Synchronized Multi-Drone Delivery via Public Transport Mobile Depots: A Dual-Phase Hybrid MILP and RL Framework

Mohamed Ali HAJJEM
Université de Lorraine
LGIPM
F-57000 Metz, France
mohamed-ali.hajjem@univ-lorraine.fr

Sadok TURKI
Université de Lorraine
LGIPM
F-57000 Metz, France
sadok.turki@univ-lorraine.fr

Nathalie SAUER
Université de Lorraine
LGIPM
F-57000 Metz, France
nathalie.sauer@univ-lorraine.fr

***Abstract*—The integration of Unmanned Aerial Vehicles (UAVs) with Bus Rapid Transit (BRT) networks offers a highly efficient paradigm for urban last-mile delivery. Operationalizing this system requires reconciling rigid timetables, stochastic demand, and non-linear energy consumption. This paper introduces a novel dual-phase optimization framework to address the Stochastic Public Transport-Based Drone Delivery Problem. Phase 1 employs a Mixed-Integer Linear Programming (MILP) model to establish a deterministic baseline schedule for planned parcels, strictly enforcing spatial-temporal synchronization with the mobile depot. Phase 2 introduces a Q-Learning dynamic controller to manage the stochastic arrival of high-yield express parcels. The Markov Decision Process (MDP) uses a custom reward function that balances financial yield against an aerodynamic power model and synchronization penalties. Computational experiments, simulating operations under variable meteorological conditions, demonstrate that the proposed hybrid architecture significantly increases total shift revenue while maintaining a zero-failure rate for drone-bus rendezvous.**

***Index Terms*—Vehicle Routing Problem with Drones (VRP-D); Public Transport Integration; Mixed-Integer Linear Programming (MILP); Reinforcement Learning; Aerodynamic Energy Modeling; Last-Mile Logistics.**

I. INTRODUCTION

Trending contactless commerce has placed unprecedented pressure on last-mile delivery systems, which represent a cost-intensive segment of the supply chain. The proliferation of delivery vans has led to a range of adverse urban impacts, including increased travel times for commuters, air pollution, and the degradation of urban infrastructure [1]. To address these challenges, the logistics industry and academic community have turned toward multi-modal solutions, which aim to provide cleaner, greener, and more accessible delivery services [2], [3].

One of the most innovative approaches currently under investigation is the use of public transport vehicles as mobile depots for drones [4], [5].

However, unlike traditional truck-drone pairings, where the truck can often wait for a returning drone or adjust its route, a BRT bus operates on a strict schedule [1], [5]. This necessitates a high degree of synchronization between the drones and the mobile depots [8]. Furthermore, the system must be resilient to stochastic disturbances [6], [9], such as the arrival of high-yield express parcels mid-shift, which require immediate routing decisions without violating the capacity constraints of the bus or the physical endurance limits of the drones.

The primary scientific challenge explored in this study is the reconciliation of a deterministic routing baseline with a stochastic demand environment. The problem is formulated in two distinct phases:

- **Phase 1** uses an exact Mixed-Integer Linear Programming (MILP) formulation to solve the vehicle routing problem with drones (VRP-D) for a set of pre-known, scheduled parcels. This ensures that the initial routes are optimized for minimum cost and maximum energy efficiency while adhering to the rigid BRT timetable.
- **Phase 2** introduces a dynamic insertion mechanism based on Q-Learning, a form of Reinforcement Learning (RL) that allows the system to adapt to express parcel arrivals in real-time [10]. The Q-Learning agent evaluates whether to accept, reject, or queue a new request based on the current state of the environment, including bus capacity, remaining time, and battery levels, which are themselves subject to environmental physics [11].

The scientific contributions of this research are fourfold:

1) It introduces a comprehensive MILP model for the VRP-D that explicitly incorporates the constraints of a public transport schedule and the synchronization requirements of a drone fleet.
2) It proposes a hybrid methodology that transitions from a deterministic optimization to an adaptive learning-based controller for dynamic demand management.
3) It integrates an energy consumption model that accounts for wind vectors, ambient temperature, humidity, and payload weight, providing a more realistic assessment of drone endurance than distance-based models [12].
4) It provides a framework for the specific off-peak operational window, highlighting the unique constraints of single-bus capacity and multi-drone synchronization.

II. Literature Review

The literature on drone-aided delivery has evolved rapidly since the introduction of the Flying Sidekick Traveling Salesman Problem [13]. As the field has matured, researchers have explored multi-truck and multi-drone systems, transitioning toward the integration of drones with various types of ground vehicles.

A. Deterministic Baseline and OR in VRP-D

The use of MILP for the deterministic VRP-D has reached a high level of sophistication. Modern formulations often treat the ground vehicle as a moving depot. The VRP with Release Times and Reloading at Mobile Satellites models the synchronization between delivery vehicles and mobile satellites, highlighting that en-route reloads can significantly reduce total distance and fleet size [8]. In the context of public transport, the problem becomes a three-echelon location-routing problem, where the bus acts as a link between fixed distribution centers and delivery points [7], [9]. Using fixed-route buses for freight can reduce vehicle mileage and fuel consumption by more than 5% [7]. However, the rigidity of bus schedules introduces a synchronization constraint, where the drone must arrive at a stop before the bus departs, a requirement heavily dependent on the drone's energy consumption profile [1].

B. Dynamic Insertion and RL in Logistics

The stochastic nature of modern logistics has led to learning-based approaches complementing traditional optimization. RL agents have demonstrated the ability to learn complex policies for hold vs. serve decisions, delaying requests to allow for future route consolidation and thereby maximizing the total customers served [10]. Integrating RL with integer programming represents a significant advancement, allowing real-time responses to urgent requests while periodically planning resource allocation.

C. UAV Energy Physics and Environmental Factors

A realistic drone routing model requires a energy consumption function. Traditional models treating energy as a linear function of distance are insufficient for urban environments experiencing high variance in battery drain. The literature identifies various physical models addressing lift-to-drag ratios, hovering power, and avionics losses, with recent emphasis on the impact of air density [14]. Recent research emphasizes the specific impact of air density (ρ), directly influenced by temperature, humidity, and altitude [12]. High temperatures and humidity reduce air density, requiring faster propeller rotation to generate the same thrust and increasing battery consumption. This relationship is quantified by the density altitude metric [12].

D. Research Gap and Contributions

Although there have been significant advancements in VRP-D and energy modeling, critical gaps remain. Most current models for public transport-based delivery assume infinite flexibility or focus on rural contexts [7]. There is a lack of research combining: 1) Strict operative limits tailored to time windows, 2) Single-bus capacity and parcel-sorting limits, 3) Managing a synchronized swarm of drones launching and landing on a moving vehicle, and 4) Applying hybrid MILP-to-RL methodologies to strict bus schedules influenced by complex energy physics. This study explicitly addresses these gaps by synthesizing these specific operational constraints into a unified stochastic framework.

III. System Conceptualization

The methodology is structured around a dual-phase optimization framework designed to bridge the gap between long-term reliability and short-term agility. The operational environment consists of a single BRT vehicle B traversing a fixed sequence of stops $\mathcal{S} = \{1, \ldots, n\}$ and a homogeneous fleet of drones $\mathcal{K} = \{1, \ldots, m\}$ launched from the vehicle's roof. The core of the system is a hybrid architecture that reconciles deterministic scheduling with stochastic demand.

A. Operational Boundaries and Assumptions

To formalize the problem space and isolate the stochastic variables, this framework establishes the following boundary conditions:

- The BRT timetable (A_s, D_s) is treated as a rigid temporal boundary, isolating system stochasticity exclusively to express parcel arrivals ($\mathcal{P}_{exp}$) rather than transit variability.
- The drone fleet $\mathcal{K}$ possesses uniform structural (W_{max}) and energetic (Q_{drone}) parameters, deliberately restricting the state space dimensionality for the Q-Learning agent.

B. Linear Mathematical Formulation

The deterministic phase establishes a baseline for pre-planned deliveries $\mathcal{P}_{plan}$, ensuring that the initial routing schedule is optimized for total operational cost while adhering to the rigid temporal constraints of the public transport network. The following notation and constraints formalize the interaction between the moving mobile depot and the aerial fleet.

1) Notation: All sets, parameters, and decision variables used in the MILP model are defined in Table I.

2) Objective Function: The objective function seeks to minimize the weighted sum of cumulative travel distance and non-linear environmental energy consumption across the entire fleet $\mathcal{K}$:

$$\min Z = \sum_{k \in \mathcal{K}} \sum_{i,j \in \mathcal{N}} (c_{dist}.P_{i,j}.x_{i,j,k} + c_{energy}.E_{i,j,k}) \quad (1)$$

The weight coefficients c_{dist} and c_{energy} serve as normalization factors. The coefficient c_{dist} [€/m] represents the marginal operational cost attributed to distance, encompassing hardware depreciation and maintenance intervals. Simultaneously, c_{energy} [€/J] serves as the bridge for energy consumption, reflecting the economic impact of battery throughput and electricity costs.

TABLE I
SETS, PARAMETERS, AND VARIABLES

Sets	
$\mathcal{N}$	Set of all nodes (stops $\mathcal{S}$ and customer locations $\mathcal{C}$)
$\mathcal{S}$	Set of BRT stops enabling drone synchronization
$\mathcal{K}$	Set of homogeneous drones, $\lvert\mathcal{K}\rvert = m$
$\mathcal{P}_{plan}$	Set of pre-known parcels for the deterministic phase
Parameters	
m	Total number of available drones in the fleet
T_{min}, T_{max}	Operational window boundaries [s]
C_{bus}	Capacity of the bus parcel sorting racks [units]
W_{max}	Maximum drone payload capacity [kg]
W_i	Weight of parcel destined for customer i [kg]
Q_{drone}	Maximum drone battery capacity [J]
ρ	Local air density, modeled as $f(T, P_v)$ [kg/m^3]
$\mathbf{V}_w$	Wind vector affecting ground speed [m/s]
A_s, D_s	Arrival and departure time of the bus at stop $s \in \mathcal{S}$ [s]
$P_{i,j}$	Spatial distance between nodes i and j [m]
v_g	Nominal drone ground speed [m/s]
η	Combined propulsion and battery efficiency factor [%]
$P_{induced}$	Induced power required for lift [W]
$P_{parasitic}$	Parasitic power to overcome body drag [W]
$P_{profile}$	Profile power to overcome blade drag [W]
δ_i	Service time at customer node i [s]
$\mathcal{E}$	Set of synchronization edges
M	A large positive constant for Big-M constraints
Decision Variables	
$x_{i,j,k} \in \{0,1\}$	1 if drone k travels from node i to node j
$y_{s,i,k} \in \{0,1\}$	1 if drone k launches from stop s to serve customer i
$z_{i,s',k} \in \{0,1\}$	1 if drone k returns to stop s' after serving customer i
$t_{i,k} \geq 0$	Arrival time of drone k at node $i \in \mathcal{N}$ [s]
$E_{i,j,k}$	Energy consumed by drone k amid node i and j [J]
$e_{i,k} \geq 0$	Remaining energy level of drone k at node $i \in \mathcal{N}$ [J]
$u_{i,k} \geq 0$	Remaining payload on $k \in \mathcal{K}$ upon arrival at $i \in \mathcal{C}$ [kg]

By defining these coefficients as monetary conversion factors, the formulation effectively transforms a bi-objective trade-off into a single-objective cost minimization problem. This prevents the large numerical values typically associated with energy from disproportionately dominating the optimization landscape, ensuring that the solver identifies a truly balanced equilibrium between path efficiency and battery preservation.

3) Main constraints:

a) Flow Conservation and Routing Logic: A mathematical mapping of the mobile depot's trajectory is essential to prevent sub-tours and guarantee structural network feasibility. Because the BRT vehicle continuously traverses the set of stops $\mathcal{S}$, standard circular flow constraints are fundamentally insufficient. To properly model the dispatch and recovery of the m-drone fleet, the flow conservation constraints are decoupled into explicit launch, transit, and recovery phases:

$$\sum_{k\in\mathcal{K}} \left(\sum_{s\in\mathcal{S}} y_{s,i,k} + \sum_{j\in\mathcal{C}, j\neq i} x_{j,i,k} \right) = 1, \quad \forall i \in \mathcal{P}_{plan} \tag{2}$$

$$\sum_{s\in\mathcal{S}} y_{s,i,k} + \sum_{j\in\mathcal{C}, j\neq i} x_{j,i,k} = \sum_{j\in\mathcal{C}, j\neq i} x_{i,j,k} + \sum_{s'\in\mathcal{S}} z_{i,s',k}, \quad \forall i \in \mathcal{C}, k \in \mathcal{K} \tag{3}$$

Constraint (2) mandates that each planned customer $i \in \mathcal{P}_{plan}$ is serviced exactly once by a single drone k, either via direct bus launch ($y_{s,i,k}$) or sequentially ($x_{j,i,k}$). Constraint (3) enforces strict flow conservation: a drone arriving at customer i must depart to another customer ($x_{i,j,k}$) or return to a valid BRT recovery stop ($z_{i,s',k}$). Decoupling the depot boundary variables (y, z) from intermediate flight variables (x) prevents fleet duplication and ensures no drone remains stranded.

b) Temporal Tracking and Mobile Synchronization: To guarantee that the aerial fleet operates strictly within the designated service period, all node arrival times are bounded by the global operational window parameters.

$$T_{min} \leq t_{i,k} \leq T_{max}, \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{4}$$

Arrival time at node j must account for the preceding arrival at i, service drop-off δ_i, and flight duration. Flight time is defined as the spatial distance $P_{i,j}$ divided by nominal ground speed v_g, dynamically adjusted for the projected wind vector $\mathbf{V}_w$:

$$t_{j,k} \geq t_{i,k} + \delta_i + \frac{P_{i,j}}{v_g - \mathbf{V}_w} - M(1 - x_{i,j,k}), \quad \forall i, j \in \mathcal{C}, k \in \mathcal{K} \tag{5}$$

To ensure temporal continuity across the entire sortie, the timeline is chained from the launch stop s to the recovery stop s' and enforced on the initial dispatch leg ($y_{s,i,k}$) and the final recovery leg ($z_{i,s',k}$):

$$t_{i,k} \geq t^{launch}_{s,k} + \frac{P_{s,i}}{v_g \pm \mathbf{V}_w} - M(1 - y_{s,i,k}), \quad \forall s \in \mathcal{S}, i \in \mathcal{C}, k \in \mathcal{K} \tag{6}$$

$$t^{return}_{s',k} \geq t_{i,k} + \delta_i + \frac{P_{i,s'}}{v_g \pm \mathbf{V}_w} - M(1 - z_{i,s',k}), \quad \forall i \in \mathcal{C}, s' \in \mathcal{S}, k \in \mathcal{K} \tag{7}$$

The core of the multi-modal coordination relies on spatiotemporal synchronization between the drones and the moving BRT depot. A drone k can only be dispatched from stop s to customer i if the launch occurs within the scheduled dwell time of the bus (between its arrival A_s and departure D_s).

$$t^{launch}_{s,k} \geq A_s - M\left(1 - y_{s,i,k}\right), \quad \forall s \in \mathcal{S}, i \in \mathcal{C}, k \in \mathcal{K} \tag{8}$$

$$t^{launch}_{s,k} \leq D_s + M\left(1 - y_{s,i,k}\right), \quad \forall s \in \mathcal{S}, i \in \mathcal{C}, k \in \mathcal{K} \tag{9}$$

Similarly, if drone k is assigned to rendezvous with the bus at a future stop s after serving customer i, its arrival time must be strictly less than or equal to the bus's scheduled departure time from that specific platform. This ensures the drone lands safely and is not stranded.

$$t^{return}_{s,k} \geq A_s - M\left(1 - z_{i,s,k}\right), \quad \forall s' \in \mathcal{S}, i \in \mathcal{C}, k \in \mathcal{K} \tag{10}$$

$$t^{return}_{s,k} \leq D_s + M\left(1 - z_{i,s,k}\right), \quad \forall s' \in \mathcal{S}, i \in \mathcal{C}, k \in \mathcal{K} \tag{11}$$

4) Physical and Energy Constraints:

a) Capacity and Structural Boundaries: Network capacity is bounded by BRT and drone structural limits. Drone dispatches from stop s cannot exceed the bus sorting rack capacity C_{bus} (12), and every delivery must respect the drone payload ceiling W_{max}.

$$\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{C}} y_{s,i,k} \leq C_{bus}, \quad \forall s \in \mathcal{S} \tag{12}$$

$$u_{i,k} \leq W_{max}, \quad \forall i \in \mathcal{C}, k \in \mathcal{K} \tag{13}$$

$$u_{j,k} \leq u_{i,k} - W_i + M(1 - x_{i,j,k}), \quad \forall i,j \in \mathcal{C}, k \in \mathcal{K} \tag{14}$$

$$u_{i,k} \geq W_i, \quad \forall i \in \mathcal{C}, k \in \mathcal{K} \tag{15}$$

b) Energy Physics and Environmental Constraints: The state of charge $e_{j,k}$ drops non-linearly based on parasitic, induced, and profile power requirements [12], [14], [15]:

$$\begin{aligned} e_{j,k} \leq e_{i,k} - \frac{1}{\eta}\Big[&P_{parasitic}(\rho) + P_{induced}(W_i, \rho) \\ &+ P_{profile}(\rho, \mathbf{V}_w)\Big] \cdot \frac{P_{i,j}}{v_g \pm \mathbf{V}_w} \\ &+ M(1 - x_{i,j,k}) \end{aligned} \tag{16}$$

A Battery Safety Buffer guarantees that the remaining energy does never fall below 20% of the maximum capacity Q_{drone}, preventing voltage drops during the mobile-landing phase [16]:

$$e_{i,k} \geq 0.2 Q_{drone}, \quad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{17}$$

The energy formulation integrates real-world meteorology to ensure feasible operational bounds. Air density $\rho := f(T, P_v)$ directly dictates aerodynamic lift [12]. Furthermore, the majorated payload W_i inflates the induced power requirement $P_{induced}$, as heavier parcels demand exponentially higher rotor thrust. Simultaneously, the prevailing wind vector $\mathbf{V}_w$ [15] acts as a double penalty: it increases profile drag $P_{profile}$ during headwinds and extends total flight time by altering the effective ground speed over the segment $P_{i,j}$ [14]. Additionally, the battery discharge efficiency η is scaled according to external thermal conditions, ensuring the operational range estimations remain mathematically sound and physically viable.

C. Q-Learning Dynamic Insertion Formulation

To address stochastic mid-shift express arrivals $\mathcal{P}_{exp}$, Phase 2 dynamically updates the deterministic MILP baseline via a Q-Learning-optimized Markov Decision Process [11].

1) State Space (S): The state $s \in S$ provides a comprehensive snapshot of the system's operational and environmental conditions at decision epoch t. It is defined as the following tuple:

$$s = \{s_{next}, C_{rem}, \{e_k\}_{k=1}^{4}, T_{rem}, \mathcal{Q}_{exp}, \rho, \mathbf{V}_w\} \tag{18}$$

Where s_{next} represents the upcoming BRT stop, C_{rem} is the remaining parcel capacity on the bus, $\{e_k\}$ tracks the energy levels of the drone fleet $\mathcal{K}$, and T_{rem} monitors the countdown to the temporal boundary. Real-time telemetry for air density ρ and wind $\mathbf{V}_w$ ensures the agent accounts for current aerodynamic resistance.

2) Action Space (A): Upon the arrival of an express parcel p, the agent selects an action $a \in A := \{\text{Accept}, \text{Queue}, \text{Reject}\}$. Accept$(p, k, s_{land})$ immediately dispatches drone k with a planned recovery at stop s_{land}. Queue(p) defers the decision to the next state transition for potential consolidation [10], while Reject(p) permanently discards requests violating energetic or temporal safety bounds.

3) Reward Function (R): The following piecewise reward function balances financial yield against physical and temporal risks:

$$R(s,a) = \begin{cases} \alpha \cdot \text{Yield}(p) - \beta \Delta E_{drone} - \gamma.\text{Risk}_{sync} & \text{if Accept} \\ -\delta \cdot \text{Yield}(p) & \text{if Reject} \\ 0 & \text{if Queue} \end{cases} \tag{19}$$

Where $\alpha, \beta, \gamma, \delta$ are weighting coefficients. The risk penalty $\gamma.\text{Risk}_{sync}$ is non-linear, activating sharply when the predicted return time t_{return} approaches the bus departure limit D_s, thus training the agent to prioritize synchronization over marginal revenue.

4) Q-Value Update Rule: The agent optimizes its policy π^* through the off-policy Bellman equation [17]. By iterating through states, the agent learns the long-term value of its actions using learning rate λ and discount factor θ:

$$Q(s,a) \leftarrow Q(s,a) + \lambda[r + \theta \max_{a'} Q(s', a') - Q(s,a)] \tag{20}$$

The execution logic for this update within the dynamic delivery environment is detailed in Algorithm 1.

This learning-based approach ensures that the drone fleet remains adaptive. Through training, the agent learns to proactively reject or queue parcels when environmental telemetry $(\rho, \mathbf{V}_w)$ suggests that high aerodynamic power requirements would threaten the 20% battery safety buffer or the non-negotiable BRT rendezvous.

IV. Computational Experiments and Results

To validate the proposed dual-phase framework, computational experiments were conducted using a simulated 10-stop BRT corridor modeled after the Mettis transit network in Metz (France) featuring dedicated transit lanes that guarantee the strict temporal fidelity of the mobile depot, with a fleet of $m = 4$ drones over an off-peak morning shift (09:30–10:30). The deterministic MILP baseline (Phase 1) was solved using `PuLP`, a standard pure-Python optimization library, while the dynamic Q-Learning controller (Phase 2) was trained over 10,000 episodes using a custom Python `gymnasium` environment. Environmental telemetry (air density and wind velocity) was modeled to reflect standard seasonal variations [18].

Algorithm 1: Dynamic Express Parcel Insertion Controller

Input: MILP Baseline Schedule, Discretized Q-table, State s
Output: Optimized Dynamic Policy π^*

while $T_{rem} > 0$ ***and*** $C_{rem} > 0$ **do**
  Observe $s = \{s_{next}, C_{rem}, \{e_k\}, T_{rem}, \mathcal{Q}_{exp}, \rho, \mathbf{V}_w\}$;
  Identify new express parcel arrival $p \in \mathcal{P}_{exp}$;
  Select action $a \in \{$Accept, Queue, Reject$\}$ via ϵ-greedy policy;
  if $a = Accept$ **then**
    Calculate predicted t_{return} and energy drain ΔE_{drone};
    if $t_{return} \geq 0.95 D_s$ **then**
      Execute transition and receive reward R with $\gamma \text{Risk}_{sync}$;
    else
      Execute transition and receive reward R with $-\beta \Delta E_{drone}$;
  Update $Q(s,a)$ via Bellman equation;
  $s \leftarrow s'$;

A. Deterministic Baseline Synchronization

Phase 1 successfully optimized 4-drone fleet utilization while strictly adhering to the non-negotiable BRT timetable bounds.

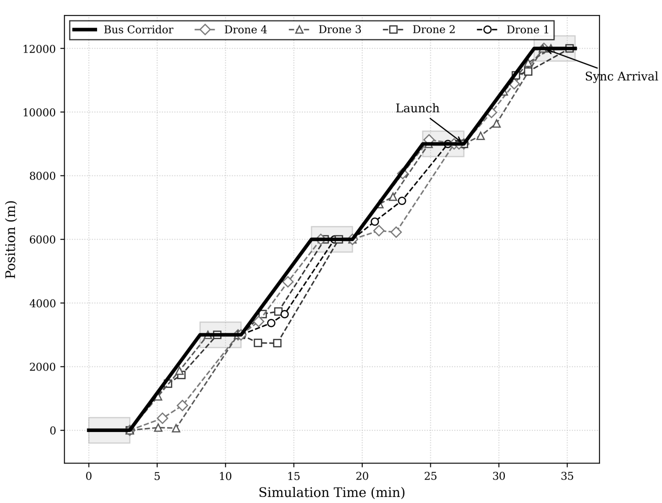

Fig. 1. Spatio-Temporal Synchronization of the Drone Swarm with the BRT Mobile Depot.

Fig. 1 demonstrates that all drone launch and recovery timestamps successfully fall within the BRT's dwell time boundaries ($A_s \leq t_{s,k} \leq D_s$). This deterministic schedule generates operational slack (idle time and excess rack capacity) which the Phase 2 dynamic controller targets.

B. Impact of Meteorological Variables on Range

Departing from linear approximations, the physics engine calculates parasitic, induced, and profile drag based on atmospheric variables.

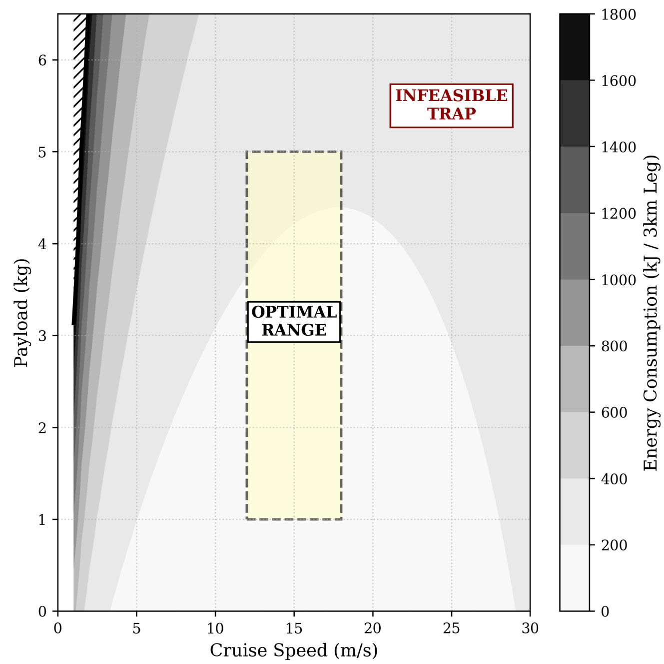

Fig. 2. Aerodynamic Energy Drain Sensitivity (1km Arc, 2kg Payload).

Fig. 2 demonstrates the non-linear sensitivity of the fleet's battery consumption to cruise speed (v) and payload mass. Because parasitic drag increases with the cube of velocity ($P \propto v^3$), high-speed operations rapidly drain the battery, pushing the system into an Infeasible Trap where energy requirements exceed the 1.8 MJ capacity constraint. Conversely, the model identifies an Optimal Range (approximately 12 to 18 m/s) where induced power (lift) and parasitic power (drag) are balanced. This operational envelope ensures that multi-drop sorties carrying maximum payloads remain within the mathematical bounds of the MILP constraints.

C. Q-Learning Agent Convergence

To dynamically insert express parcels ($\mathcal{P}_{exp}$), the agent optimizes $R(s,a)$ over 10,000 episodes using a learning rate $\lambda = 0.1$, discount factor $\theta = 0.95$, and ϵ-greedy exploration decay [19].

Fig. 3 tracks the cumulative episodic reward. Early in the training phase ($\epsilon > 0.6$), the agent frequently incurs the massive $\gamma \text{Risk}_{sync}$ penalty by attempting overly ambitious deliveries that result in missed rendezvous. However, as the policy π^* stabilizes around episode 7,000, the moving average smooths. The agent successfully learns to reject or queue parcels when weather telemetry indicates a risk to the temporal or energy constraints.

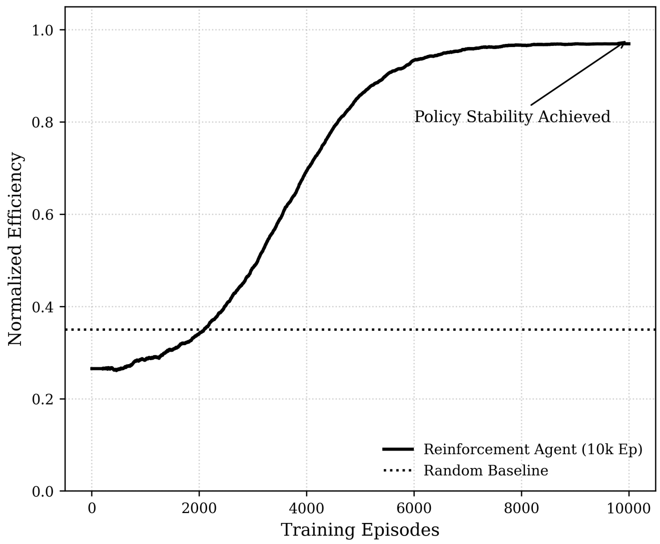

Fig. 3. Q-Learning Agent Convergence and Policy Optimization.

V. Discussion & Practical Implications

A. Meteorological Sensitivity and Density Altitude

Deploying this dual-phase Integrated Passenger and Freight Transport (IPFT) framework demonstrates sensitivity to density altitude shifts. A 10°C ambient temperature increase decreases air density (ρ), forcing higher propeller RPMs for equivalent lift and accelerating battery drain [12]. The Q-Learning agent natively absorbs these atmospheric perturbations by compressing the effective operational range, shifting toward a conservative policy (favoring Queue or Reject) to preserve the 20% safety buffer and guarantee bus rendezvous.

B. Pulsed Delivery Rhythm and Synchronization

Continuous drone launch and recovery loops from a moving BRT risk platform congestion. The framework mitigates this bottleneck via a pulsed delivery rhythm. Enforcing MILP temporal bounds (A_s, D_s) and penalizing high-risk return margins (P_{sync}) in the RL controller organically interlaces the drone sorties. This synchronization prevents simultaneous landings at any given bus stop (S_{next}), ensuring that the physical parcel sorting capacity (Cap_{bus}) and human-operator bandwidth are not overwhelmed during the constrained off-peak window.

VI. Conclusion

This study presents a novel dual-phase IPFT framework, effectively reconciling the deterministic rigidity of BRT schedules with the stochastic arrival of high-yield contactless commerce demands. By coupling an exact MILP baseline for pre-planned routing with a dynamic Q-Learning controller for real-time express parcel insertion, the system optimizes the operational window without disrupting passenger transit. Furthermore, integrating environmental physics to model battery degradation ensures that multi-drone launch and recovery synchronization remains robust against energy depletion. Ultimately, this hybrid OR-RL architecture establishes a foundation for future physical field tests in meteorologically variable urban environments.

References

[1] M. Lupi, D. Conte, and A. Farina, "Last mile urban freight delivery: A new system based on platoons of automated vehicles," *Transp. Policy*, vol. 164, pp. 42-59, Apr. 2025.

[2] A. Tighazoui, S. Turki, C. Sauvey, and N. Sauer, "Optimal design of a manufacturing-remanufacturing-transport system within a reverse logistics chain," *Int. J. Adv. Manuf. Technol.*, vol. 101, no. 5, pp. 1773-1791, Apr. 2019.

[3] S. Turki, M. Jemmali, and L. Hidri, "Optimal control for unreliable manufacturing and remanufacturing system under environmental subsidy and overrun penalty mechanism," *IEEE Trans. Eng. Manag.*, 2025.

[4] H. Huang, A. V. Savkin, and C. Huang, "When drones take public transport: Towards low cost and large range parcel delivery," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, Jul. 2019, pp. 1657-1660.

[5] T. Deng, X. Xu, Z. Zou, W. Liu, D. Wang, and M. Hu, "Multidrone parcel delivery via public vehicles: A joint optimization approach," *IEEE Internet of Things J.*, vol. 11, no. 6, pp. 9312-9323, Mar. 2023.

[6] H. Huang, A. V. Savkin, and C. Huang, "Drone routing in a time-dependent network: Toward low-cost and large-range parcel delivery," *IEEE Trans. Ind. Inform.*, vol. 17, no. 2, pp. 1526-1534, Feb. 2020.

[7] F. Bruzzone, F. Cavallaro, and S. Nocera, "The integration of passenger and freight transport for first-last mile operations," *Transp. Policy*, vol. 100, pp. 31-48, Jan. 2021.

[8] R. Soto-Concha, D. Morillo-Torres, J. W. Escobar, J. F. Mena-Reyes, and R. Linfati, "Mixed-Integer Linear Programming Models for the Vehicle Routing Problem with Release Times and Reloading at Mobile Satellites," *Mathematics*, vol. 13, no. 22, p. 3638, Nov. 2025.

[9] H. Huang, A. V. Savkin, and C. Huang, "Reliable path planning for drone delivery using a stochastic time-dependent public transportation network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 4941-4950, Aug. 2020.

[10] S. Jia, H. Cui, R. Chen, and Q. Meng, "Dynamic container drayage with uncertain request arrival times and service time windows," *Transp. Res. Part B Methodol.*, vol. 166, pp. 237-258, Dec. 2022.

[11] T. H. Jiang and Y. C. Chang, "Machine Learning-Enhanced Last-Mile Delivery Optimization: Integrating Deep Reinforcement Learning with Queueing Theory for Dynamic Vehicle Routing," *Appl. Sci.*, vol. 15, no. 21, p. 11320, Nov. 2025.

[12] H. Ercan, F. A. Ayaz, and H. Ulucan, "Effect of temperature, pressure and humidity on battery consumption in unmanned aerial vehicles," *J. Aviat.*, vol. 9, no. 1, pp. 5-12, Jan. 2025.

[13] C. C. Murray and A. G. Chu, "The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery," *Transp. Res. Part C Emerg. Technol.*, vol. 54, pp. 86-109, May 2015.

[14] C. Muli, S. Park, and M. Liu, "A comparative study on energy consumption models for drones," in *Global IoT Summit*, Cham: Springer International Publishing, 2022, pp. 199-210.

[15] H. Luo, Z. Liang, M. Zhu, X. Hu, and G. Wang, "Integrated optimization of unmanned aerial vehicle task allocation and path planning under steady wind," *PLoS One*, vol. 13, no. 3, p. e0194690, Mar. 2018.

[16] N. Akram, M. Khoshrangbaf, M. Challenger, and O. Dagdeviren, "Energy Consumption Modeling and Flight Time Analysis of Micro Drones," *IEEE Access*, vol. 13, pp. 109854-109865, 2025.

[17] J. Li, H. Modares, T. Chai, F. L. Lewis, and L. Xie, "Off-policy reinforcement learning for synchronization in multiagent graphical games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2434-2445, Oct. 2017.

[18] A. Ulazia, J. Sáenz, G. Ibarra-Berastegi, S. J. González-Rojí, and S. Carreno-Madinabeitia, "Global estimations of wind energy potential considering seasonal air density changes," *Energy*, vol. 187, Art. no. 115938, Nov. 2019.

[19] Y. Lyu, Z. Tang, Y. Li, B. Liu, M. Deng, and G. Wu, "A Multi-Agent Deep Reinforcement Learning Method with Diversified Policies for Continuous Location of Express Delivery Stations Under Heterogeneous Scenarios," *ISPRS Int. J. Geo-Inf.*, vol. 14, no. 12, p. 461, Dec. 2025.