J = 0$ marks parity with the conventional suite and negative values mark improvement on the weighted composite (without the centring, parity would sit at 0.20). The weights express an operator preference that treats energy and delay as the primary costs and delivery and mission count as secondary benefits; because the ranking of configurations could depend on this choice, recomputes J under five alternative weight sets, and the Pareto analyses weight-free. Four engineering constraints define the admissible region, evaluated per UAV at mission-arrival time; violation of any one invalidates the mission:

$$g_1: \text{SOC}_{\text{arrival}} \geq 0.10 \text{ (energy reserve)} \quad (17a)$$

$$g_2: \text{PDR}_{30s} \geq 0.40 \text{ (reception)} \quad (17b)$$

$$g_3: \text{no near miss within 60 s (safety)} \quad (17c)$$

$$g_4: S_{\text{HC}} \leq 0.85 \text{ (weather and communication envelope)} \quad (17d)$$

A near miss is a three-dimensional separation of less than 50 m between two en-route aircraft; aircraft in charging mode or within the 100 m terminal area of a vertiport are excluded because vertiport queuing is a managed terminal operation. A Pareto tracker maintains the set of non-dominated 4-tuples (E, D, R, M) across suites and ablations, where configuration q dominates configuration p if q is at least as good on all four objectives and strictly better on at least one.

2.8 Calibration of the Policy Coefficients

The second element replaces engineering judgement for ten policy coefficients that would otherwise be fixed by hand: the MADRL spread cap, the PPO transmit-power floor and ceiling, the severity-gate base of the hand-off rule, the replanning cooldown and its minimum-SOC gate, the FedAvg charging-threshold centre and span, the cruise-factor span, and the consensus-pressure decay. Because the first six coefficients influence only the EDRL stack and the last four only the FHAP stack, they are calibrated as two independent problems of dimensions six and four, respectively. The optimizer is the standard $(\mu/\mu_w, \lambda)$ -CMA-ES of Hansen [31] with rank- μ covariance update, evolution paths, and cumulative step-size adaptation, run in normalized box coordinates $[0,1]^n$ with clipping, population $\lambda = 8$, parent number $\mu = 4$, initial step $\sigma_0 = 0.3$, and 12 generations, so 96 oracle evaluations per run:

$$\theta_k = m_g + \sigma_g \mathbf{B}_g \mathbf{D}_g \mathbf{z}_k, \quad \mathbf{z}_k \sim \mathcal{N}(0, \mathbf{I}), \quad k = 1, \dots, \lambda \quad (18a)$$

$$m_{g+1} = m_g + \sigma_g \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}, \quad \mathbf{C}_{g+1} = (1 - c_1 - c_\mu) \mathbf{C}_g + c_1 \mathbf{p}_c \mathbf{p}_c^\top + c_\mu \sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda} \mathbf{y}_{i:\lambda}^\top \quad (18b)$$

where m_g , σ_g , and $\mathbf{C}_g = \mathbf{B}_g \mathbf{D}_g^2 \mathbf{B}_g^\top$ are the mean, step size, and covariance of the search distribution, $\mathbf{y}_{i:\lambda} = (\theta_{i:\lambda} - m_g) / \sigma_g$ are the ranked steps, w_i are the log-rank recombination weights, and $\mathbf{p}_c$ and the step-size path follow the default learning rates of [32]. Each candidate is evaluated by a reduced-cost oracle: a density-preserving mini-scenario with 400 UAVs, 13 base stations, 2 vertiports, a 2.77 km square, and 1,800 s of operation under the full physics of Sections 4.2 to 4.5, run with one fixed environment seed so that all candidates see identical realizations (common random numbers). The oracle returns the four-headline metrics, which Eq. (16) reduces to J with oracle-scale normalizers measured on the conventional suite under the same seed, so that the oracle objective and the full-scale objective share one definition. Three controls bound the value of the optimizer: the hand-set defaults are evaluated on the same oracle; a random search with the same budget of 96 evaluations is run for each problem; and both CMA-ES and random search are repeated with three optimizer seeds. The vector reported as θ^* is the best CMA-ES solution by oracle J (selection uses the oracle only), and reports convergence, the stability of the calibrated coordinates across seeds, and the transfer of oracle rankings to full scale. Because the oracle shrinks the service area to preserve aircraft density, it also shortens delivery legs (1.2 km against 5.1 km at full scale); shows that this distortion explains the failed transfer of the FHAP charging centre and recalibrates the four FHAP coefficients on a geometry-preserving oracle (the full 12 km square with 240 base stations and 40 vertiports, 750 aircraft, 5,400 s, and two environment seeds) with the same CMA-ES settings.

2.9 Physics-Grounded Training and Evaluation Protocol

The third element closes the loop between training and evaluation. A protocol that trains each component on a few hundred transitions from a single rollout, with a random transition-level split, cannot establish generalization. The present protocol trains each component on a lightweight physics-grounded MDP whose state transitions reuse the path loss, fading, interference, SINR, energy, and routing functions of Eqs. (1) to (15), and it organizes the data by fleet instance. Ten mini-fleet instances of 200 aircraft on a density-preserving 2.25 km square with 8 base stations are generated from independent seeds; instances 0 to 6 are used for training, instance 7 for validation (checkpoint selection and the hyperparameter choices of Table 3), and instances 8 and 9 for the single final test evaluation. Because the independent unit is the fleet instance (with different layouts, weather trajectories, and fading streams), transitions from one trajectory never straddle the training and evaluation sets. Table 4 records the transition counts per component; in total 1,198,800 physics-grounded transitions were used for training. Features and targets are standardized on the training instances before learning, the pre-processing and feature-extraction stage that guideline-based signal classification pipelines apply before the learner [33]. For every component the held-out test score is reported together with a random policy and a reference rule (least-loaded channel with the lowest power tier for the DQN, nominal power for PPO, no spreading for MADRL, best-SINR next hop for IQMR, no detour for the path planner, centralized and local-only training for FedAvg, and an equally tuned fixed-gain PID for ANN-PID), so that the value added by learning is measured. A synthetic-reward control trains the same architectures on random-walk states with sinusoidal or constant-target rewards unrelated to the physics and deploys the resulting components in the same suites, which tests the physics-grounded claim.

TABLE IV. PHYSICS-GROUNDED TRAINING DATA: TRANSITIONS PER COMPONENT, INSTANCE-LEVEL SPLITS, TRAINING TIME ON THE REFERENCE HARDWARE, AND HELD-OUT EVALUATION.

Component	Transition (arrays)	Training transitions	Validation	Test	Training time (s)	Held-out test score (learned / random / reference rule)
DQN	state (11), action (1), reward (1), next_state (11), instance, t	126,000	3,000 per check	6,000	27	0.856 / 0.773 / 0.933
PPO	state (6), action (1), reward (1), instance, t	480,000	12,000 per check	24,000	35	0.874 / n/a / 0.825
MADRL	joint_obs (80), joint_act (20),	72,000	1,200 per check	2,400	20	0.671 / 0.678 / 0.686

Component	Transition (arrays)	Training transitions	Validation	Test	Training time (s)	Held-out test score (learned / random / reference rule)
	joint_reward (1), instance, t					
IQMR	state, action, reward, next_state (indices), instance, t	252,000	12,000 per check	24,000	7	0.916 / 0.882 / 0.917
RLP	state, action, alt_action, reward, next_state, instance, t	252,000	12,000 per check	24,000	4	-0.184 / -0.273 / -0.271
FedAvg	features (16), targets (2), zone, instance	16,800	2,400	4,800	2	MSE 0.401 / mean predictor 0.996 / centralized 0.411, local-only 0.431
ANN-PID	closed-loop episodes (300 steps) with wind, set-point, error, control	1,440,000 plant steps	4 episodes	8 episodes	126	cost 0.0326 / default PID 0.2221 / tuned fixed PID 0.0307

When the deep block (DQN, PPO, MADRL) and the classical block (IQMR, path planner) coexist in the EDRL suite, an adaptive hand-off threshold decides which block drives the trajectory layer. Two training-quality scalars q_{DRL} and q_{RL} , defined as the normalized validation scores of the two blocks after training (fixed thereafter), modulate the gate around its calibrated base $\tau_0 = e3_severity_base$:

$$\tau(q_{\text{DRL}}, q_{\text{RL}}) = \text{clip}(\tau_0 - 0.10q_{\text{DRL}} + 0.05q_{\text{RL}}, 0.35, 0.75) \quad (19)$$

The gate compares τ with the composite severity $S(t)$ in Eq. (9): a well-trained deep block widens its operating envelope, while a well-trained classical block is brought in earlier as conditions harden. A compliance audit completes the protocol: every policy returns only input-side quantities (channel, transmit power, speed factor, noise factor, lateral drift, heading blend, charging threshold, station choice), and no policy can scale a reported metric; the audit is enforced by the simulation loop, which reads no other policy outputs.

2.10 Experimental Design, Statistics, and Reproducibility

To guarantee paired comparisons, the same fleet configurations, weather conditions, fading processes and mission requirements were used for all the realizations of the experiments, which were carried out using seven UAV protocol suites. Training and calibration of the model were finished prior to the full-scale evaluation, and further ablation, sensitivity and robustness, and charging-policy experiments were carried out to examine the performance differences seen. Analyses added after the main results were considered were clearly defined as post hoc analyses. Communication reliability, energy consumption, mission completion, collision avoidance, and computational efficiency were taken into account when evaluating.

Mean values, standard deviations, 95% confidence interval, paired t-test, Wilcoxon signed-rank test and Holm adjusted significance level were used to assess statistical reliability [34,35]. Practical equivalence tests were conducted with two one-sided tests (TOST) applied, with specific performance margins defined. Python/NumPy was used to implement fleet-scale experiments and for complementary packet-level routing checks, NS-3 version 3.43 was used. Reproducibility was achieved by using fixed random seeds, specifying the configurations of the simulations, training the models, and recording the outputs of the experiments.

3. RESULTS and DISCUSSION

3.1. Training Outcomes, Reference Policies, and Training Controls

Table 4 shows the training results of the 7 components, summarized in Fig. 2. The test return of DQN is 0.856, which is higher than random policy (0.773) and lower than the reference rule (0.933). There was an improvement in performance with PPO compared to nominal power control (0.874 vs. 0.825), but MAPPO performance was worse than the no-spreading baseline (0.671 vs. 0.686). IQMR was found to perform similarly to the best-SINR rule (0.916 vs. 0.917), whereas the path planner showed an improvement in its return when compared to the no-detour baseline (-0.184 vs. -0.271). FedAvg made

a prediction error of 0.401, which is close to centralized training (0.411) and better than local-only training (0.431). Likewise, the cost of the controller was less with ANN-PID than with default PID (0.0326 vs. 0.2221), but not better than an equally fine-tuned fixed-gain PID (0.0307). The results, overall, indicate learning gains at the component level and show that a number of traditional reference strategies are as effective as or more effective than others (Table 5 and Fig. 3).

TABLE V. CLOSED-LOOP CONTROL BENCHMARK ON EIGHT HELD-OUT DISTURBANCE EPISODES (SET-POINT CHANGES EVERY 50 STEPS, WIND FROM THE PHYSICS SAMPLER): ANN-PID GAIN SCHEDULER AGAINST FIXED-GAIN PID REFERENCES.

Controller	ISE	IAE	Overshoot	Settling (steps)	Cost
ANN-PID gain scheduler	0.0150	0.0637	0.316	10.5	0.0326
fixed-gain PID, tuned with the same strategy and budget	0.0105	0.0538	0.358	5.7	0.0307
fixed-gain PID, default gains (1.5, 0.2, 0.1)	0.1273	0.3030	1.875	43.4	0.2221

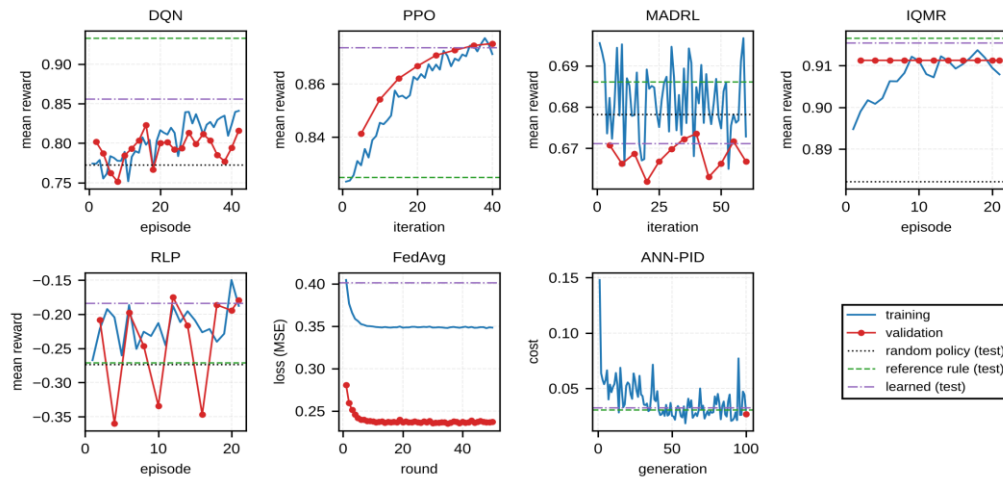


Fig. 2. Training and validation curves of the seven learned components with held-out test, random-policy, and reference-rule scores.

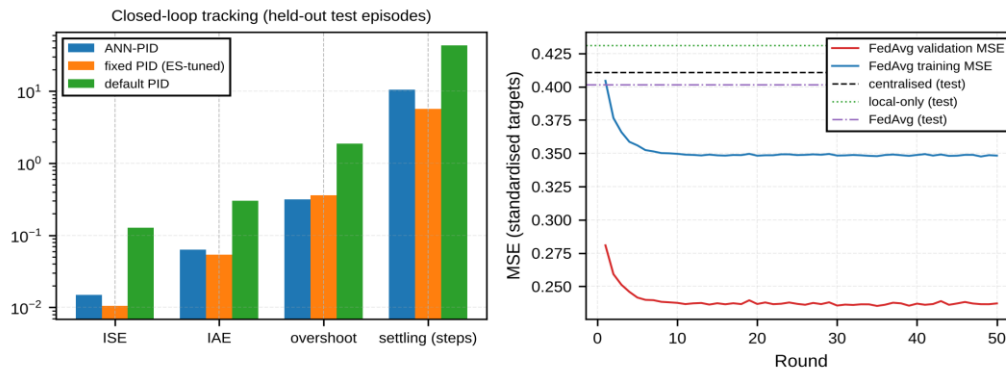


Fig. 3. ANN-PID against tuned and default fixed-gain PID on held-out episodes (left); FedAvg error by round against centralized and local-only training (right).

Two control experiments were carried out in order to investigate the effect of physics-based training and training data volume. Physics-grounded training resulted in better interpretability of held-out evaluations than synthetic-reward training, but did not yield measurable enhancements that were beyond environmental variability. Reducing the training budget to 1% and 10% yielded DQN scores of 0.795 and 0.853 (full: 0.856), PPO scores of 0.839 and 0.872 (full: 0.874), and IQMR scores of 0.900 and 0.909 (full: 0.916). This was not a problem at the full scale performance, with no noticeable difference in seed level variability between these scenarios, suggesting that training beyond 10% of the original budget was not seen to have any significant impact in the evaluated scenarios.

3.2. Calibration of the Coefficients

Table 6 lists the ten coefficients, their bounds, the hand-set defaults, the solutions found by three CMA-ES seeds and three random-search seeds, and the selected θ^* ; Fig. 4 shows the convergence of the best oracle objective against the evaluation budget. For the six EDRL coefficients (Table 7), the defaults scored $J = -0.169$ on the oracle (oracle reference values $E = 428.7$ W, $D = 66.2$ ms, $R = 0.932$, $M = 1.92$), CMA-ES reached -0.268 ± 0.007 over three seeds, and random search reached -0.259 ± 0.001 with the same 96 evaluations. For the four FHAP coefficients, the defaults scored -0.047 , CMA-ES -0.062 ± 0.005 , and random search -0.061 ± 0.005 . The entire calibration required 1,154 oracle evaluations over 24 min.

TABLE VI. THE TEN CALIBRATED COEFFICIENTS: BOUNDS, DEFAULTS, CMA-ES AND RANDOM-SEARCH SOLUTIONS (THREE SEEDS EACH, 96 ORACLE EVALUATIONS PER RUN), AND THE SELECTED θ^* .

Coefficient	Bounds	Default	CMA-ES seeds 1, 2, 3	Random search seeds 1, 2, 3	θ^*
e1_spread_cap ($\text{m}\cdot\text{s}^{-1}$)	[0.60, 2.00]	1.20	1.43, 1.50, 2.00	0.85, 0.76, 2.00	2.00
e2_TX_floor (dBm)	[15.00, 20.00]	17.00	16.35, 19.42, 17.05	16.33, 15.68, 18.84	17.05
e2_TX_ceil (dBm)	[24.00, 30.00]	26.00	24.00, 30.00, 30.00	28.51, 28.52, 26.01	30.00
e3_severity_base	[0.400, 0.700]	0.550	0.671, 0.658, 0.700	0.600, 0.620, 0.683	0.700
e6_replan_cooldown (s)	[30, 180]	90	95, 30, 155	121, 80, 137	155
e6_min_SOC_replan	[0.150, 0.350]	0.250	0.269, 0.262, 0.333	0.170, 0.271, 0.341	0.333
f1_charge_thresh	[0.220, 0.360]	0.280	0.235, 0.360, 0.248	0.316, 0.284, 0.247	0.360
f1_span	[0.020, 0.080]	0.040	0.023, 0.046, 0.041	0.065, 0.048, 0.047	0.046
f1_speed_span	[0.020, 0.080]	0.050	0.052, 0.052, 0.059	0.050, 0.039, 0.053	0.052
f4_pressure_decay	[0.700, 0.950]	0.850	0.926, 0.752, 0.841	0.880, 0.721, 0.769	0.752

TABLE VII. ORACLE-SCALE COMPOSITE OBJECTIVE J (CENTRED, LOWER IS BETTER) OF THE DEFAULTS, CMA-ES, AND RANDOM SEARCH.

Problem	Defaults	CMA-ES (3 seeds)	Random search (3 seeds)	Selected
EDRL (6-D)	-0.1693	-0.2585, -0.2725, -0.2742 (mean -0.2684)	-0.2585, -0.2585, -0.2614 (mean -0.2595)	-0.2742 (seed 3)
FHAP (4-D)	-0.0472	-0.0567, -0.0683, -0.0614 (mean -0.0621)	-0.0659, -0.0547, -0.0619 (mean -0.0608)	-0.0683 (seed 2)

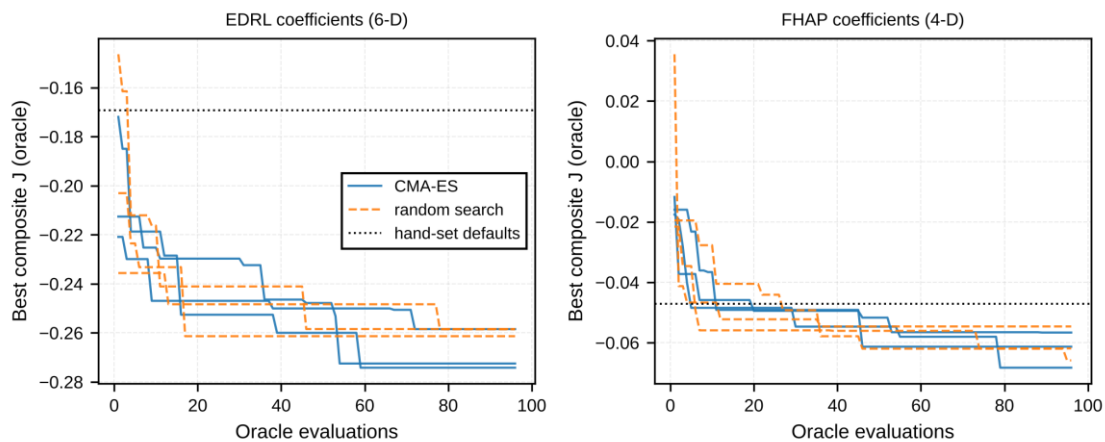


Fig. 4. Calibration convergence of CMA-ES and random search against the defaults for the EDRL and FHAP coefficient subsets.

The EDRL objective was improved by around 0.099 J units with the use of calibration; the CMA-ES performance was only 0.009 J units better than the random search on average, which is not enough to conclude that the CMA-ES is the better optimizer. By contrast, FHAP calibration was not very sensitive with objective differences less than 0.021 J units.

Additionally, optimizing parameters also failed to be consistent across optimization seeds. The analysis of the transfer from Oracle to full-scale (see Fig. 5) showed good ranking agreement for EDRL (Spearman $\rho = 0.89$, $p = 0.007$) but not for FHAP (Spearman $\rho = -0.39$, $p = 0.383$). The short-delivery distances from the reduced-scale oracle to the FHAP were blamed for the poor transfer, and led to the adoption of geometry-preserving recalibration. The results presented here justify the use of environment-in-the-loop calibration for EDRL and show the constraints of reduced-scale optimization for charging-dependent FHAP policies.

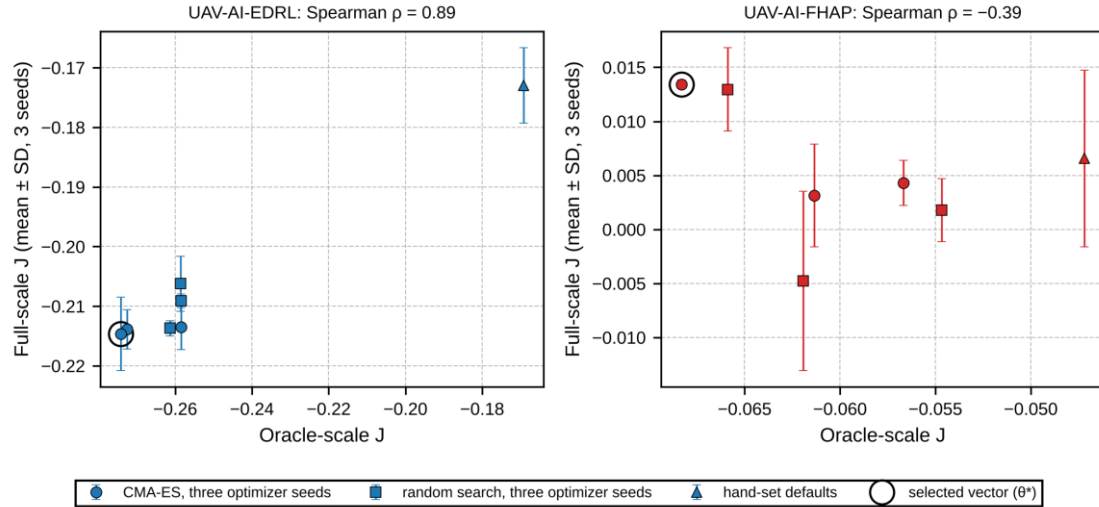


Fig. 5. Oracle-to-full-scale transfer of the seven candidate vectors: composite J at oracle scale against full scale (mean $\pm$ SD over three seeds).

3.3. Headline Comparison of the Seven Suites

Table 8 reports the full-scale metrics of all suites as means $\pm$ standard deviations over the six seeds, Table 9 the paired differences against the conventional suite, and Fig. 6 the seed-level values. Fig. 7 shows the five-hour traces of delivery, delay, power, and state of charge as seed means with ± 1 standard deviation bands, along with the weather trajectory in Fig. 8.

TABLE VIII. FULL-SCALE METRICS OF THE SEVEN SUITES (MEAN $\pm$ STANDARD DEVIATION OVER SIX ENVIRONMENT SEEDS; RADIO METRICS OVER ON-DUTY AIRCRAFT).

Metric	Conv.	QR	DRL	FL	EDRL	FHAP	Comp.
Mean end-to-end delay (ms)	66.0 $\pm$ 0.1	25.3 $\pm$ 0.0	66.0 $\pm$ 0.1	65.9 $\pm$ 0.1	25.3 $\pm$ 0.0	66.0 $\pm$ 0.1	25.3 $\pm$ 0.1
Packet delivery ratio	0.931 $\pm$ 0.001	0.933 $\pm$ 0.001	0.931 $\pm$ 0.000	0.932 $\pm$ 0.001	0.933 $\pm$ 0.000	0.932 $\pm$ 0.001	0.933 $\pm$ 0.001
Mean power (W)	407.4 $\pm$ 4.1	407.4 $\pm$ 4.1	373.8 $\pm$ 5.0	386.0 $\pm$ 6.7	371.8 $\pm$ 7.7	388.7 $\pm$ 5.9	362.1 $\pm$ 3.6
Gated missions per UAV	3.25 $\pm$ 0.02	3.25 $\pm$ 0.02	3.32 $\pm$ 0.02	2.73 $\pm$ 0.03	3.33 $\pm$ 0.03	2.74 $\pm$ 0.04	2.63 $\pm$ 0.04
Energy per gated mission (Wh)	640.5 $\pm$ 6.2	640.5 $\pm$ 6.2	573.3 $\pm$ 4.9	721.3 $\pm$ 11.0	570.8 $\pm$ 4.7	727.4 $\pm$ 13.7	698.7 $\pm$ 8.8
Mission success ratio	0.474 $\pm$ 0.004	0.474 $\pm$ 0.004	0.491 $\pm$ 0.003	0.519 $\pm$ 0.006	0.491 $\pm$ 0.004	0.514 $\pm$ 0.005	0.500 $\pm$ 0.001
Mission completion (%)	85.2 $\pm$ 0.3	85.2 $\pm$ 0.3	86.2 $\pm$ 0.2	77.1 $\pm$ 0.6	86.2 $\pm$ 0.7	77.1 $\pm$ 0.5	75.3 $\pm$ 0.8
Mean SINR (dB)	32.1 $\pm$ 0.4	32.1 $\pm$ 0.4	32.1 $\pm$ 0.4	32.2 $\pm$ 0.3	32.1 $\pm$ 0.3	32.2 $\pm$ 0.3	32.1 $\pm$ 0.3
Mean RSSI (dBm)	-56.3 $\pm$ 0.3	-56.3 $\pm$ 0.3	-57.2 $\pm$ 0.3	-56.2 $\pm$ 0.3	-57.2 $\pm$ 0.3	-56.2 $\pm$ 0.3	-57.1 $\pm$ 0.3
Spectral efficiency (bps/Hz)	7.88 $\pm$ 0.01	7.88 $\pm$ 0.01	7.84 $\pm$ 0.01	7.88 $\pm$ 0.00	7.83 $\pm$ 0.01	7.88 $\pm$ 0.00	7.84 $\pm$ 0.01
Link availability (%)	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0	100.0 $\pm$ 0.0
Mean flight time (min)	180.4 $\pm$ 1.1	180.4 $\pm$ 1.1	181.3 $\pm$ 2.0	178.3 $\pm$ 0.8	180.6 $\pm$ 2.3	180.3 $\pm$ 2.7	178.6 $\pm$ 1.6

Metric	Conv.	QR	DRL	FL	EDRL	FHAP	Comp.
Trajectory deviation (m)	6,747.9 ± 142.8	6,747.9 ± 142.8	6,427.7 ± 75.9	5,999.9 ± 51.1	6,383.9 ± 55.8	6,039.5 ± 54.5	5,658.6 ± 35.2
Near-miss events per UAV	115.4 ± 1.2	115.4 ± 1.2	94.2 ± 0.7	79.0 ± 0.7	94.5 ± 0.9	79.7 ± 0.6	78.0 ± 1.0
Charges served per UAV	6.30 ± 0.14	6.30 ± 0.14	6.98 ± 0.03	8.42 ± 0.08	7.01 ± 0.04	8.38 ± 0.06	9.10 ± 0.06
Mean transmit power (dBm)	23.0 ± 0.0	23.0 ± 0.0	22.0 ± 0.0	23.0 ± 0.0	22.0 ± 0.0	23.0 ± 0.0	22.0 ± 0.0
Same-channel neighbours (en route)	2.90 ± 0.02	2.90 ± 0.02	2.39 ± 0.03	2.66 ± 0.02	2.40 ± 0.02	2.65 ± 0.03	2.31 ± 0.03
Policy time (μs per UAV-step)	0.1 ± 0.0	2.5 ± 0.0	2.5 ± 0.0	0.6 ± 0.0	5.1 ± 0.2	0.9 ± 0.0	5.3 ± 0.5

TABLE IX. PAIRED DIFFERENCES OF THE HEADLINE METRICS AGAINST UAV-CONVENTIONAL OVER SIX SEEDS: MEAN DIFFERENCE WITH 95% CI, RELATIVE CHANGE, PAIRED t , HOLM-ADJUSTED, AND WILCOXON p -VALUES, AND EFFECT SIZE d_z (ABBREVIATIONS AS IN TABLE 8; THE REMAINING METRICS ARE IN THE RELEASED ANALYSIS OUTPUT).

Metric	Suite	Difference [95% CI]	Change (%)	p (paired t)	p (Holm)	p (Wilcoxon)	d_z
Mean end-to-end delay (ms)	QR	-40.65 [-40.67, -40.64]	-61.6	< 0.001	< 0.001	0.031	-2,849.27
Mean end-to-end delay (ms)	DRL	-0.03 [-0.08, 0.03]	-0.0	0.319	1.000	0.312	-0.45
Mean end-to-end delay (ms)	FL	-0.04 [-0.08, 0.01]	-0.1	0.103	1.000	0.094	-0.81
Mean end-to-end delay (ms)	EDRL	-40.64 [-40.71, -40.57]	-61.6	< 0.001	< 0.001	0.031	-588.32
Mean end-to-end delay (ms)	FHAP	-0.01 [-0.13, 0.11]	-0.0	0.847	1.000	0.844	-0.08
Mean end-to-end delay (ms)	Comp.	-40.65 [-40.73, -40.58]	-61.6	< 0.001	< 0.001	0.031	-566.08
Packet delivery ratio	QR	0.0015 [0.0015, 0.0016]	0.2	< 0.001	< 0.001	0.031	42.84
Packet delivery ratio	DRL	-0.0005 [-0.0010, 0.0001]	-0.0	0.071	1.000	0.094	-0.94
Packet delivery ratio	FL	0.0006 [-0.0002, 0.0014]	0.1	0.116	1.000	0.438	0.78
Packet delivery ratio	EDRL	0.0012 [0.0005, 0.0019]	0.1	0.006	0.266	0.031	1.84
Packet delivery ratio	FHAP	0.0003 [-0.0006, 0.0013]	0.0	0.386	1.000	0.688	0.39
Packet delivery ratio	Comp.	0.0015 [0.0008, 0.0022]	0.2	0.002	0.113	0.031	2.32
Mean power (W)	QR	0.00 [0.00, 0.00]	0.0	n/a	1.000	n/a	0.00
Mean power (W)	DRL	-33.65 [-38.16, -29.13]	-8.3	< 0.001	< 0.001	0.031	-7.82
Mean power (W)	FL	-21.47 [-30.42, -12.51]	-5.3	0.002	0.082	0.031	-2.52
Mean power (W)	EDRL	-35.63 [-46.81, -24.46]	-8.7	< 0.001	0.025	0.031	-3.35
Mean power (W)	FHAP	-18.75 [-25.82, -11.69]	-4.6	0.001	0.055	0.031	-2.79

Metric	Suite	Difference [95% CI]	Change (%)	<i>p</i> (paired <i>t</i>)	<i>p</i> (Holm)	<i>p</i> (Wilcoxon)	<i>d_z</i>
Mean power (W)	Comp.	−45.36 [−52.27, −38.44]	−11.1	< 0.001	< 0.001	0.031	−6.88
Gated missions per UAV	QR	0.00 [0.00, 0.00]	0.0	n/a	1.000	n/a	0.00
Gated missions per UAV	DRL	0.07 [0.04, 0.10]	2.1	0.003	0.132	0.031	2.23
Gated missions per UAV	FL	−0.52 [−0.57, −0.47]	−16.1	< 0.001	< 0.001	0.031	−10.80
Gated missions per UAV	EDRL	0.07 [0.03, 0.11]	2.2	0.006	0.261	0.031	1.86
Gated missions per UAV	FHAP	−0.52 [−0.56, −0.47]	−15.9	< 0.001	< 0.001	0.031	−12.44
Gated missions per UAV	Comp.	−0.62 [−0.66, −0.58]	−19.1	< 0.001	< 0.001	0.031	−15.23
Energy per gated mission (Wh)	QR	0.00 [0.00, 0.00]	0.0	n/a	1.000	n/a	0.00
Energy per gated mission (Wh)	DRL	−67.23 [−73.62, −60.84]	−10.5	< 0.001	< 0.001	0.031	−11.04
Energy per gated mission (Wh)	FL	80.78 [65.81, 95.74]	12.6	< 0.001	0.002	0.031	5.67
Energy per gated mission (Wh)	EDRL	−69.75 [−78.59, −60.91]	−10.9	< 0.001	< 0.001	0.031	−8.28
Energy per gated mission (Wh)	FHAP	86.91 [72.60, 101.22]	13.6	< 0.001	0.001	0.031	6.37
Energy per gated mission (Wh)	Comp.	58.14 [50.65, 65.64]	9.1	< 0.001	< 0.001	0.031	8.14
Near-miss events per UAV	QR	0.00 [0.00, 0.00]	0.0	n/a	1.000	n/a	0.00
Near-miss events per UAV	DRL	−21.24 [−22.39, −20.09]	−18.4	< 0.001	< 0.001	0.031	−19.40
Near-miss events per UAV	FL	−36.39 [−37.44, −35.34]	−31.5	< 0.001	< 0.001	0.031	−36.45
Near-miss events per UAV	EDRL	−20.95 [−22.52, −19.37]	−18.1	< 0.001	< 0.001	0.031	−13.93
Near-miss events per UAV	FHAP	−35.68 [−36.82, −34.54]	−30.9	< 0.001	< 0.001	0.031	−32.77
Near-miss events per UAV	Comp.	−37.37 [−38.29, −36.45]	−32.4	< 0.001	< 0.001	0.031	−42.74
Mean SINR (dB)	QR	0.00 [0.00, 0.00]	0.0	n/a	1.000	n/a	0.00
Mean SINR (dB)	DRL	0.08 [0.04, 0.11]	0.2	0.003	0.156	0.031	2.13
Mean SINR (dB)	FL	0.15 [0.06, 0.24]	0.5	0.008	0.311	0.031	1.76
Mean SINR (dB)	EDRL	0.02 [−0.07, 0.12]	0.1	0.528	1.000	0.438	0.28
Mean SINR (dB)	FHAP	0.15 [0.02, 0.27]	0.5	0.029	1.000	0.062	1.24
Mean SINR (dB)	Comp.	0.06 [−0.08, 0.21]	0.2	0.317	1.000	0.562	0.45

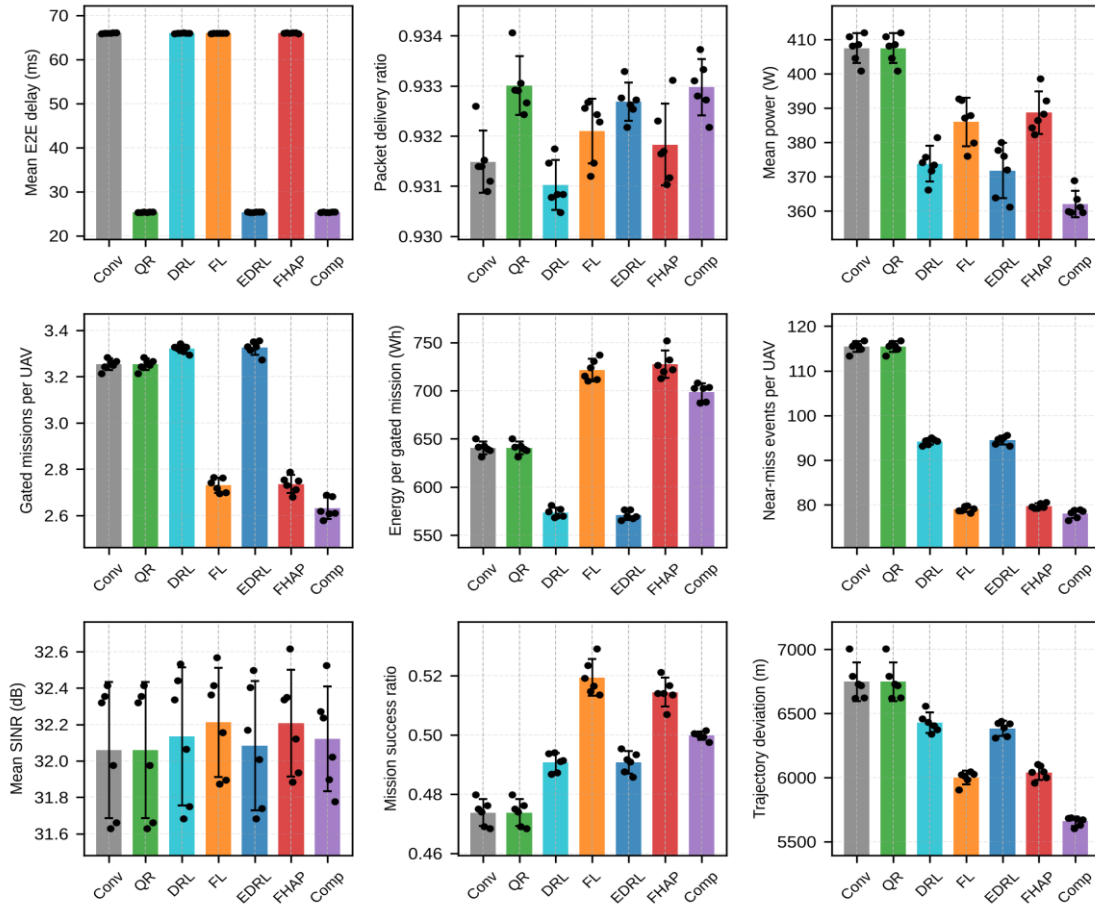


Fig. 6. Headline metrics of the seven suites: seed means with 95% confidence intervals and individual seeds.

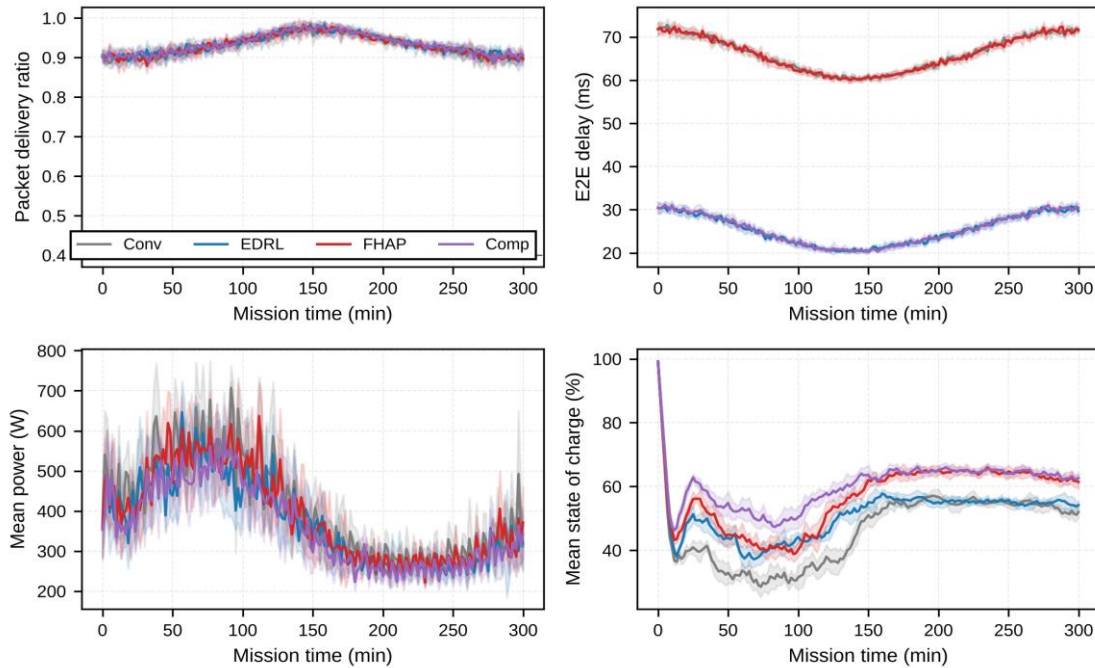


Fig. 7. Five-hour traces of delivery ratio, delay, power, and state of charge (seed mean $\pm$ one standard deviation).

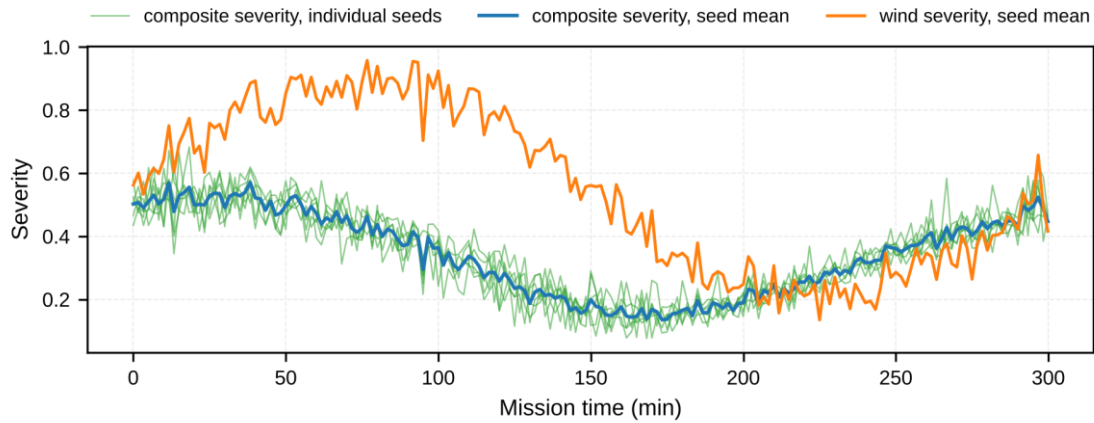


Fig. 8. Composite and wind severity over the mission for the six environment seeds.

Comparative results show a significant contribution from routing, flight control and energy management. The proactive IQMR forwarding was able to reduce the mean end-to-end delay for the UAV-QR, UAV-AI-EDRL, and composite suite by 61.6%, 66.0%, and 66.6%, respectively, whereas the configurations with AODV-based forwarding showed negligible mean end-to-end delay reductions. This improvement was not the result of learning the next-hop as is commonly believed, but rather was due to the proactive forwarding abstraction; the applicability of this improvement at the packet level is investigated separately. The number of packets delivered did not vary significantly for any suite. The mean power reductions for UAV-DRL and UAV-AI-EDRL were 8.3% and 8.7% respectively, while the energy per gated mission reductions were 10.5% and 10.9% respectively, and the number of near-miss events reduced by 18.4% and 18.1% respectively. The improvements were mainly attributed to the SOC-aware throttling and collision-avoidance features, and the remaining classical EDRL features were not seen to provide any significant improvement under the conditions used.

Compared to this, UAV-AI-FHAP decreased mean power by 4.6%, near-miss events by 30.9%, but gated mission throughput by 15.9% and increased energy per completed gated mission by 13.6%. This trade-off mainly came from the premature charging decisions due to the calibrated threshold. The composite architecture had the highest power saving (11.1%) and the closest approach to near-miss (32.4%) as well as improvement in routing delay, but had 19.1% less power in mission throughput. Subsequent analysis demonstrated that the reduction in the FHAP mission deficit using geometry-preserving recalibration was 2.2%, and the mission deficit was eliminated using leg-aware charging when compared to the original conventional baseline. Overall, the results show that while the gains from better communication efficiency, energy efficiency and flight safety do not directly correlate to mission throughput, mission-constrained cross-layer evaluation is key.

3.4. Mission Metrics, Gates, and Charging Logistics

The mission metrics have distinct denominators. Gated missions per UAV divides the number of arrivals that passed all four gates of Eq. (17) by the total number of provisioned aircraft, including off-duty aircraft, so it measures fleet throughput per provisioned airframe. The mission success ratio divides gated arrivals by all arrivals, and therefore measures the quality of arrivals rather than their number: a suite can raise throughput while lowering the ratio if its additional arrivals occur under worse conditions. Mission completion is the share of aircraft that completed at least two gated missions and one charge over the five hours, and its on-duty variant restricts the denominator to aircraft on duty at the end of the run. Table 10 gives the gate-failure breakdown and the charging statistics.

TABLE X. ARRIVALS, GATED MISSIONS, SHARE OF ARRIVALS FAILING THE NEAR-MISS AND ENERGY GATES (THE DELIVERY AND WEATHER GATES FAILED NO ARRIVAL), CHARGING REQUESTS AND SERVICES, STATE OF CHARGE AT THE CHARGING DIVERT, SHARE OF ON-DUTY AIRCRAFT IN CHARGING MODE, EN-ROUTE SPEED, AND DUTY CYCLE (SEED MEANS).

Suite	Arrivals / UAV	Gated / UAV	Fail g_3 near miss (%)	Fail g_1 SOC (%)	Charges requested / UAV	Charges served / UAV	SOC at divert	Charging share (%)	En-route speed ($m \cdot s^{-1}$)	Duty on
UAV-Conventional	6.87	3.25	52.6	0.00	12.14	6.30	0.18	60.7	14.48	0.601
UAV-QR	6.87	3.25	52.6	0.00	12.14	6.30	0.18	60.7	14.48	0.601
UAV-DRL	6.77	3.32	50.9	0.00	11.13	6.98	0.19	55.1	12.54	0.604

Suite	Arrivals / UAV	Gated / UAV	Fail g_3 near miss (%)	Fail g_1 SOC (%)	Charges requested / UAV	Charges served / UAV	SOC at divert	Charging share (%)	En-route speed (m·s ⁻¹)	Duty on
UAV-FL	5.26	2.73	48.1	0.00	12.91	8.42	0.35	62.3	13.65	0.594
UAV-AI-EDRL	6.78	3.33	50.9	0.00	11.17	7.01	0.19	54.8	12.54	0.602
UAV-AI-FHAP	5.32	2.74	48.6	0.00	12.91	8.38	0.35	62.5	13.66	0.601
UAV-AI-Composite	5.26	2.63	50.0	0.00	12.67	9.10	0.35	57.4	12.20	0.595

The near-miss gate is the binding gate for every suite: 52.6% of the conventional suite's arrivals failed it, the delivery and weather gates failed no arrival, and the energy gate failed fewer than 0.01% of arrivals (Table 10). At 52 aircraft per km² and a 50 m separation standard, a five-hour mission produces more than a hundred near-miss events per aircraft, so the gate filters about half of all arrivals in every suite; it is a strong constraint on the mission count but a blunt discriminator between suites, and the manuscript therefore reports near-miss events per aircraft as a separate safety metric. The trade-off is visible in the denominators: UAV-AI-FHAP has the higher success ratio (0.514 versus 0.474) because its aircraft arrive with fewer recent near misses, yet it completes fewer gated missions because it attempts fewer arrivals (5.32 versus 6.87 per aircraft) after spending more time charging. Charging demand follows the thresholds: the conventional and EDRL suites request a charge at 20% state of charge and are served about 6.30 to 7.01 times per aircraft, while the federated zone thresholds between 0.32 and 0.40 raise the served charges to 8.38. Charging time itself differs little between suites: 62.5% of on-duty aircraft in UAV-AI-FHAP were in charging mode at a given step, compared with 60.7% for the conventional suite, so the FHAP deficit has to be explained by where on a delivery leg the aircraft are diverted.

3.5. Diagnosis and Mitigation of the FHAP Mission Deficit

Three mechanisms could produce the mission deficit of the federated suite: more time spent charging, slower en-route flight, and flight lost on delivery legs that a charging divert interrupts. Additional runs that log the progress of every interrupted leg (three seeds) show that the number of charge cycles drives the deficit and that the slower flight adds little: at a centre of 0.20, with the same cruise factors, the deficit is only 1.3% (Table 11). With the threshold center at 0.36, each cycle replenishes about 60% of the battery instead of 75%, so the fleet completed 34.2% more charge cycles than the conventional suite (8.4 against 6.3 per aircraft). Every threshold divert interrupts the current leg, on average, after 42% of its length, and the aircraft receives a new target after charging, so the flight time already spent on that leg is lost. With legs of 5.1 km at full scale, the flight on interrupted legs amounted to 45% of the en-route distance of the conventional suite and 52% of that of UAV-AI-FHAP, whose en-route distance was also 9.7% shorter because more of its aircraft were charging and its wind-aware cruise factors flew 5.7% slower (Table 10). The productive distance, the en-route flight that ended in a delivery, fell by 21.3%, in line with the 21.7% fewer arrivals.

The calibration oracle could not see this cost. To preserve aircraft density it shrinks the service area, which shortens the legs to 1.2 km, so an interrupted leg loses little flight, and a higher centre keeps more aircraft hovering on charging pads below cruise power, which lowers the mean-power term of Eq. (16). therefore preferred the highest centre (Fig. 9 a; $J = -0.068$ at 0.36 against -0.009 at 0.22). A 5,400 s horizon on the same reduced area did not change the picture (J within 0.010 across centres), whereas a geometry-preserving oracle with the full 12 km square, 240 base stations, 40 vertiports, 750 aircraft, 5,400 s, and two environment seeds reproduced the full-scale preference, with J rising monotonically from -0.034 at 0.22 to 0.012 at 0.36. The failed transfer is therefore caused by the area reduction of the oracle and not by its horizon. CMA-ES on the geometry-preserving oracle (290 evaluations) placed the centre between 0.220 and 0.239 in all three optimizer seeds (selected vector: centre 0.22, span 0.03, cruise-factor span 0.06, pressure decay 0.76), and a variant whose energy term is the energy per gated mission instead of mean power, which removes the reward for time on the pads, placed it between 0.220 and 0.225.

TABLE XI. MITIGATION OF THE FHAP MISSION DEFICIT AT FULL SCALE (SIX SEEDS): GATED MISSIONS, POWER, ENERGY PER GATED MISSION, NEAR-MISS EVENTS, AND CHARGING STATISTICS, WITH THE PAIRED CHANGE AGAINST UAV-CONVENTIONAL (MEAN [95% CI]).

Configuration	Gated / UAV	Change vs Conv. (%)	Power (W)	Change vs Conv. (%)	Energy/mission change (%)	Near-miss change (%)	Charging share (%)	SOC at divert
UAV-Conventional, fixed threshold 0.20 (reference)	3.25	(reference)	407.4	(reference)	(reference)	(reference)	60.7	0.18
UAV-AI-FHAP, θ^* (centre 0.36)	2.74	-15.9 [-17.2, -14.6]	388.7	-4.6 [-6.3, -2.9]	13.6	-30.9	62.5	0.35

Configuration	Gated / UAV	Change vs Conv. (%)	Power (W)	Change vs Conv. (%)	Energy/mission change (%)	Near-miss change (%)	Charging share (%)	SOC at divert
UAV-AI-FHAP, centre 0.32	2.90	-10.8 [-12.0, -9.7]	393.7	-3.4 [-5.2, -1.6]	8.1	-28.8	61.6	0.31
UAV-AI-FHAP, centre 0.28	3.02	-7.1 [-8.6, -5.6]	390.4	-4.2 [-6.7, -1.7]	2.4	-27.1	60.1	0.27
UAV-AI-FHAP, centre 0.24	3.13	-3.8 [-5.2, -2.4]	391.4	-3.9 [-5.7, -2.2]	-0.6	-26.1	59.3	0.23
UAV-AI-FHAP, centre 0.20	3.21	-1.3 [-2.5, -0.0]	392.2	-3.7 [-5.3, -2.1]	-3.0	-25.6	58.9	0.20
UAV-AI-FHAP, recalibrated (geometry-preserving oracle)	3.18	-2.2 [-4.0, -0.5]	389.0	-4.5 [-7.1, -1.9]	-2.0	-25.6	59.1	0.21
UAV-AI-FHAP, recalibrated (energy per mission)	3.16	-3.0 [-4.3, -1.6]	395.9	-2.8 [-4.4, -1.2]	0.1	-26.7	59.6	0.21
UAV-AI-FHAP, leg-aware rule (federated estimate)	3.71	13.9 [13.2, 14.7]	392.4	-3.7 [-5.3, -2.1]	-15.0	-29.2	61.1	0.29
UAV-AI-FHAP, leg-aware rule (physics estimate)	3.65	12.1 [10.2, 14.1]	391.9	-3.8 [-5.7, -2.0]	-14.7	-29.8	61.2	0.31
UAV-Conventional, leg-aware rule	3.62	11.4 [9.9, 12.9]	405.8	-0.4 [-1.6, 0.8]	-10.5	-3.6	63.6	0.31
UAV-AI-Composite, θ^*	2.63	-19.1 [-20.4, -17.8]	362.1	-11.1 [-12.8, -9.4]	9.1	-32.4	57.4	0.35
UAV-AI-Composite, leg-aware rule	3.42	5.1 [3.9, 6.4]	367.2	-9.9 [-10.7, -9.1]	-14.9	-31.9	56.9	0.32

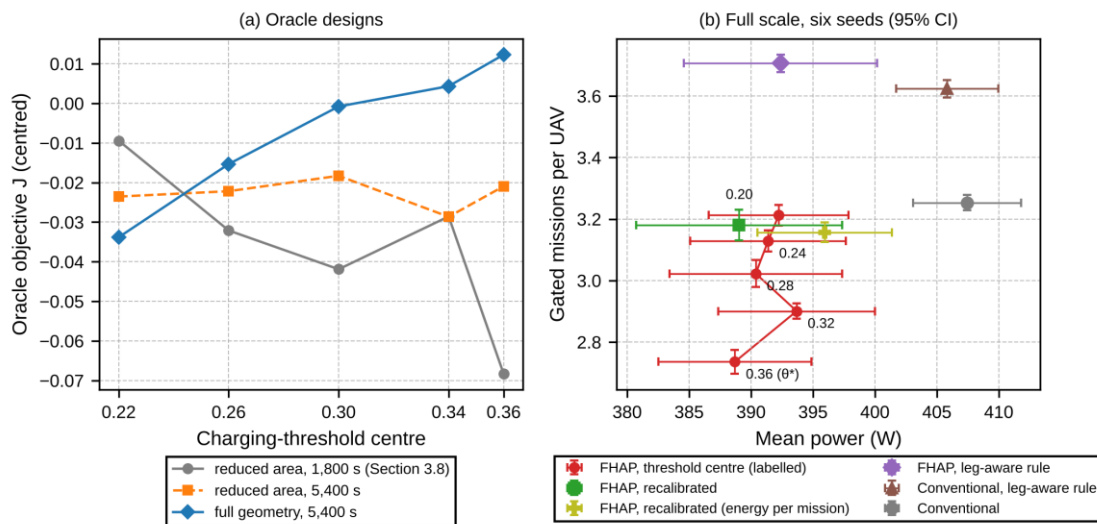


Fig. 9. FHAP charging: (a) oracle objective against the charging-threshold centre for three oracle designs; (b) gated missions against mean power at full scale for the threshold sweep, the recalibrated vectors, and the leg-aware rule.

Table 11 and Fig. 9b report the full-scale consequences. Lowering the centre traced a monotonic trade-off: from 0.36 to 0.20, the mission deficit against the conventional suite shrank from 15.9% to 1.3%, while the power saving remained between 3.4% and 4.6%, and the near-miss reduction between 25.6% and 30.9%. The recalibrated vector reduced the deficit to 2.2% [0.5, 4.0] with 4.5% lower power and 25.6% fewer near misses than the conventional suite. The leg-aware rule removed the deficit: with the federated energy estimate, UAV-AI-FHAP completed 13.9% [13.2, 14.7] more gated missions than the conventional suite and 35.5% more than with θ^* , at 3.7% lower power, 15.0% less energy per gated mission, and 29.2%

fewer near misses. The rule is not specific to the federated suite. Applied to the conventional suite, it raised gated missions by 11.4%; against this stronger baseline, UAV-AI-FHAP, with the same rule, completed 2.3% [0.8, 3.7] more gated missions at 3.3% lower power and with 26.5% fewer near misses. With the physics estimate of the energy rate, the rule completed 1.6% [0.0, 3.1] fewer gated missions than with the federated estimate, so most of the gain comes from deciding at the end of a leg and not from the learned energy model. The composite suite with the rule completed 5.1% more gated missions than the conventional suite, while consuming 9.9% less power. The mission deficit. thus a property of the calibrated threshold rule and of the oracle that selected it, and it can be removed while the energy and safety gains of the federated suite are kept.

3.6. Component-Level and Layer-Level Ablation at Full Scale

Table 12 removes each learned or rule-based component in turn from the full EDRL and FHAP suites and reports the paired change against the full suite over the same environment seeds (six for every component), together with the training controls, and Fig. 10 plots the component effects with their confidence intervals; Table 13 classifies every component with the equivalence tests and sets its effect against its cost; Table 14 and Fig. 11 give the layer-level ablation with the Pareto membership of every configuration.

TABLE XII. COMPONENT-LEVEL ABLATION AND TRAINING CONTROLS AT FULL SCALE: PAIRED CHANGE AGAINST THE FULL SUITE (MEAN [95% CI] OVER THE LISTED SEEDS).

Configuration	Seeds	Delay (ms)	PDR	Power (W)	Missions / UAV	Energy/mission (Wh)	Near misses / UAV
EDRL without consensus damping	6	-0.0 [-0.1, 0.1]	0.0002 [-0.0005, 0.0009]	6.5 [-6.0, 19.0]	0.0 [-0.0, 0.0]	3.1 [-5.9, 12.0]	1.2 [-0.1, 2.5]
EDRL without DQN (static random channel, nominal tier)	6	-0.0 [-0.1, 0.0]	0.0004 [0.0001, 0.0007]	2.6 [-8.2, 13.4]	-0.0 [-0.0, 0.0]	2.8 [-6.3, 11.9]	0.2 [-0.9, 1.3]
EDRL with fixed gate (no Eq. (19) adaptation)	6	-0.1 [-0.1, 0.0]	0.0003 [-0.0004, 0.0010]	2.0 [-10.5, 14.5]	-0.0 [-0.0, 0.0]	2.5 [-4.1, 9.2]	-0.3 [-1.1, 0.5]
EDRL without IQMR (AODV)	6	40.7 [40.7, 40.7]	-0.0020 [-0.0020, -0.0019]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
EDRL without MAPPO spread	6	0.0 [-0.1, 0.1]	-0.0000 [-0.0010, 0.0010]	-2.3 [-15.1, 10.5]	-0.1 [-0.2, -0.1]	17.8 [11.5, 24.2]	28.3 [26.4, 30.2]
EDRL without PPO (tier power only)	6	-0.0 [-0.1, 0.0]	-0.0002 [-0.0009, 0.0005]	-0.7 [-5.9, 4.6]	-0.0 [-0.1, 0.0]	2.0 [-2.9, 7.0]	-0.6 [-2.3, 1.2]
EDRL without cost-aware replanning	6	0.0 [0.0, 0.0]	0.0000 [0.0000, 0.0000]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
EDRL without path planner	6	-0.1 [-0.1, 0.0]	0.0003 [-0.0004, 0.0010]	2.0 [-10.5, 14.5]	-0.0 [-0.0, 0.0]	2.5 [-4.1, 9.2]	-0.3 [-1.1, 0.5]
EDRL without SOC throttle	6	-0.0 [-0.1, 0.0]	0.0003 [-0.0002, 0.0009]	34.3 [24.3, 44.2]	0.2 [0.1, 0.2]	25.8 [18.0, 33.6]	-9.4 [-10.6, -8.1]
EDRL with hand-set defaults instead of θ^*	6	-0.0 [-0.1, -0.0]	0.0005 [-0.0003, 0.0012]	-27.3 [-35.8, -18.8]	-1.1 [-1.2, -1.0]	214.1 [179.8, 248.3]	56.8 [48.5, 65.2]
FHAP with centralized model	6	-0.0 [-0.1, 0.1]	-0.0001 [-0.0008, 0.0007]	-0.4 [-6.2, 5.5]	-0.0 [-0.1, 0.0]	3.1 [-11.7, 17.8]	-0.9 [-1.7, -0.0]
FHAP with local-only models	6	-0.0 [-0.2, 0.1]	0.0002 [-0.0011, 0.0016]	0.6 [-7.4, 8.7]	-0.0 [-0.1, 0.0]	3.4 [-17.0, 23.9]	-1.5 [-2.4, -0.7]
FHAP with tuned fixed-gain PID	6	-0.0 [-0.1, 0.1]	0.0003 [-0.0006, 0.0011]	-2.7 [-15.8, 10.4]	-0.0 [-0.1, 0.0]	-6.1 [-21.4, 9.1]	-0.7 [-1.4, 0.0]

Configuration	Seeds	Delay (ms)	PDR	Power (W)	Missions / UAV	Energy/mission (Wh)	Near misses / UAV
FHAP without reactive avoidance	6	0.0 [−0.1, 0.1]	0.0001 [−0.0008, 0.0010]	−1.3 [−9.2, 6.7]	−0.2 [−0.2, −0.1]	50.4 [28.3, 72.5]	39.2 [37.9, 40.4]
FHAP without charging consensus (nearest vertiport)	6	−0.0 [−0.1, 0.0]	0.0002 [−0.0008, 0.0012]	−6.1 [−19.0, 6.9]	0.2 [0.2, 0.3]	−63.7 [−83.2, −44.2]	8.1 [6.8, 9.4]
FHAP without FedAvg setpoints (fixed threshold 0.20, no zone cruise factors)	6	0.0 [−0.1, 0.1]	−0.0003 [−0.0014, 0.0008]	21.3 [9.7, 33.0]	0.5 [0.5, 0.6]	−82.4 [−93.8, −70.9]	0.5 [−0.2, 1.2]
FHAP with hand-set defaults instead of θ^*	6	−0.0 [−0.2, 0.2]	0.0001 [−0.0015, 0.0016]	6.4 [−6.3, 19.1]	0.2 [0.2, 0.3]	−52.1 [−72.1, −32.2]	2.6 [1.6, 3.7]
EDRL: synthetic-reward training (previous protocol)	3	−0.0 [−0.1, 0.1]	−0.0012 [−0.0034, 0.0009]	5.2 [−39.6, 50.0]	−0.1 [−0.6, 0.4]	28.9 [−98.8, 156.7]	0.9 [−9.3, 11.1]
EDRL: 1% of the training budget	2	−0.0 [−0.1, 0.0]	0.0000 [−0.0086, 0.0087]	6.5 [−122.6, 135.6]	0.0 [−0.0, 0.1]	−3.0 [−68.2, 62.1]	0.9 [−7.2, 9.0]
EDRL: 10% of the training budget	2	−0.0 [−0.0, −0.0]	−0.0000 [−0.0012, 0.0011]	−1.9 [−130.7, 126.8]	0.0 [−0.1, 0.2]	−19.0 [−83.2, 45.2]	3.1 [−0.9, 7.2]
FHAP: synthetic-reward training (previous protocol)	3	−0.0 [−0.2, 0.1]	−0.0002 [−0.0013, 0.0010]	7.7 [−16.9, 32.2]	−0.0 [−0.2, 0.1]	14.7 [−5.9, 35.3]	−0.9 [−5.5, 3.8]
FHAP: 1% of the training budget	2	−0.0 [−0.5, 0.4]	0.0004 [−0.0055, 0.0063]	10.8 [−52.4, 73.9]	−0.1 [−0.7, 0.5]	38.6 [−24.9, 102.2]	−0.8 [−14.3, 12.6]
FHAP: 10% of the training budget	2	−0.0 [−0.1, 0.0]	0.0003 [0.0001, 0.0005]	5.9 [−100.2, 112.0]	−0.0 [−0.1, 0.1]	8.8 [−228.6, 246.2]	0.2 [−5.7, 6.1]

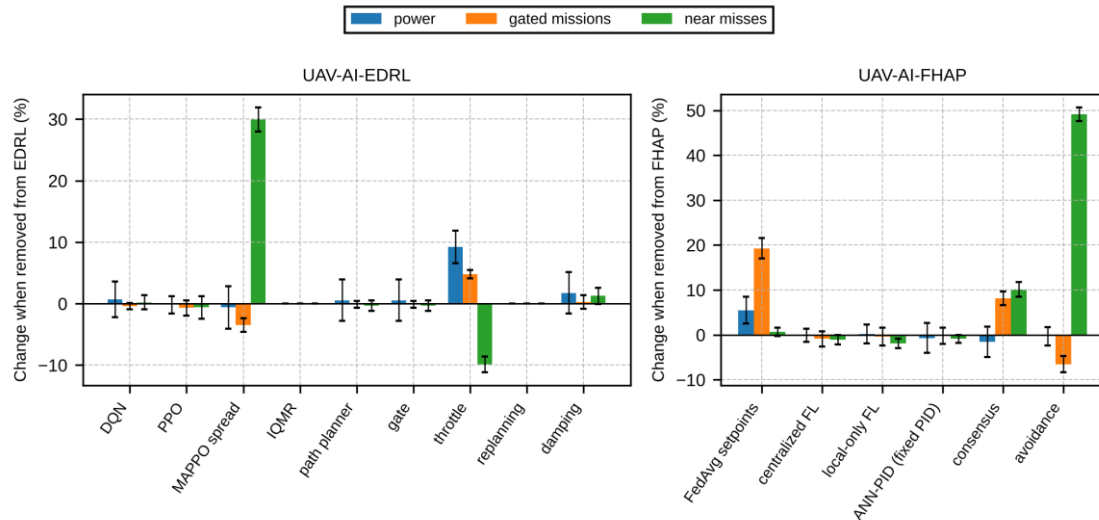


Fig. 10. Component ablation: change of power, gated missions, and near-miss events when each component is removed from EDRL (left) and FHAP (right), with 95% confidence intervals over six seeds.

TABLE XIII. CONTRIBUTION AND COST OF EVERY COMPONENT: PARAMETERS, TRAINING TIME, INFERENCE COST AT 7,500 AIRCRAFT, LARGEST CHANGE ON REMOVAL (%), MEAN [95% CI] OVER SIX SEEDS), EQUIVALENCE VERDICT, AND RANGE OF THE LARGEST REMOVAL EFFECT OVER TRAINING SEEDS 1 TO 3 (%), THREE ENVIRONMENT SEEDS EACH). D DELAY, R DELIVERY RATIO, E POWER, M GATED MISSIONS, EPM ENERGY PER GATED MISSION, NM NEAR-MISS EVENTS.

Component	Parameters	Training (s)	Inference (μ s / UAV-step)	Largest change on removal (%)	Verdict	Training seeds 1 to 3 (%)
UAV-AI-EDRL						
DQN channel and tier selection	5,968	27	0.173	E 0.7 [−2.2, 3.6]	inconclusive (E)	NM −7.9 to −0.2, same sign
PPO transmit-power refinement	9,347	35	0.619	M −0.7 [−1.9, 0.5]	equivalent to removal	EPM −4.6 to 0.2, sign varies
MAPPO spread controller	13,955	20	0.626	NM 29.9 [27.9, 31.9]	changes outcomes (M, EPM, NM)	NM 16.0 to 29.6, same sign
IQMR proactive forwarding	512	7	1.303	D 160.5 [160.5, 160.6]	changes outcomes (D)	D 160.6 to 160.9, same sign
Q-learning path planner	230	4	0.073	E 0.5 [−2.8, 3.9]	inconclusive (E)	NM −17.3 to −0.6, same sign
adaptive hand-off gate (rule)	rule	0	< 0.001	E 0.5 [−2.8, 3.9]	inconclusive (E)	n/a
SOC-aware throttle (rule)	rule	0	0.004	NM −9.9 [−11.2, −8.6]	changes outcomes (E, M, EPM, NM)	n/a
cost-aware replanning (rule)	rule	0	< 0.001	D 0.0 [0.0, 0.0]	equivalent to removal	n/a
consensus damping (rule)	rule	0	0.005	E 1.7 [−1.6, 5.1]	inconclusive (E)	n/a
UAV-AI-FHAP						
FedAvg zone setpoints	32	2	0.006	M 19.3 [17.0, 21.6]	changes outcomes (M, EPM)	M 18.9 to 20.2, same sign
ANN-PID gain scheduler	67	126	0.068	NM −0.9 [−1.8, 0.0]	inconclusive (E, EPM)	E −0.4 to 1.6, sign varies
charging consensus (rule)	rule	0	0.005	NM 10.1 [8.5, 11.8]	changes outcomes (M, EPM, NM)	n/a
reactive avoidance (rule)	rule	0	0.113	NM 49.1 [47.6, 50.7]	changes outcomes (M, NM)	n/a
FedAvg aggregation, against centralized training	32	2	0.006	NM −1.1 [−2.2, −0.0]	inconclusive (M, EPM)	n/a
FedAvg aggregation, against local-only training	32	2	0.006	NM −1.9 [−3.0, −0.9]	inconclusive (EPM)	n/a

TABLE XIV. LAYER-LEVEL ABLATION AT FULL SCALE (THREE-SEED MEANS): OBJECTIVES, CENTRED COMPOSITE J OF EQ. (16), AND PARETO MEMBERSHIP.

Configuration	E (W)	D (ms)	PDR	Missions / UAV	J	Pareto
EDRL (full)	373.9	25.33	0.9327	3.332	−0.2147	yes
EDRL without PHY	372.6	25.31	0.9324	3.329	−0.2154	yes
EDRL without MAC	376.4	25.31	0.9331	3.329	−0.2128	yes
EDRL without routing	373.9	66.00	0.9308	3.332	−0.0293	no
EDRL without trajectory	369.7	25.30	0.9327	3.217	−0.2108	yes

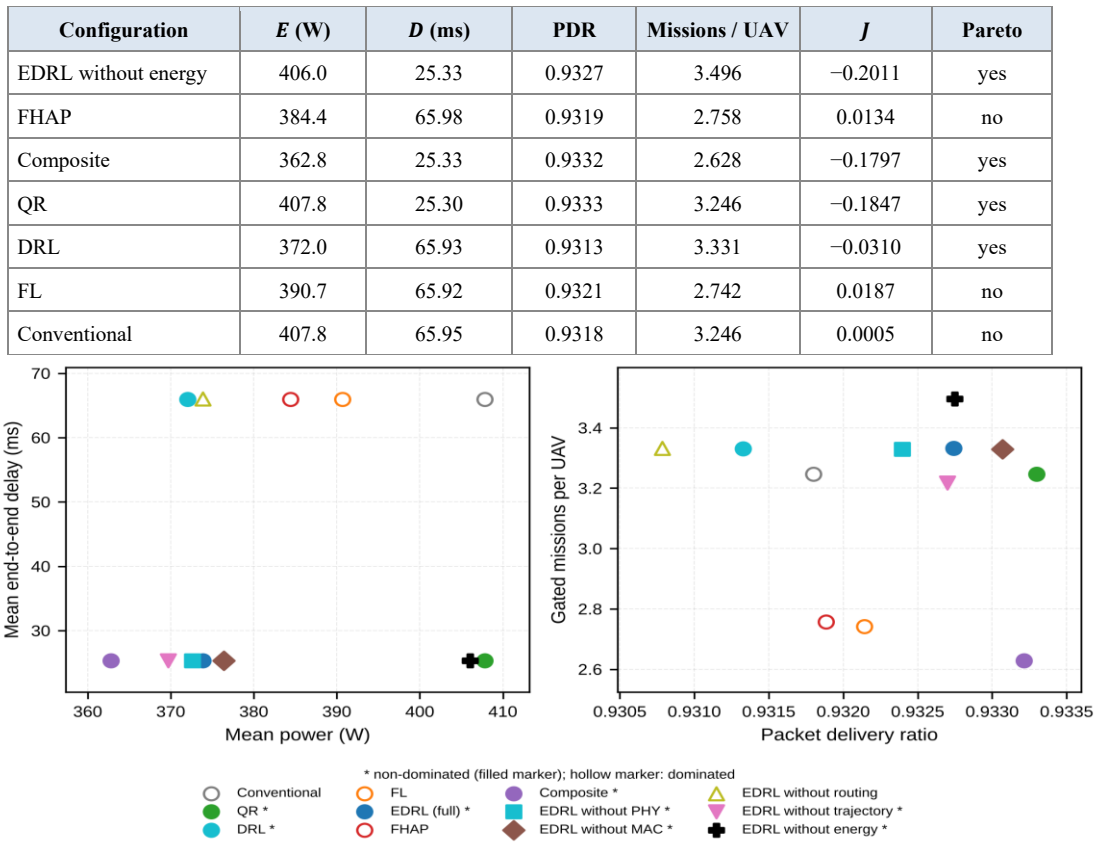


Fig. 11. Pareto projections of the suites and layer ablations: power against delay (left) and delivery against gated missions (right); filled markers are non-dominated.

The component ablation separates the mechanisms that change suite-level outcomes from those that do not (Tables 14 and 15). Removing IQMR returns the delay to the AODV level (66.0 ms), lowers the delivery ratio by 0.0020, and changes nothing else; because the best-SINR rule and a random rank give the same delay, this contribution belongs to proactive forwarding and not to the learned table. Removing the SOC-aware throttle raises mean power by 9.2% and gated missions by 4.8%, while lowering near misses by 9.9%: the throttle buys energy at the expense of mission throughput, and EDRL without its energy layer is itself Pareto-efficient (Table 14). Removing the MAPPO spread raises near-miss events by 29.9% and lowers gated missions by 3.5% through the near-miss gate, making it the only deep component with a measurable effect; its magnitude depends on the training seed (Table 13). In the FHAP suite the reactive avoidance rule carries the near-miss reduction (49.1% more near misses without it), and two mechanisms cost missions: the federated setpoints at θ^* (removing them raises gated missions by 19.3% at 5.5% higher power) and the charging consensus, which sends aircraft to farther stations (without it, gated missions rise by 8.1% and near misses by 10.1%). The equivalence tests classify the PPO power refinement and the replanning rule as equivalent to their removal on all six metrics and the DQN channel selection (power), the path planner (power), the adaptive gate (power), the consensus damping (power), the ANN gain scheduler (power, energy per mission), federated against centralized training (gated missions, energy per mission), and federated against local-only training (energy per mission) as inconclusive: none of these changes was significant, but the confidence intervals of the metrics in parentheses extend beyond the margins. The calibration row measures the joint effect of the six default coefficients and explains why the classical specialist is silent: with the hand-set gate base of 0.55 ($\tau = 0.50$) the planner acts on 270 of the 1,800 steps, on which its base-station-ward heading blend diverts every aircraft from its target, and the spread cap falls from 2.0 to 1.2 $\text{m} \cdot \text{s}^{-1}$; together these changes cut gated missions by 32.9% and raise energy per mission by 37.5%. Removing the whole spread controller costs only 3.5% of gated missions, so most of this loss comes from planner diversions. The main effect of calibrating the EDRL coefficients was to raise the gate base to 0.70 and silence a harmful component.

The training-seed check qualifies one of these verdicts. Removing the path planner changed nothing measurable with training seed 1, but with training seeds 2 and 3 it raised gated missions by 10.8% and 10.7% and lowered near-miss events by 17.3% and 10.6% (Table 13): the severity gate opens on one to four steps per run, and on those steps the planners of training seeds 2 and 3 move every flying aircraft one altitude band down or up, which concentrates the fleet in fewer bands, whereas the planner of training seed 1 holds the band at these severities. The removal effects of IQMR forwarding, of the spread controller

on near misses (16.0% to 29.6%), and of the FedAvg setpoints on gated missions (18.9% to 20.2%) kept their sign across the three training seeds.

The inference costs are small for every component. The throttle and the avoidance rule cost 0.004 and 0.11 μ s per aircraft-step, the spread controller 0.63 μ s with 13,955 parameters, and proactive forwarding 1.30 μ s, almost all of it for ranking the next-hop candidates, which the best-SINR rule needs as well (1.31 μ s); the DQN, PPO, and path planner add 5,968, 9,347, and 230 parameters and 0.17, 0.62, and 0.07 μ s per aircraft-step without a measurable benefit. The case for removing these components, therefore, rests on their training cost and on the training-seed risk shown above, not on inference time. Two reduced configurations test this attribution. EDRL with only best-SINR proactive forwarding, the spread controller, and the throttle matched the full suite within the margins on five of the six metrics, and its power change of -0.6% $[-4.1, 2.8]$ was not significant. FHAP with the tuned fixed-gain PID, the avoidance rule, and the leg-aware rule with the physics estimate, but without FedAvg, the ANN gain scheduler, or the consensus, matched FHAP with the federated leg-aware rule on delay, delivery, and near misses and completed 3.8% more gated missions at 3.2% $[1.2, 5.3]$ higher power, consistent with the omitted cruise factors (Table 12); against the conventional suite it completed 18.2% more gated missions with 31.2% fewer near misses and a power change of -0.6% . The incremental benefit of the remaining learning machinery is therefore not detectable at the suite level in this scenario: the reduced EDRL configuration keeps the gains of the full suite, and the reduced FHAP configuration gives up the power saving for missions while keeping the near-miss reduction.

3.7. Training Variability Against Environmental Variability

Table 15 crosses three training seeds with three environment seeds for the EDRL and FHAP suites and decomposes the variance of each headline metric.

TABLE XV. TWO-WAY VARIANCE DECOMPOSITION (THREE TRAINING SEEDS $\times$ THREE ENVIRONMENT SEEDS): GRAND MEAN, STANDARD DEVIATION OF THE TRAINING-SEED MEANS, STANDARD DEVIATION OF THE ENVIRONMENT-SEED MEANS, AND THE SHARES OF THE TOTAL SUM OF SQUARES.

Suite	Metric	Grand mean	SD over training seeds	SD over environment seeds	Training share (%)	Environment share (%)	Interaction (%)
UAV-AI-EDRL	delay (ms)	25.31	0.03	0.01	43.4	3.5	53.2
UAV-AI-EDRL	PDR	0.9333	0.0005	0.0003	56.2	22.0	21.8
UAV-AI-EDRL	power (W)	371.83	14.34	2.85	78.4	3.1	18.5
UAV-AI-EDRL	missions per UAV	3.11	0.19	0.10	68.9	19.8	11.3
UAV-AI-FHAP	delay (ms)	65.92	0.05	0.01	79.3	5.7	15.0
UAV-AI-FHAP	PDR	0.9322	0.0004	0.0001	55.5	3.2	41.3
UAV-AI-FHAP	power (W)	388.20	4.20	1.34	82.8	8.5	8.7
UAV-AI-FHAP	missions per UAV	2.73	0.02	0.02	46.3	32.3	21.4

For the delay and delivery metrics, both sources of variability are negligible relative to the suite effects: the three training seeds and three environment seeds shift the EDRL delay by less than 0.1 ms and its delivery ratio by less than 0.002. For the energy and mission metrics, the training seed dominates: for EDRL, the training-seed component accounts for 78% of the variance in mean power and 69% of the variance in gated missions, against 3% and 20% for the environment seed, and the mean power of EDRL ranged from 357 to 385 W and its gated missions from 2.96 to 3.33 per aircraft across training seeds. Training seeds 2 and 3, therefore, completed fewer gated missions than the conventional suite (3.25), so the small mission gain of EDRL reported in Table 9 for training seed 1 is not reproduced across training seeds, and the manuscript does not claim it. EDRL's near-miss events per aircraft were 94.3, 112.6, and 106.3 for training seeds 1 to 3 against 115.1 for the conventional suite on the same environment seeds, so the near-miss reduction ranges from 2.2% to 18.0%, whereas the power saving holds for every training seed (5.6% to 12.6%). The component runs. locate the source of this sensitivity in the trajectory layer: the path planners of training seeds 2 and 3 shift every flying aircraft by one altitude band on the few steps

at which the severity gate opens, and removing the planner raised the gated missions of these seeds by 10.8% and 10.7% (Table 13). For FHAP, the training seed also dominates the (smaller) variance in power, while its mission count is stable to within 0.1 per aircraft across both seed types, because its mission deficit is set by the charging center rather than by a learned component. Reporting a single training seed would have hidden a source of variability larger than the environmental one for the energy and mission metrics of the EDRL suite.

3.8. Robustness Under Distribution Shift

Table 16 and Fig. 12 evaluate the conventional, EDRL, and FHAP suites under nine conditions that differ from the training and calibration environment, each on the six environment seeds of the principal comparison. The conditions vary network density (half base-station density), environmental conditions (harsher channel, severe weather), mobility and demand patterns (Gauss-Markov velocity noise raised from 0.8 to 3.0 m·s⁻¹, and hotspot demand that draws 60% of the mission targets around four centres with an 800 m spread), and charging demand (24 instead of 40 vertiports, a duty cycle of 0.70 to 0.90, a 90 Wh instead of a 120 Wh battery, and half the charging rate). A stress regime combines the harsher channel with half base-station density and was run with the nominal components and with components retrained in that regime.

TABLE XVI. SUITE METRICS UNDER NINE DISTRIBUTION SHIFTS AND THE STRESS REGIME (SEED MEANS) AND THE PAIRED CHANGE OF EACH AI SUITE AGAINST THE CONVENTIONAL SUITE UNDER THE SAME CONDITION (%; * MARKS A 95% CONFIDENCE INTERVAL THAT EXCLUDES ZERO).

Condition	Seeds	Conv. power (W)	Conv. gated / UAV	EDRL power (%)	EDRL gated (%)	EDRL near misses (%)	FHAP power (%)	FHAP gated (%)	FHAP near misses (%)
nominal scenario	6	407.4	3.25	-8.7*	2.2*	-18.1*	-4.6*	-15.9*	-30.9*
harsher channel (NLOS 14 dB, K 2 dB)	6	407.4	3.25	-8.9*	2.3*	-18.0*	-4.5*	-15.6*	-31.5*
half base-station density (120)	6	406.9	3.22	-8.5*	3.6*	-18.7*	-3.6*	-14.5*	-31.0*
severe weather (amplitude $\times 1.4$, wind $\times 1.3$)	6	445.5	2.93	-10.1*	1.6	-16.3*	-5.7*	-16.6*	-32.6*
fewer vertiports (24)	6	411.0	3.10	-10.0*	5.5*	-23.3*	-4.8*	-16.3*	-35.1*
high duty cycle (0.70 to 0.90)	6	507.7	3.42	-6.5*	7.8*	2.7*	-2.4*	-14.0*	-22.6*
hotspot demand (60% of targets at four centres)	6	401.5	2.93	-7.0*	4.2*	-22.7*	-3.4*	-12.9*	-35.4*
erratic mobility (velocity noise 3.0 m·s ⁻¹)	6	418.8	2.96	-10.3*	-0.5	5.9*	-5.3*	-13.6*	-6.4*
small battery (90 Wh)	6	395.8	2.14	-9.6*	8.5*	-17.2*	-4.7*	-19.5*	-40.2*
slow charging (6% of capacity per minute)	6	387.0	1.71	-8.9*	12.0*	-13.7*	-4.4*	-19.3*	-47.0*
stress regime, nominal components	6	406.9	3.22	-8.6*	3.5*	-18.3*	-2.9*	-14.9*	-31.4*
stress regime, components retrained under stress	6	406.9	3.22	-9.1*	3.6*	-18.0*	-4.3*	-15.6*	-31.5*

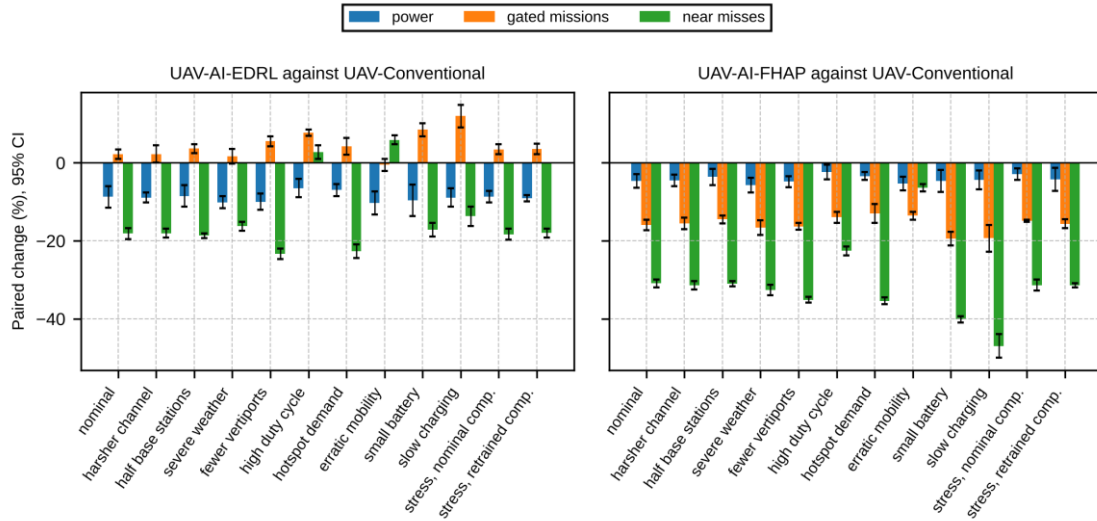


Fig. 12. Change of the EDRL (left) and FHAP (right) suites against the conventional suite under nine distribution shifts and the stress regime (95% confidence intervals over six seeds).

The routing and energy results of UAV-AI-EDRL held under every condition (Table 16, Fig. 13). Its delay reduction lay between 61.3% and 61.9% under the nine shifts and the stress regime, and the delivery ratios of the three suites differed by less than 0.3% under every shift: the harsher channel and the halved base-station density lowered the mean SINR of the conventional suite from 32.1 dB to 30.7 and 29.6 dB and its delivery ratio from 0.931 to 0.927 and 0.929, which confirms that delivery in this scenario is bounded by the environmental loss terms and not by the link budget. The EDRL power saving ranged from 6.5% under the high duty cycle to 10.3% under erratic mobility, with a confidence interval that excludes zero under all nine shifts, and its gated-mission change was positive, with a confidence interval above zero, under seven shifts and largest where energy was scarce (8.5% with the small battery and 12.0% with slow charging), because the energy saved by the throttle shortens the time on the pads. The near-miss reduction, which depends on the MAPPO spread controller, held under seven shifts (13.7% to 23.3%) but reversed under erratic mobility (+5.9% [4.8, 7.0]), where velocity noise of $3.0 \text{ m} \cdot \text{s}^{-1}$ masks the lateral drift of at most $2 \text{ m} \cdot \text{s}^{-1}$ that the controller applies, and under the high duty cycle (+2.7%), where the EDRL aircraft attempted 8.3 arrivals per aircraft against 7.5 for the conventional suite and the additional flying offset the separation gained.

The FHAP profile was reproduced across all shifts, with confidence intervals that exclude zero: power 2.4% to 5.7% lower, near misses 6.4% to 47.0% fewer, and gated missions 12.9% to 19.5% fewer than the conventional suite. The mission deficit was largest where charging was the bottleneck, 19.5% with the small battery and 19.3% with slow charging, because every early divert then costs more time on the pads, which is consistent with the charging diagnosis. Erratic mobility cut the near-miss reduction of FHAP to 6.4%, the smallest of all conditions, because the velocity noise also masks the $0.8 \text{ m} \cdot \text{s}^{-1}$ drift of the avoidance rule. In the stress regime, the mean SINR of the conventional suite fell to 28.2 dB, and its delivery ratio to 0.921, and both AI suites kept their profiles with the nominal components (EDRL: power -8.6%, near misses -18.3%; FHAP: power -2.9%, gated missions -14.9%, near misses -31.4%). Retraining the learned components in that regime produced one significant change in the delivery ratio of EDRL (+0.57% [0.41, 0.72]), with the retrained power control raising the mean transmit power from 22.0 to 22.2 dBm; no other metric in either suite showed a significant difference between nominal and retrained components. Six seeds per condition yield the minimum detectable effects in and the nine conditions perturb the parametric models' parameters; the experiment shows which results survive these perturbations and does not establish robustness beyond them.

3.9. Scalability and Computational Cost

Table 17 and Fig. 13 report the three principal suites at fleet sizes from 750 to 15,000 aircraft, with the infrastructure fixed, so that aircraft density rises from 5.2 to 104 km^{-2} , along with the measured computational cost for each suite at 7,500 aircraft.

TABLE XVII. FLEET-SIZE SWEEP WITH FIXED INFRASTRUCTURE (SIX SEEDS UP TO 7,500 AIRCRAFT, THREE AT 15,000) AND COMPUTE COST FROM SEQUENTIAL RUNS ON AN IDLE MACHINE: WALL TIME PER RUN, POLICY TIME PER AIRCRAFT-STEP, AND PEAK MEMORY.

N	Seeds	Suite	Density (200 m)	Delay (ms)	PDR	Power (W)	Missions / UAV	Near misses / UAV	Wall time (s)	Policy time (μ s)	Memory (MB)
750	6	UAV-Conventional	1.16	65.9	0.933	405.2	6.21	18.9	4	0.09	52
750	6	UAV-AI-EDRL	1.01	25.3	0.934	372.5	6.17	11.5	7	2.15	100
750	6	UAV-AI-FHAP	1.17	65.9	0.933	384.8	5.23	11.3	5	0.42	100
1,500	6	UAV-Conventional	2.33	65.9	0.933	402.7	5.71	34.8	6	0.06	101
1,500	6	UAV-AI-EDRL	2.07	25.3	0.933	372.1	5.68	22.8	11	1.78	103
1,500	6	UAV-AI-FHAP	2.29	65.9	0.933	383.7	4.77	21.7	7	0.29	103
3,000	6	UAV-Conventional	4.67	65.9	0.933	406.5	4.82	61.0	11	0.04	105
3,000	6	UAV-AI-EDRL	4.17	25.3	0.934	370.7	4.88	43.0	22	1.93	107
3,000	6	UAV-AI-FHAP	4.47	65.9	0.933	383.2	4.09	39.2	12	0.23	107
7,500	6	UAV-Conventional	11.59	66.0	0.931	407.4	3.25	115.4	26	0.03	125
7,500	6	UAV-AI-EDRL	10.55	25.3	0.933	371.8	3.33	94.5	50	1.62	128
7,500	6	UAV-AI-FHAP	10.62	66.0	0.932	388.7	2.74	79.7	29	0.25	128
15,000	3	UAV-Conventional	22.99	66.0	0.931	405.8	2.01	171.9	59	0.03	191
15,000	3	UAV-AI-EDRL	21.02	25.3	0.932	373.1	2.07	159.8	105	1.63	194
15,000	3	UAV-AI-FHAP	20.66	65.9	0.931	388.7	1.68	126.4	65	0.34	194

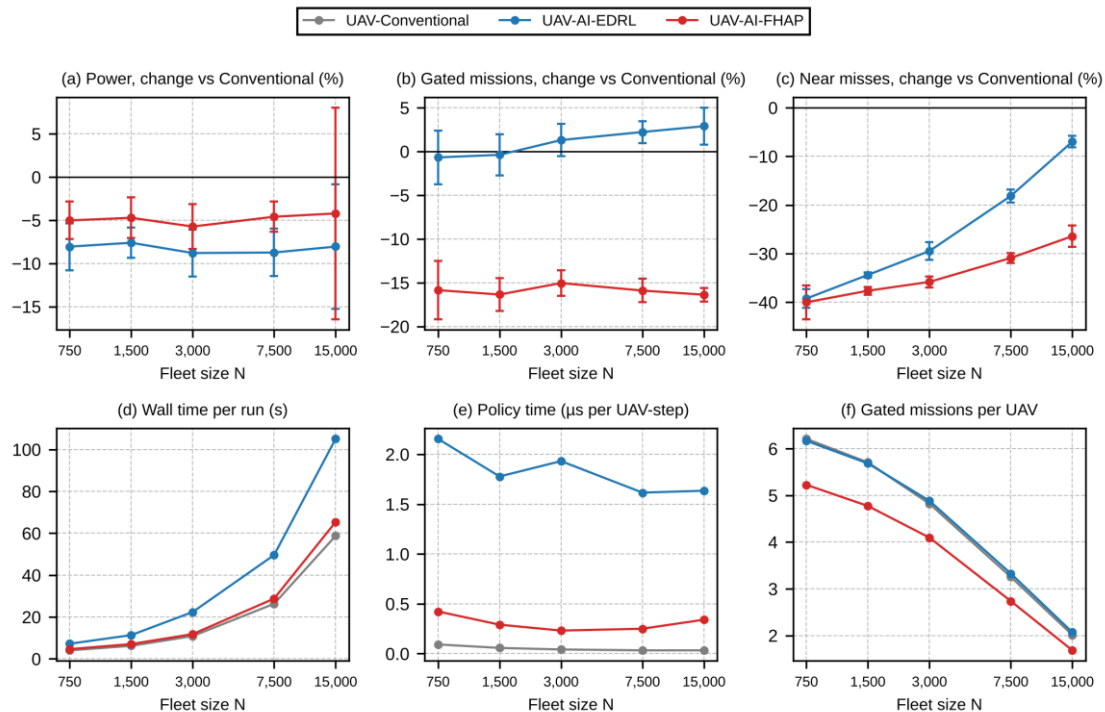


Fig. 13. Fleet-size sweep from 750 to 15,000 aircraft: change of power, gated missions, and near-miss events against the conventional suite with 95% confidence intervals (top), compute cost and gated missions (bottom).

Across a twentyfold range of fleet size, no significant suite effect changed sign (Table 17, Fig. 13), but the safety effects shrank with density. The EDRL suite reduces delay by 61.6% at 750 aircraft and by 61.6% at 15,000, and the delivery ratio of every suite stays between 0.931 and 0.934 at every size, although the mean SINR of the conventional suite falls from 34.0

dB to 31.2 dB as the same-channel neighbour count rises from 0.29 to 5.75: the interference term of Eq. (6) saturates at three same-channel neighbours and the links stay above 30 dB, so density does not reach the delivery metric in this scenario. Density reaches the mission metrics instead. Gated missions per aircraft fall from 6.21 to 2.01 for the conventional suite while arrivals stay near 6.8 per aircraft, because the share of arrivals that pass the near-miss gate falls from 0.912 to 0.292 as the near-miss count per aircraft grows from 18.9 to 171.9. The near-miss reductions of both AI suites shrink with density, for EDRL from 39.3% at 750 aircraft to 7.0% at 15,000, because a lateral drift of at most $2 \text{ m} \cdot \text{s}^{-1}$ separates aircraft with about one neighbour within 200 m but not with 23, and for FHAP from 40.0% to 26.4%. The EDRL power saving stays between 7.6% and 8.8%, and its gated-mission change ranges from -0.7% to 2.9% across all sizes, so the energy result is consistent across fleet sizes. The FHAP power saving lies between 4.6% and 5.7% up to 7,500 aircraft and is 4.2% at 15,000 aircraft, where its confidence interval over three seeds includes zero, while its mission deficit persists at 16.4%.

The compute columns of Table 17 come from sequential runs on an otherwise idle 2.1 GHz machine (one seed per size). Wall time grows close to linearly with fleet size: a five-hour run takes 4.0 s for the conventional suite at 750 aircraft and 58.7 s at 15,000, 7.2 s and 105.1 s for EDRL, and 4.6 s and 65.4 s for FHAP. The policy time per aircraft-step at 7,500 aircraft is $1.62 \mu\text{s}$ for EDRL, $0.25 \mu\text{s}$ for FHAP, and $0.03 \mu\text{s}$ for the conventional suite; the EDRL cost is dominated by the ranking of next-hop candidates and the forward passes of the PPO and MAPPO networks (Table 13), and it is $1.63 \mu\text{s}$ at 15,000 aircraft, so the per-aircraft decision cost does not grow with the fleet. Peak resident memory reaches 194 MB at 15,000 aircraft and decreases at smaller sizes. The training stage of the seven components took 5.6 min in total (Table 4), and the federated aggregation exchanged 435 kB of model parameters over its 50 rounds (2×34 participating clients $\times$ 32 parameters $\times$ 4 bytes per round).

3.10. Sensitivity of the Composite Objective to Its Weights

Table 18 and Fig. 14 recompute the composite ranking of the seven suites across six weight sets, using the seed means from Table 8. With the default weights, the suites score UAV-Conventional 0 (reference); UAV-QR -0.185 ± 0.004 ; UAV-EDRL -0.029 ± 0.005 ; UAV-FL 0.016 ± 0.003 ; UAV-AI-EDRL -0.216 ± 0.005 ; UAV-AI-FHAP 0.018 ± 0.006 ; UAV-AI-Composite -0.180 ± 0.002 , and the weight-free Pareto set among the seven suites is UAV-QR, UAV-AI-EDRL, and UAV-AI-Composite.

TABLE XVIII. RANK (AND CENTRED J) OF EACH SUITE UNDER ALTERNATIVE WEIGHT SETS (ENERGY/DELAY/DELIVERY/MISSIONS).

Weights	Conv.	QR	DRL	FL	EDRL	FHAP	Comp.	Best
default (0.30/0.30/0.20/0.20)	5 (0.000)	2 (−0.185)	4 (−0.029)	6 (0.016)	1 (−0.216)	7 (0.018)	3 (−0.180)	UAV-AI-EDRL
equal (0.25 each)	5 (0.000)	2 (−0.154)	4 (−0.026)	6 (0.027)	1 (−0.182)	7 (0.028)	3 (−0.134)	UAV-AI-EDRL
energy-heavy (0.50/0.20/0.15/0.15)	6 (0.000)	3 (−0.123)	4 (−0.044)	5 (−0.002)	1 (−0.170)	7 (0.001)	2 (−0.150)	UAV-AI-EDRL
delay-heavy (0.20/0.50/0.15/0.15)	5 (0.000)	2 (−0.308)	4 (−0.020)	6 (0.013)	1 (−0.329)	7 (0.015)	3 (−0.302)	UAV-AI-EDRL
mission-heavy (0.15/0.15/0.20/0.50)	5 (0.000)	2 (−0.093)	3 (−0.023)	6 (0.072)	1 (−0.117)	7 (0.073)	4 (−0.014)	UAV-AI-EDRL
delivery-heavy (0.15/0.15/0.50/0.20)	5 (0.000)	2 (−0.093)	4 (−0.016)	6 (0.024)	1 (−0.111)	7 (0.025)	3 (−0.072)	UAV-AI-EDRL

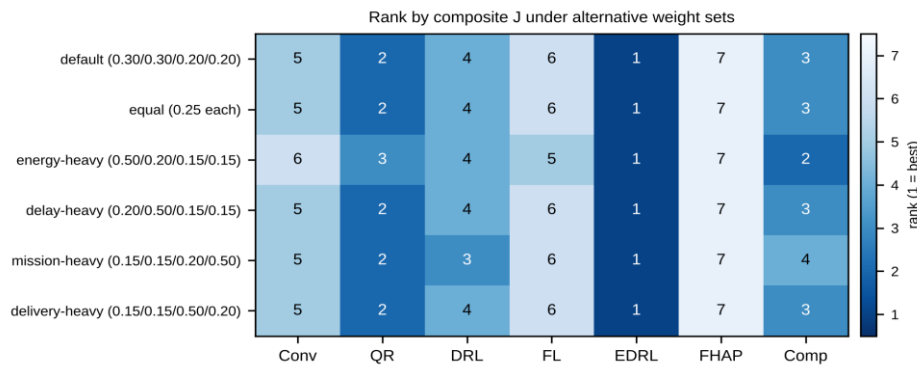


Fig. 14. Suite rank by the composite objective under six weight sets (1 = best).

The ranking is stable at the top and bottom within the tested range: UAV-AI-EDRL ranks first and UAV-AI-FHAP last across all six weight sets. UAV-QR ranks second except when energy receives half of the weight, where the composite takes second place; the composite falls to fourth behind UAV-DRL when missions receive half of the weight, and UAV-FL moves ahead of the conventional suite only under the energy-heavy weights. The weight-free Pareto set among the seven suites consists of UAV-QR, UAV-AI-EDRL, and UAV-AI-Composite: the conventional suite is dominated by UAV-QR, which matches it on power and missions and improves delay and delivery, and UAV-DRL, UAV-FL, and UAV-AI-FHAP are dominated by UAV-AI-EDRL (on the three-seed means of Table 14, UAV-DRL is non-dominated). The conclusions drawn from the composite objective, therefore, do not depend on the particular weights, although the magnitude of the gaps does.

3.11. Packet-Level Ns-3 Study of the Routing Abstraction

The routing yields the paper's largest single result, so its two structural assumptions were tested in NS-3 3.43 [36]: that per-hop loss is governed by the received SNR, and that reactive route discovery introduces a delay that proactive forwarding avoids. The scenario (Table 1) used 60 UAV nodes under three-dimensional random-waypoint mobility at $15 \text{ m}\cdot\text{s}^{-1}$ in a 1,200 m square with altitudes between 50 and 110 m, 9 fixed sinks on 25 m masts, an 802.11a ad hoc link with log-distance path loss and Nakagami fading, and 64-byte UDP packets at 2 packets per second from every UAV to its nearest sink, over transmit powers from 11 to 26 dBm with three runs each. Four forwarding schemes were compared: AODV [5]; OLSR, the proactive protocol evaluated for UAV ad hoc networks in [37], with its default timers and with fast timers; and a zero-overhead proactive scheme that recomputes minimum-hop routes every second from the true node positions and transmits no control packet, the packet-level counterpart of the forwarding assumed by the abstraction. Every scheme ran with standard ARP and with ARP caches populated before the run, which removes address-resolution traffic and waiting. Delivery is measured against the packets generated by the sources, counted before the socket, so packets dropped for lack of a route count as losses for every scheme, and the mean SNR of received frames is recorded so that the closed-form loss map can be evaluated at the same operating point (144 runs; Table 19, Fig. 15).

TABLE XIX. NS-3 PACKET-LEVEL RESULTS BY FORWARDING SCHEME AND ADDRESS RESOLUTION (DELIVERY AND DELAY: MEANS OVER THREE RUNS AND THREE TRANSMIT POWERS PER BAND; HOPS AND FRAMES: MEANS OVER ALL SIX POWERS) AND THE PREDICTIONS OF THE CLOSED-FORM ABSTRACTION AT THE MEASURED SNR AND HOP COUNT.

Forwarding	Address resolution	Delivery, 11 to 17 dBm	Delivery, 20 to 26 dBm	Delay (ms), 11 to 17 dBm	Delay (ms), 20 to 26 dBm	Hops	Frames per second
AODV	ARP	0.109	0.516	262.3	52.6	1.42	3,214
OLSR, default timers	ARP	0.114	0.651	86.6	92.2	1.38	474
OLSR, fast timers	ARP	0.142	0.688	189.4	160.7	1.65	726
zero-overhead proactive	ARP	0.228	0.742	264.0	91.0	1.76	611
AODV	pre-resolved	0.185	0.669	113.1	16.2	1.57	2,211
OLSR, default timers	pre-resolved	0.142	0.753	1.9	2.6	1.49	551
OLSR, fast timers	pre-resolved	0.187	0.826	2.2	2.7	1.80	773
zero-overhead proactive	pre-resolved	0.398	0.908	4.6	2.4	2.23	821
abstraction, reactive	n/a	0.949	0.946	66.9	63.7	n/a	n/a
abstraction, proactive	n/a	0.934	0.963	32.5	25.7	n/a	n/a

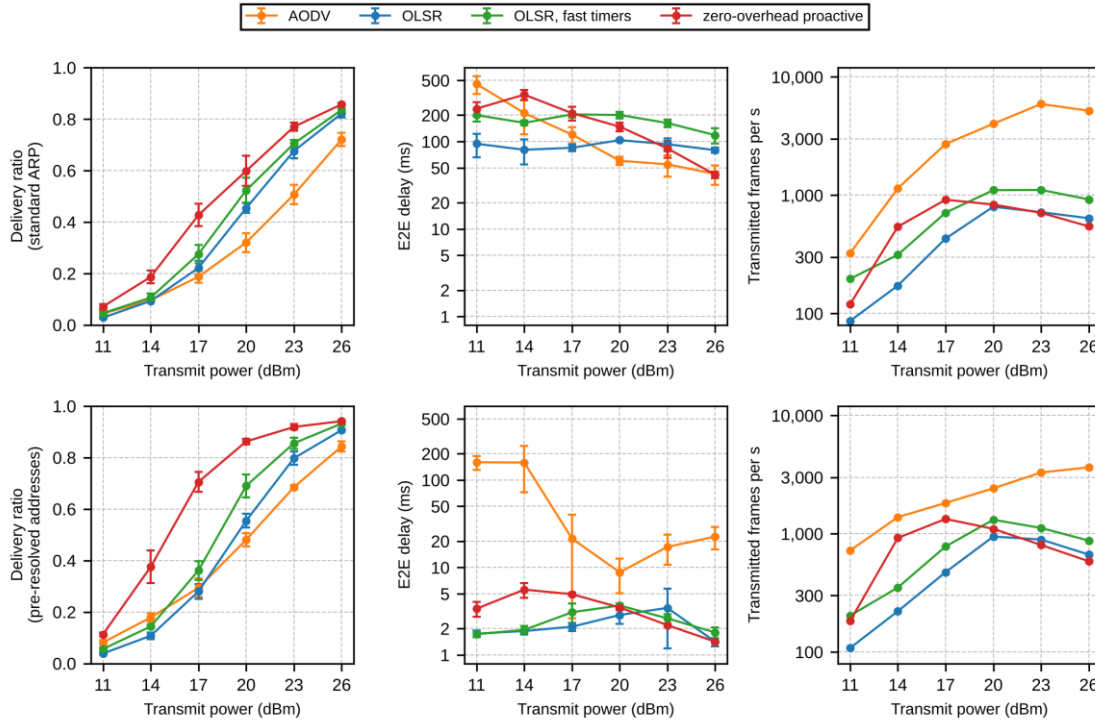


Fig. 15. NS-3 study: delivery ratio, end-to-end delay, and transmitted frames against transmit power for four forwarding schemes with standard ARP (top) and pre-resolved addresses (bottom).

Four results follow. First, the direction of the power effect is reproduced: delivery increased with transmit power across all schemes (Spearman correlation: 1.00). Second, address resolution, which the abstraction does not represent, dominated the packet-level delay: populating the ARP caches before the run cut the mean delay at 20 to 26 dBm from 52.6 to 16.2 ms for AODV and from 92.2 to 2.6 ms for OLSR, because with standard ARP the packets that follow a next-hop change wait for an address exchange, which is repeated after a one-second timeout when the broadcast request or its reply is lost. Third, with pre-resolved addresses the delay advantage of proactive forwarding appeared for the standard proactive protocol as well as for the zero-overhead scheme: at 20 to 26 dBm OLSR took 2.6 ms with default timers and 2.7 ms with fast timers, and the zero-overhead scheme 2.4 ms, against 16.2 ms for AODV (reductions of 83% to 85%), and at 11 to 17 dBm, where routes broke and discovery was repeated, AODV took 113.1 ms against 1.9 to 4.6 ms for the proactive schemes. With standard ARP the ordering depended on the scheme and the power: OLSR was faster than AODV at 11 to 17 dBm (86.6 against 262.3 ms) but slower at 20 to 26 dBm, fast timers raised its delay to 160.7 ms and its frame rate from 474 to 726 frames per second, and the zero-overhead scheme was no faster than AODV at either power (264.0 and 91.0 ms against 262.3 and 52.6 ms). A recent search-and-rescue study likewise reports AODV ahead on delivery and OLSR ahead on delay under its mobility models [38], so the ordering of standard protocols depends on the operating point. Fourth, no scheme reached the delivery predicted by the loss map at the measured SNR (0.88 to 0.98): the closest were the proactive schemes with pre-resolved addresses at 20 to 26 dBm (0.826 for OLSR with fast timers and 0.908 for the zero-overhead scheme), and at 11 to 17 dBm no scheme delivered more than 0.40; the schemes lost packets to route breaks under $15 \text{ m} \cdot \text{s}^{-1}$ mobility, contention, address resolution, and the absence of a route, none of which the abstraction represents, and the SNR of received frames is a survivor statistic.

The NS-3 evidence, therefore, separates the two parts of the routing claim. The direction, proactive forwarding that avoids the discovery delay, holds at packet level for OLSR and for zero-overhead forwarding when address resolution imposes no waiting; the magnitude of the full-scale reduction is a property of the abstraction, because at packet level the change against AODV ranged from a reduction of 98% (OLSR with default timers, pre-resolved addresses, 11 to 17 dBm) to an increase of 206% (OLSR with fast timers, standard ARP, 20 to 26 dBm).

4. CONCLUSION

This study evaluated seven UAV protocol suites, including conventional, reinforcement learning-based, federated ANN-PID, hybrid, and composite architectures, within a dense urban delivery scenario involving 7,500 aircraft and six independent environmental realizations. A unified five-layer optimization objective incorporating four mission-validity constraints was

developed alongside CMA-ES-based parameter calibration and physics-grounded training. Controlled experiments, paired statistical analyses, and component-level ablations were conducted to identify the mechanisms responsible for system-level performance improvements. The results demonstrated that performance gains were primarily associated with specific routing and control mechanisms rather than increased architectural complexity. Proactive forwarding reduced end-to-end delay by 61.6%, although packet-level NS-3 experiments indicated that this advantage depended on pre-resolved link-layer addresses. The UAV-AI-EDRL architecture reduced mean power consumption by 8.7% and near-miss events by 18.1%, primarily through SOC-aware throttling and collision-avoidance control. However, ablation analyses revealed no measurable suite-level benefits from several learned components, including DQN, PPO, path planning, ANN-PID, and federated aggregation. Furthermore, ANN-PID did not outperform an equally tuned conventional PID controller. The federated architecture reduced energy consumption and near-miss events but initially decreased completed missions by 15.9% because of premature charging decisions. Geometry-preserving recalibration reduced this deficit to 2.2%, while leg-aware charging achieved a 13.9% mission improvement relative to the original conventional baseline, decreasing to 2.3% when the same charging strategy was applied to that baseline. These findings emphasize the importance of mission-aware optimization and fair baseline comparisons when evaluating hybrid intelligent systems. The conclusions remain limited to the simulated physical models, environmental conditions, and fleet configurations. Although robustness experiments supported several performance improvements, the near-miss advantage of EDRL was not maintained under certain mobility and duty-cycle conditions. Future research should investigate full-scale packet-level validation using cellular air-to-ground communication, incorporate vertiport capacity and charging queues, and evaluate ANN-PID through hardware-in-the-loop experiments and fault injection. Overall, the study demonstrates that carefully selected cross-layer mechanisms can provide greater practical value than integrating additional learning components without measurable system-level benefits.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research received no external funding.

Acknowledgment

None.

References

- [1] N. Cheng, S. Wu, X. Wang, Z. Yin, C. Li, W. Chen, and F. Chen, "AI for UAV-assisted IoT applications: A comprehensive review," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14438–14461, 2023, doi: 10.1109/JIOT.2023.3268316.
- [2] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Toward autonomous multi-UAV wireless network: A survey of reinforcement learning-based approaches," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 3038–3067, 2023, doi: 10.1109/COMST.2023.3323344.
- [3] W. Meng, X. Zhang, L. Zhou, H. Guo, and X. Hu, "Advances in UAV path planning: A comprehensive review of methods, challenges, and future directions," *Drones*, vol. 9, no. 5, Art. no. 376, 2025, doi: 10.3390/drones9050376.
- [4] A. Puente-Castro, D. Rivero, A. Pazos, and E. Fernandez-Blanco, "A review of artificial intelligence applied to path planning in UAV swarms," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 153–170, 2022, doi: 10.1007/s00521-021-06569-4.
- [5] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing," *Internet Engineering Task Force*, RFC 3561, Jul. 2003.
- [6] A. Rovira-Sugranes, A. Razi, F. Afghah, and J. Chakareski, "A review of AI-enabled routing protocols for UAV networks: Trends, challenges, and future outlook," *Ad Hoc Netw.*, vol. 130, Art. no. 102790, 2022, doi: 10.1016/j.adhoc.2022.102790.
- [7] L. Zhang, X. Ma, Z. Zhuang, H. Xu, V. Sharma, and Z. Han, "Q-learning aided intelligent routing with maximum utility in cognitive UAV swarm for emergency communications," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3707–3723, 2023, doi: 10.1109/TVT.2022.3221538.
- [8] R. Ding, J. Chen, W. Wu, J. Liu, F. Gao, and X. Shen, "Packet routing in dynamic multi-hop UAV relay network: A multi-agent learning approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10059–10072, 2022, doi: 10.1109/TVT.2022.3182335.

- [9] X. Qiu, L. Xu, P. Wang, Y. Yang, and Z. Liao, “A data-driven packet routing algorithm for an unmanned aerial vehicle swarm: A multi-agent reinforcement learning approach,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 10, pp. 2160–2164, 2022, doi: 10.1109/LWC.2022.3195963.
- [10] N. Zhao, Z. Ye, Y. Pei, Y.-C. Liang, and D. Niyato, “Multi-agent deep reinforcement learning for task offloading in UAV-assisted mobile edge computing,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 6949–6960, 2022, doi: 10.1109/TWC.2022.3153316.
- [11] Z. Ning, Y. Yang, X. Wang, Q. Song, L. Guo, and A. Jamalipour, “Multi-agent deep reinforcement learning based UAV trajectory optimization for differentiated services,” *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5818–5834, 2024, doi: 10.1109/TMC.2023.3312276.
- [12] B. Brik, A. Ksentini, and M. Bouaziz, “Federated learning for UAVs-enabled wireless networks: Use cases, challenges, and open problems,” *IEEE Access*, vol. 8, pp. 53841–53849, 2020, doi: 10.1109/ACCESS.2020.2981430.
- [13] Y. Shen, Y. Qu, C. Dong, F. Zhou, and Q. Wu, “Joint training and resource allocation optimization for federated learning in UAV swarm,” *IEEE Internet Things J.*, vol. 10, no. 3, pp. 2272–2284, 2023, doi: 10.1109/JIOT.2022.3152829.
- [14] J. Tang, J. Nie, Y. Zhang, Z. Xiong, W. Jiang, and M. Guizani, “Multi-UAV-assisted federated learning for energy-aware distributed edge training,” *IEEE Trans. Netw. Service Manag.*, vol. 21, no. 1, pp. 280–294, 2024, doi: 10.1109/TNSM.2023.3298220.
- [15] Y. Jing, Y. Qu, C. Dong, W. Ren, Y. Shen, Q. Wu, and S. Guo, “Exploiting UAV for air-ground integrated federated learning: A joint UAV location and resource optimization approach,” *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 3, pp. 1420–1433, 2023, doi: 10.1109/TGCN.2023.3242999.
- [16] C. N. K. Reddy and A. M., “Hybrid intelligent routing with optimized learning (HIROL) for adaptive routing topology management in FANETs,” *Int. J. Electr. Electron. Eng.*, vol. 11, no. 11, pp. 30–43, 2024, doi: 10.14445/23488379/IJEEE-V11I11P104.
- [17] S. A. A. Shah, X. N. Fernando, and R. Kashef, “A survey on artificial-intelligence-based Internet of Vehicles utilizing unmanned aerial vehicles,” *Drones*, vol. 8, no. 8, Art. no. 353, 2024, doi: 10.3390/drones8080353.
- [18] M. Sajid, H. Mittal, S. Pare, and M. Prasad, “Routing and scheduling optimization for UAV assisted delivery system: A hybrid approach,” *Appl. Soft Comput.*, vol. 126, Art. no. 109225, 2022, doi: 10.1016/j.asoc.2022.109225.
- [19] C. Deng, X. Fang, and X. Wang, “UAV-enabled mobile-edge computing for AI applications: Joint model decision, resource allocation, and trajectory optimization,” *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5662–5675, 2023, doi: 10.1109/JIOT.2022.3151619.
- [20] I. A. Najm, M. Ismail, and A. A.-R. Tareq Rahem, “Wireless implementation selection in higher institution learning environment,” *J. Theor. Appl. Inf. Technol.*, vol. 67, no. 2, pp. 477–484, 2014.
- [21] 3GPP, “Study on enhanced LTE support for aerial vehicles,” 3rd Generation Partnership Project, Tech. Rep. TR 36.777, version 15.0.0, Dec. 2017.
- [22] M. B. Alshawki, I. A. Najm, and A. K. Hamoud, “Congestion control prediction model for 5G environment based on supervised and unsupervised machine learning approach,” *IEEE Access*, vol. 12, pp. 91127–91139, 2024, doi: 10.1109/ACCESS.2024.3416863.
- [23] A. T. Rahem, M. Ismail, N. F. Abdullah, and I. A. Najm, “New mathematical model to find the shortest path based on Boolean algebra operations for networks,” in *Proc. IEEE 3rd Int. Symp. Telecommun. Technol. (ISTT)*, Kuala Lumpur, Malaysia, 2016, pp. 112–114, doi: 10.1109/ISTT.2016.7918095.
- [24] V. Mnih et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015, doi: 10.1038/nature14236.
- [25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017, doi: 10.48550/arXiv.1707.06347.
- [26] C. Yu, A. Velu, E. Vinitsky, J. Gao, Y. Wang, A. Bayen, and Y. Wu, “The surprising effectiveness of PPO in cooperative multi-agent games,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 35, 2022, pp. 24611–24624.
- [27] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

- [28] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Agüera y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS)*, PMLR, vol. 54, 2017, pp. 1273–1282.
- [29] A. K. Hamoud, A. S. Alasady, W. A. Awadh, J. M. Dahr, M. B. M. Kamel, A. M. Humadi, and I. A. Najm, “A comparative study of supervised/unsupervised machine learning algorithms with feature selection approaches to predict student performance,” *Int. J. Data Min. Model. Manag.*, vol. 15, no. 4, pp. 393–409, 2023, doi: 10.1504/IJDMMM.2023.134590.
- [30] N. Hansen, “The CMA evolution strategy: A tutorial,” *arXiv preprint arXiv:1604.00772*, 2016, doi: 10.48550/arXiv.1604.00772.
- [31] S. Al-Yousif, I. A. Najm, H. S. Talab, N. H. Al Qahtani, M. Alfiras, O. Y. M. Al-Rawi, W. S. Al-Dayyeni, A. A. A. Alrawi, M. J. Mnati, M. Jarrar, F. Ghabban, N. A. Al-Shareefi, M. M. Jaber, A. H. Saleh, N. Md Tahir, H. T. Najim, and M. Taher, “Intrapartum cardiotocography trace pattern pre-processing, features extraction and fetal health condition diagnoses based on RCOG guideline,” *PeerJ Comput. Sci.*, vol. 8, Art. no. e1050, 2022, doi: 10.7717/peerj-cs.1050.
- [32] A. Patterson, S. Neumann, M. White, and A. White, “Empirical design in reinforcement learning,” *J. Mach. Learn. Res.*, vol. 25, Art. no. 318, pp. 1–63, 2024.
- [33] R. Agarwal, M. Schwarzer, P. S. Castro, A. Courville, and M. G. Bellemare, “Deep reinforcement learning at the edge of the statistical precipice,” in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 34, 2021, pp. 29304–29320.
- [34] D. Lakens, “Equivalence tests: A practical primer for t tests, correlations, and meta-analyses,” *Soc. Psychol. Pers. Sci.*, vol. 8, no. 4, pp. 355–362, 2017, doi: 10.1177/1948550617697177.
- [35] I. A. Najm, N. M. Hammash, M. Ismail, and P. Fatah, “Web based application visualization for comprehensive data structures,” *J. Theor. Appl. Inf. Technol.*, vol. 77, no. 2, pp. 171–177, 2015.
- [36] G. F. Riley and T. R. Henderson, “The ns-3 network simulator,” in *Modeling and Tools for Network Simulation*, K. Wehrle, M. Güneş, and J. Gross, Eds. Berlin, Germany: Springer, 2010, pp. 15–34, doi: 10.1007/978-3-642-12331-3_2.
- [37] A. H. Wheeb, R. Nordin, A. Abu Samah, and D. Kanellopoulos, “Performance evaluation of standard and modified OLSR protocols for uncoordinated UAV ad-hoc networks in search and rescue environments,” *Electronics*, vol. 12, no. 6, Art. no. 1334, 2023, doi: 10.3390/electronics12061334.
- [38] S. Patel and K. Raj, “Characterizing the performance of UAVNET routing protocols in search-and-rescue operations,” *Wireless Pers. Commun.*, vol. 146, no. 7, pp. 4339–4361, 2026, doi: 10.1007/s11277-026-12140-y.