Transportation Research Part E 210 (2026) 104825

Contents lists available at ScienceDirect

Transportation Research Part E

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

Air-ground multimodal transport planning for joint passenger mobility and parcel delivery: integration of drones, aircraft, and ground vehicles

Yimeng Zhang [a,b,c], Chenjie Yang [a], Haoning Xi [d,*], Songhan Peng [a], Junjie Yang [a], Mi Gan [a,b], Xiaobo Liu [a,b], Ruixue Ai [e]

[a] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China
[b] National and Local Joint Engineering Research Center of Integrated Transportation Intelligence, Southwest Jiaotong University, Chengdu, China
[c] Intelligent Comprehensive Transportation Key Laboratory of Sichuan Province, Southwest Jiaotong University, Chengdu, China
[d] Newcastle Business School, The University of Newcastle, Australia
[e] University of Oslo, Oslo, Norway

ARTICLE INFO

Keywords:
Multimodal transport
Air-ground transport planning
Joint passenger mobility and parcel delivery
Low-altitude economy
Adaptive large neighborhood search

ABSTRACT

Drones and electric Vertical Takeoff and Landing aircraft (eVTOLs) offer potential to improve transport efficiency and flexibility by circumventing ground transport constraints such as traffic congestion and infrastructure limitations. However, aerial vehicles alone often cannot meet transport demands, necessitating integration with ground-based systems. To address this challenge, this study investigates the Integrated Air-Ground Multimodal Transport Planning Problem (IAG-MTPP) for joint passenger mobility and parcel delivery. The IAG-MTPP integrates fixed-route and flexible-route vehicles, accommodating operational scenarios involving eVTOL aircraft, drones, and ground vehicles. The proposed IAG-MTPP is formulated as a Mixed-Integer Linear Programming (MILP) model that optimizes multimodal operations to minimize transshipment costs, delay penalties, and carbon emissions, while also incorporating capital costs and battery-energy constraints for electric air and ground vehicles. The model is tailored to the complexities of air-ground vehicle routing, considering transfer operations, inter-modal coordination, and routing flexibility. It enables switches between parcel-only, passenger-only, and integrated operations, providing adaptability to evolving transport demands. To solve large-scale instances, we customize an Adaptive Large Neighborhood Search (ALNS) algorithm with insertion, removal, and swap operators, feasibility checks, and an operator-selection scheme. Benchmarking against a commercial exact solver and multiple heuristic algorithms demonstrates the robustness and scalability of the proposed ALNS. The effectiveness and applicability of the proposed model and algorithm are validated through numerical experiments in scenarios including emergency rescue operations in Jiuzhaigou and urban transport in Chengdu, China. The results demonstrate that the integrated air-ground system optimizes multimodal routing while reducing transport costs and improving service coverage compared to alternatives. The proposed ALNS algorithm solves large-scale instances efficiently where commercial exact solvers fail, supporting real-world deployment in large networks and high-demand settings. This study offers insights and guidelines to support efficient resource allocation in the rapidly evolving low-altitude economy.

* Corresponding author.
E-mail address: Alice.Xi@newcastle.edu.au (H. Xi).

https://doi.org/10.1016/j.tre.2026.104825
Received 2 August 2025; Received in revised form 16 February 2026; Accepted 17 March 2026
Available online 31 March 2026

1. Introduction

In recent years, low-altitude air mobility has emerged as a transformative development in transport research, offering substantial advantages in urban logistics, emergency response, and regional connectivity (Ren and Wang, 2025; Huang et al., 2024; Xiang et al., 2024). The integration of drones, electric Vertical Takeoff and Landing (eVTOL) aircraft, and ground vehicles (GVs) into coordinated air-ground transport systems is quickly moving from theoretical exploration to practical application, driven by significant global investments and infrastructure initiatives. For example, the UK government has allocated over £20 million to accelerate the deployment of flying taxis and commercial drone services, aiming for operational readiness by 2028 (UK Government, 2025). This investment focuses on establishing the essential infrastructure and regulatory frameworks required to integrate eVTOLs and drones into existing transport networks. Supporting these objectives, the Future Flight and Land Infrastructure Programme (FFLIP) in Oxfordshire has demonstrated scalable, high-power charging infrastructure capable of simultaneously supporting eVTOLs, drones, and electric GVs (Petalite, 2025). Internationally, similar efforts are underway. In Taiwan, a partnership involving SkyDrive, 7A Drones, and the Industrial Technology Research Institute (ITRI) is pioneering eVTOL applications for emergency medical transport, significantly improving healthcare access to remote islands (SkyDrive Inc., 2025). Likewise, Michigan in the United States is actively developing Advanced Air Mobility (AAM) infrastructure to integrate drones and eVTOLs seamlessly, thereby enhancing regional connectivity and economic growth (Michigan Economic Development Corporation, 2024). Collectively, these developments highlight a clear global trajectory toward integrated multimodal air-ground transport.

Given the distinctive operational characteristics and complementary advantages of eVTOLs, drones, and ground vehicles, it is imperative to propose a model that strategically leverages their integration (Doo et al., 2021). eVTOLs offer rapid vertical takeoff and landing capabilities suitable for congested urban areas and challenging terrains, drones provide flexibility and efficiency for small-scale or urgent deliveries, and ground vehicles ensure stability, high payload capacity, and cost-effectiveness for large-scale operations. By systematically combining these transport modes, stakeholders can capitalize on their strengths to enhance overall system resilience, reliability, and responsiveness, especially in complex and dynamically changing environments. Previous research, e.g., Wang and Sheu (2019) and Zhai et al. (2024), highlights the rising importance of low-altitude operations within modern transport systems. Although prior studies have investigated air-ground vehicle routing in specific contexts, existing methodologies typically remain specialized, limiting their general applicability in real-world scenarios.

Additionally, providing joint passenger and parcel transport within an integrated air-ground transport framework is crucial under contemporary urban conditions, where demand patterns fluctuate significantly, and efficient resource utilization becomes critical (Miyoshi et al., 2025). Joint transport allows for the pooling of passenger and parcel flows, improving load factors, reducing empty mileage, and ultimately decreasing congestion and environmental impacts (Kellermann et al., 2020; Lu et al., 2022; Zhan et al., 2023). Such integrated operations not only generate economic efficiencies through optimized capacity use but also enhance service flexibility and responsiveness to urgent or time-sensitive demands, demonstrating considerable potential for sustainable urban development. Despite conceptual recognition, joint passenger-parcel transport has seen limited practical incorporation into transport planning models due to operational complexities and logistical constraints. Studies such as those by Sacramento et al. (2019) and Schermer et al. (2019) predominantly focus on drone-specific or parcel-exclusive scenarios, overlooking efficiencies achievable through integrated passenger-parcel operations. Similarly, the framework proposed by Thomas et al. (2024), though addressing multi-drone scenarios, does not resolve critical multimodal integration challenges. Consequently, existing approaches often fail to simultaneously optimize routing, scheduling, and resource allocation, leading to suboptimal operational performance, higher costs, and limited adaptability.

To address these research gaps and the inherent complexities of contemporary transport, this study investigates the Integrated Air-Ground Multimodal Transport Planning Problem (IAG-MTPP) for joint passenger mobility and parcel delivery. The IAG-MTPP framework is designed to integrate fixed-route and flexible-route vehicles, effectively accommodating diverse operational scenarios involving eVTOL aircraft, drones, and ground vehicles. We formulate the IAG-MTPP as a Mixed-integer Linear Programming (MILP) model aimed at optimizing fleet routing and scheduling while minimizing transshipment costs, delay penalties, and carbon emissions, explicitly incorporating capital-cost considerations and battery-energy constraints for electric air and ground vehicles. We develop a problem-specific Adaptive Large Neighborhood Search (ALNS) algorithm to handle the computational demands of large-scale, real-world scenarios. This algorithm includes tailored insertion, removal, and swap operators, rigorous feasibility checks, and an adaptive operator-selection mechanism, and its robustness is further demonstrated through benchmarking against 2-opt, Greedy, Genetic Algorithm, Simulated Annealing, and a commercial exact solver. Extensive numerical experiments, including case studies of emergency rescue missions in Jiuzhaigou and daily urban services in Chengdu, China, validate the robustness and practical applicability of the proposed model. The results offer managerial insights into vehicle flexibility, cost-speed trade-offs, and operational constraints, providing actionable guidance for stakeholders in the rapidly evolving low-altitude economy.

The remaining sections in this paper are organized as follows: Section 2 presents a brief review of the relevant literature and outlines the contributions of this study; Section 3 presents a detailed problem description, outlining the key features and assumptions of the joint transport system; Section 4 formulates the proposed IAG-MTPP as a MILP model; Section 5 introduces the customized ALNS algorithm; Section 6 presents the numerical experiments; Section 7 discusses policy implications; Section 8 presents the conclusions and outlook.

2. Literature review

We review the relevant literature on Transport planning for joint passenger mobility and parcel delivery (Section 2.1) and transport planning for low-altitude aircraft (Section 2.2). The goal is to identify existing research gaps and effectively position our contributions (Section 2.3).

2.1. Transport planning for joint passenger mobility and parcel delivery

Joint transport of passengers and parcels is increasingly recognized by academia and industry alike for its potential to improve resource allocation, reduce costs, and raise operational efficiency (Xi et al., 2023; He et al., 2025; Yang et al., 2024; Hörsting and Cleophas, 2023). Related studies have also considered individual passenger behavior in multimodal transport (Xi et al., 2025). Zheng et al. (2024) developed an economic model exploring the conversion of passenger aircraft to parcel use during the pandemic, highlighting economic flexibility and profitability from integrating passenger and parcel operations. Roa (2022) examined innovative aircraft designs for regional transport, emphasizing the integration of passenger and parcel operations to optimize existing transport infrastructure and services. Miyoshi et al. (2025) investigated dynamic fleet management in waterborne transport accommodating joint passenger and parcel demands, demonstrating through experiments that joint-purpose vessels substantially enhance efficiency and service quality compared to separate, dedicated-purpose vessels.

2.2. Transport planning for low-altitude aircraft and ground vehicles

Emerging technologies such as eVTOL aircraft and drones are reshaping urban transport, offering efficient and low-carbon alternatives to mitigate ground traffic congestion. eVTOL aircraft, central to current urban air mobility (UAM) solutions, exhibit substantial potential due to their vertical takeoff and landing capabilities coupled with intelligent path-planning technologies. For instance, the successful maiden flight of Shenzhen ZhiHang's manned flying saucer eVTOL and the technological breakthroughs demonstrated by the Joby S4 model highlight their advantages in reducing commute times and improving energy efficiency (Yan et al., 2024). The adoption of drones in transport has rapidly progressed, significantly innovating logistics systems. Mohamed and Mohamed (2025) reviewed the research on unmanned aerial vehicles (UAVs) in last-mile parcel delivery. They pointed out that compared with traditional truck delivery, UAVs can reduce carbon emissions by 71% and costs by 96.5%, with the cost per parcel ranging from 0.03 to 0.67. Moreover, the cost advantage is more significant in urban environments with high demand density. Bridgelall (2022) demonstrated how drones mitigate risks in hazardous goods transport by employing data-driven demand identification methods, thereby reducing human error and environmental hazards. Zandieh et al. (2023) emphasized drones' capabilities in routing optimization and real-time monitoring, enhancing the safety and reliability of transporting high-value goods.

In last-mile logistics, Athanasopoulos et al. (2024) showcased the efficiencies achieved through integrating drones and delivery bicycles, capitalizing on drones' ability to bypass traffic congestion and bicycles' agility in urban settings. Babaee Tirkolaee et al. (2025) further proposed the Traveling Salesman Problem with Drone and Bicycle (TSP-D-B). They constructed a MILP model to minimize the total travel time through the collaborative service of trucks, drones, and bicycles. In benchmark tests, it showed better time efficiency than the traditional TSP and TSP-D. Collaborative drone-GV logistics has been initially explored by Murray and Chu (2015) through the Flying Sidekick Traveling Salesman Problem (FSTSP), employing MILP and heuristic algorithms foundational to drone-truck collaboration. Agatz et al. (2018) extended this research with the Traveling Salesman Problem with Drone (TSP-D), analyzing time efficiency and proposing robust integer programming models combined with heuristic solutions. Su et al. (2025) innovatively proposed a freight multimodal transport system integrating buses and drones (FMTP-BD). In this system, buses transport parcels to lockers at bus stops, and then drones complete the final delivery. Their Branch-Price-and-Benders-Cut algorithm can achieve more than 6% cost savings compared with sequential optimization, while expanding the service coverage of drones.

Further advancements include Gonzalez-R et al. (2020)'s Truck-Drone Team Logistics (TDTL) model, which efficiently expanded drone operations to handle multiple customer visits per trip using an iterative greedy algorithm. Li et al. (2022) explored multi-objective truck-drone routing problems with flexible time windows, optimizing both delivery costs and customer satisfaction. Meng et al. (2023) developed the Multi-Visit Drone Routing Problem for Pickup and Delivery (MDRP-PD), effectively managing multiple pickups and deliveries per drone flight with a two-stage heuristic algorithm that outperformed benchmark methods.

Algorithmic developments also remain active. Mulumba et al. (2024) investigated drone-assisted pickup and delivery operations using an ALNS metaheuristic algorithm, validating its effectiveness through extensive experimentation. Sacramento et al. (2019) optimized the Vehicle Routing Problem with Drones (VRP-D), achieving cost reduction and workload balancing via adaptive heuristic methods. Additionally, Schermer et al. (2019) introduced hybrid heuristic solutions combining variable neighborhood search and Tabu search, significantly enhancing computational efficiency in drone-vehicle routing problems.

2.3. Research gaps and our contributions

Despite rapid development in drone–ground collaboration, existing literature remains constrained by several interrelated limitations. Existing literature on drone–ground collaboration focuses on the traditional truck-and-drone paradigm for parcel delivery (Murray and Chu, 2015; Agatz et al., 2018; Gonzalez-R et al., 2020; Luo et al., 2021; Schermer et al., 2019; Sacramento et al., 2019; Meng et al., 2023; Mulumba et al., 2024; Babaee Tirkolaee et al., 2025). These studies typically restrict the fleet to a single GV assisted by one or more drones and address *parcel-only* demand (Murray and Chu, 2015; Agatz et al., 2018; Luo et al., 2021; Mohamed

Table 1
Key studies on integrated drone, ground and air vehicle operations.

Reference	Fleet composition			Request	Priority	Model structure	Solution approach	
	Drone(s)	GV(s)	eVTOL					
Murray and Chu (2015)	1	1	-	Parcel	—	Undefined	MILP	Constructive heuristic
Agatz et al. (2018)	1	1	-	Parcel	—	Undefined	IP	Dynamic programming
Gonzalez-R et al. (2020)	1	1	-	Parcel	—	Undefined	MIP	Simulated annealing
Luo et al. (2021)	*m*	1	-	Parcel	✓	Undefined	MILP	Hybrid MOA
Schermer et al. (2019)	*m*	*n*	-	Parcel	—	Undefined	MILP	VNS–TS
Meng et al. (2023)	*m*	1	-	Parcel	—	Undefined	MILP	Two-stage heuristic
Wang and Sheu (2019)	*m*	*n*	-	Parcel	—	Undefined	MIP	Branch-and-price
Sacramento et al. (2019)	*m*	*n*	-	Parcel	—	Undefined	MIP	ALNS
Mulumba et al. (2024)	1	1	-	Parcel	—	Undefined	MILP	Constructive heuristic
Thomas et al. (2024)	*m*	1	-	Parcel	—	Undefined	MILP	RF–RRO heuristic
Zhai et al. (2024)	1	1	-	Parcel	—	Undefined	MILP	MATSP-D
Babaee Tirkolaee et al. (2025)	1	1	-	Parcel	—	Pre-/Undefined	MILP	CPLEX
Su et al. (2025)	1	*n*	-	Parcel	—	Undefined	MILP	BPBC
This study	*m*	*n*	*m*	Passenger+ Parcel	✓	Pre-/Undefined	MILP	ALNS

Notes. MILP = Mixed-integer Linear Programming; MIP = Mixed-integer Programming; IP = Integer Programming; ALNS = Adaptive Large Neighborhood Search; VNS–TS = Variable-Neighborhood Search with Tabu Search; RF–RRO = Route First–relocate/Re-optimize; BPBC = Branch-price-and-cut; GV = Ground Vehicle. "1" denotes a single vehicle of that type; m, n indicate multiple (unspecified) vehicles.

and Mohamed, 2025; Bridgelall, 2022; Zandieh et al., 2023; Su et al., 2025). Few studies, however, integrate eVTOL aircraft or enables the *simultaneous* movement of passengers and goods. Responding to these gaps, the present study models a heterogeneous fleet comprising GVs, drones, and eVTOLs and allows joint passenger–parcel requests, thereby enhancing realism for both dense urban corridors and remote regions. As shown in Table 1, in contrast to the existing literature, this study orchestrates a heterogeneous fleet of GVs, drones, and eVTOLs, handles joint passenger-parcel demand, supports both predefined and generated routes with priority rules, and proposes a general MILP formulation with a customized ALNS to manage the ensuing multimodal complexity. In addition, whereas previous work often relies on a single or undefined routing structure and omits service-priority rules, our framework accommodates both predefined and generated routes while enforcing priority constraints, enabling greater operational flexibility. Finally, we proposed a general mixed-integer programming formulation and a customized adaptive ALNS meta-heuristic, extending beyond existing ALNS and VNS-TS applications to tackle the intricacies of multimodal, multi-purpose, multi-vehicle routing across diverse operational contexts.

To bridge these gaps, this study investigates the IAG-MTPP and proposes a flexible optimization model for the coordinated deployment of eVTOL aircraft, drones, and GVs. The proposed model captures fleet-wide constraints, vertiport availability, payload limits, and synchronized time windows via a MILP model while treating mode-specific characteristics as modular parameters that can be toggled on or off. Thus, the proposed model is readily transferable across diverse operating environments and use cases. Beyond improving resource allocation and spatio-temporal efficiency, the proposed model provides analytical foundations for airspace deconfliction, energy-consumption management, and emergency-logistics planning, which are key issues for stakeholders in the burgeoning low-altitude economy.

This study makes the following key contributions: i) We cast the proposed IAG-MTPP as a MILP model that jointly schedules heterogeneous eVTOL, drone, and ground-vehicle fleets to meet simultaneous passenger and parcel demand, aiming to minimize total costs, including transshipment cost, delay penalties, and carbon emissions under both fixed-route and on-demand service regimes; ii) By integrating passenger and parcel flows within a unified framework, the proposed IAG-MTPP can switch between parcel-only, passenger-only, and combined operations without altering its mathematical formulations; iii) We customize an ALNS equipped with tailored insertion, removal, and swap operators, selected via an adaptive weight-update scheme, that delivers near-optimal solutions for realistically sized problem instances within practical runtimes; iv) Experiments on earthquake-response logistics in Jiuzhaigou and routine urban services in Chengdu, China, show that integrated air-ground operations cut total cost and delay markedly relative to single-mode baselines, and illuminate trade-offs among vehicle flexibility, speed, and costs.

3. Problem statement

In this study, we tackle contemporary logistics and emergency-response challenges by coordinating air and ground transport within a unified optimization framework. We introduce the Integrated Air–Ground Multimodal Transport Planning Problem (IAG-MTPP), where a heterogeneous fleet comprising eVTOL aircraft, unmanned drones, and ground vehicles (GVs) serves both passenger trips and parcel deliveries. The framework is designed to be modular and context-aware: specific vehicle classes, request types, or transport modes can be switched on or off to reflect the available resources and service scope, for example, excluding drones when they are not deployed or omitting parcel requests in passenger-only operations. The IAG-MTPP jointly optimizes vehicle routing, scheduling, and intermodal transfer decisions across the entire fleet to minimize total generalized operating cost, including transshipment costs incurred when loads are transferred between modes, delay penalties for services that violate promised time windows, and carbon-emission-related costs. Mathematical notations used throughout the paper are summarized in Table 2.

Table 2
Mathematical notation.

Sets	Description
W	Set of transport modes, indexed by w.
R	Set of customer requests, indexed by r.
N	Set of network stations, denoted by i or j.
O	Set of regular origin or destination stations for vehicles, $O \subseteq N$.
$\overline{O}$	Set of dummy origin or destination stations, $\overline{O} \subseteq N$.
P	Set of pickup locations, $P \subseteq N$.
D	Set of delivery locations, $D \subseteq N$.
T	Set of transshipment stations, $T \subseteq N$.
$T_{w_1}^{w_2}$	Set of stations enabling mode transfers between w_1 and w_2.
K	Set of available vehicles, labeled by k or l. K_w includes vehicles vehicles of transport mode mode w, while K_{fix} consists of vehicles with predefined operational routes.
A	Set of directed arcs. An arc from station i to j (where $i, j \in N$) is denoted by $(i, j) \in A$.
A_p	Set of pickup arcs; $(i, j) \in A_p$ implies $i \in P$.
A_d	Set of delivery arcs; $(i, j) \in A_d$ implies $j \in D$.
A_w	Feasible arc set for mode w, and A_{fix}^k denotes the fixed arcs assigned to vehicle $k \in K_{\text{fix}}$.
A_{e}, A_{u}, A_{g}	Sets of arcs exclusively available to eVTOLs, drones, and ground vehicles, respectively.
Parameters	
u_k	The load capacity of vehicle k.
q_r	The load of request r, where load refers to parcel weight or the effective weight equivalent of a passenger.
w_r	Weight of request r.
$p(r)$	Pickup location of request r ($p(r) \in P$).
$d(r)$	Delivery location of request r ($d(r) \in D$).
τ_{ij}^k	The duration (in hours) required by vehicle k to traverse arc (i, j).
$[a_{p(r)}, b_{p(r)}]$	Earliest and latest allowable pickup times for request r.
$[a_{d(r)}, b_{d(r)}]$	Earliest and latest allowable delivery times for request r.
$[a_i^k, b_i^k]$	Permissible operating time interval for fixed vehicle k at station i.
t''^k_i	Duration (in hours) of loading/unloading operations for vehicle k at station i.
v_k	Travel speed of vehicle k (in km/h).
d_{ij}^k	Vehicle k's route distance between stations i and j (km).
e_k	The CO_2 emission factor (ton/kg/km) of vehicle $k \in K$.
c_k^1	Unit transport cost of vehicle k in terms of time and weight (hour·kg).
$c_k^{1'}$	Unit transport cost of vehicle k in terms of distance and weight (km·kg).
c_k^2	Cost incurred per kg for loading or unloading operations.
c_k^3	Unit storage cost based on time and weight (kg·h).
c_k^4	Unit carbon tax (per ton basis).
c_k^5	Penalty cost incurred per kilogram per hour of delay.
q_{total}	Total volume.
Λ	Weighted vehicle capacity.
Ω	Load factor.
c_{dq}	Cost per unit kg-kilometer.
c_{total}	Total transport cost.
d_{total}	Total transport distance.
α_w	Proportion of requests served by transport mode w.
$o(k)$	Origin station of vehicle k ($o(k) \in O$).
$o'(k)$	Destination station of vehicle k ($o'(k) \in O$ or $\overline{O}$).
M	A sufficiently large constant.
B_k	Battery capacity (kWh) of vehicle k.
α_k	Energy consumption rate per unit travel time (kWh/h).
β_k	Energy consumption rate per unit travel distance (kWh/km).
Γ	Set of charging stations.
$C_{k,i}$	Accumulated energy consumption of vehicle k upon arrival at node i.
Decision variables	
x_{ij}^k	Route assignment variable for vehicle k; 1 if it uses arc (i, j), 0 otherwise.
y_{ij}^{kr}	Request assignment variable for vehicle k and request r; 1 if vehicle k transports request r using arc (i, j), 0 otherwise.
z_{ij}^k	Precedence variable for vehicle k; 1 if i is visited before j, regardless of whether the visit is immediate; 0 otherwise.
s_{ir}^{kl}	Transshipment variable; 1 if request r is transferred from vehicle k to vehicle l at station i, with $k \neq l$, 0 otherwise.
h_r	A binary decision variable that equals 1 if request r is served, and 0 otherwise.
t_i^{kr}	Time when request r, transported by vehicle k, reaches station i.
t'^{kr}_i	Time when vehicle k starts servicing request r at station i.
$\bar{t}_i^{kr}$	Time when vehicle k finishes servicing request r at station i.
t_i^k	Vehicle k's arrival time at station i.
t'^k_i	The time of the last service initiation by vehicle k at station i.
$\bar{t}_i^k$	Time when vehicle k leaves station i.
t_{ki}^{wait}	Time spent by vehicle k waiting before service at station i.
t_r^{delay}	Delivery delay duration for request r.

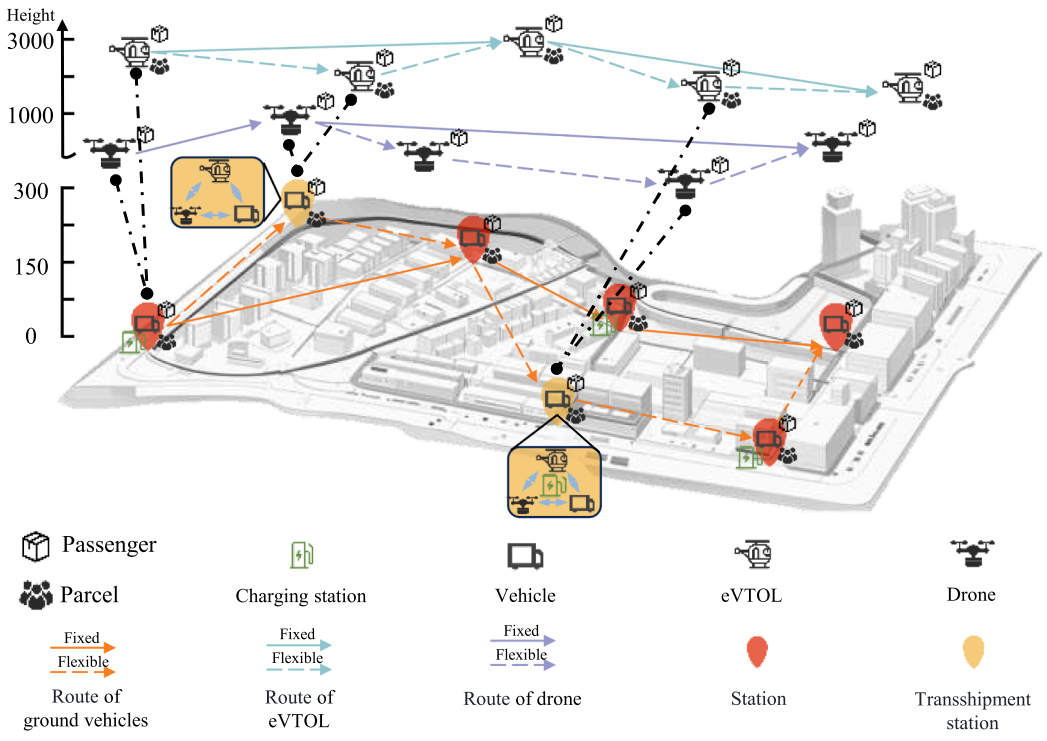


Fig. 1. Illustration of the Integrated Air-Ground Multimodal Transport Planning Problem (IAG-MTPP).

Table 2 shows the notation used in this study. Fig. 1 illustrates an integrated multimodal transport network for the IAG-MTPP, consisting of three transport modes denoted by set W: eVTOL aircraft (mode e), drones (mode u), and ground vehicles (mode g). Each mode $w \in W$ has an associated subset of vehicles $K_w \subseteq K$ in the fleet, where K is the set of all vehicles. Each vehicle k in mode w has a capacity u_k. All vehicles begin their routes at a designated origin station $o(k)$ (e.g. a depot or vertiport) and terminate at a destination station $o'(k)$ after completing their service. The fundamental rationale for integrating these modes is to leverage their complementary strengths while mitigating individual shortcomings. For instance, eVTOLs can rapidly cover medium- to long-distance aerial segments for urgent passenger trips or time-critical parcel deliveries; drones excel at point-to-point last-mile deliveries of lightweight, high-priority parcels to dispersed or hard-to-reach locations; and ground vehicles can carry bulkier loads or provide reliable service in areas where aerial options are impractical, e.g. due to no-fly zones or lack of vertiport infrastructure. By assigning each task to the most suitable mode, this heterogeneous fleet achieves a synergistic allocation of tasks that plays to each vehicle type's strengths. The modes also differ in operational flexibility: eVTOLs are capable of multi-stop tours (visiting multiple stations in one route, given a network of vertiports to land and take off), whereas drones typically operate in a point-to-point fashion (launching from and returning to a base or mobile ground station for each delivery). This difference implies that drones can be scheduled on independent missions in parallel with other vehicles, improving system-wide throughput. The proper infrastructure is assumed to be in place for each mode, e.g. a sufficient network of vertiports for eVTOL landing, and simple landing zones for drone operations.

Each customer request $r \in R$ is represented by a tuple of attributes capturing its spatial and temporal requirements and size. Formally, we define:

$$r = \left\langle p(r), d(r), [a_{p(r)}, b_{p(r)}], [a_{d(r)}, b_{d(r)}], q_r, w_r \right\rangle, \tag{1}$$

where $p(r)$ and $d(r)$ denote the request's pickup (origin) and delivery (destination) locations, respectively. The intervals $[a_{p(r)}, b_{p(r)}]$ and $[a_{d(r)}, b_{d(r)}]$ specify the allowable time windows for pickup and delivery: vehicle service for request r must begin after time $a_{p(r)}$ and, ideally, be completed by the deadline $b_{d(r)}$. Each request also has a size characterized by its quantity q_r and weight w_r (for passenger requests, these correspond to the number of passengers and their total weight). If a request r is delivered later than its specified delivery window (i.e., after $b_{d(r)}$), a delay penalty is incurred; this is captured by a nonnegative delay variable t_r^{delay} in the model, which contributes to the objective cost proportional to the lateness of the delivery. In this problem, certain requests may remain unserved. This reflects the operational realities of hybrid passenger–parcel systems, where resource availability, time-window constraints, battery limitations, or payload restrictions may prevent the system from feasibly serving all incoming requests.

The multimodal transport network is modeled as a directed graph $G = (N, A)$. The node set N represents all stations or locations relevant to the problem, and the directed arc set A represents all feasible connections between these stations. The node set N

encompasses several types of locations: pickup points (P), where customers or parcels are collected; delivery points (D), where they are dropped off; and transshipment hubs (T), where transfers between modes can occur. In addition, N includes any necessary origin/depot and destination nodes for vehicle routes. We denote by O the set of start and end points of vehicles, and by $\overline{O}$ a set of dummy nodes representing artificial start and end points. Every customer request $r \in R$ is associated with one pickup location in P and one delivery location in D that must be visited to serve that request, and naturally, each request's pickup must precede its delivery in any valid routing. The transshipment hubs $T \subset N$ are special nodes designated for mode transfer: if a request is to be handed from one vehicle to another (e.g., from a GV to a drone, or from an eVTOL to a GV), this transfer can only take place at a node in T. We further denote by $T_{w_1}^{w_2} \subseteq T$ the subset of transshipment hubs that facilitate a transfer from mode w_1 to mode w_2 (for example, a vertiport might allow transfers from an eVTOL to a truck, etc.). An inter-modal transfer at such a node implies that one vehicle of mode w_1 drops off the passenger or parcel at the hub and another vehicle of mode w_2 picks it up to continue the journey.

The directed arc set A defines all permissible movements between nodes in N. Each arc is an ordered pair (i, j) representing travel from node i to node j. To capture different kinds of movements and operations in a multimodal setting, the arc set is categorized into the following subsets: i) Pickup arcs (A_p) that terminate at a pickup location (node $j \in P$), where a vehicle arrives to pick up a customer or parcel; ii) Delivery arcs (A_d) that terminate at a delivery location (node $j \in D$), where a delivery operation is carried out; iii) Mode-specific arcs (A_w) that are traversable only by vehicles of mode w. Note that A_e, A_u, and A_g denote the sets of arcs exclusively available to eVTOLs, drones, and ground vehicles, respectively; iv) Fixed-route arcs (A_{fix}^k), which is a set of predefined arcs associated with a specific fixed-route vehicle $k \in K_{\text{fix}} \subseteq K$, representing the unchanging operational path of vehicles that follow a fixed schedule or route (e.g., a shuttle service with a predetermined route).

We define various parameters that characterize vehicle capabilities, travel times, and costs in the system. Each vehicle $k \in K$ has a maximum capacity u_k that it cannot exceed when carrying requests. Each customer request $r \in R$ has an associated demand quantity q_r (and possibly weight w_r if separate from quantity) that consumes capacity when loaded onto a vehicle. The time windows $[a_{p(r)}, b_{p(r)}]$ and $[a_{d(r)}, b_{d(r)}]$ for each request are treated as hard or soft constraints. For each vehicle k and station i, a loading/unloading time ${t''}_i^k$ may be specified, representing the time required to pick up or drop off a request at station i using vehicle k. Travel times between nodes are derived from known distances and vehicle speeds: let d_{ij}^k be the distance (km) between stations i and j for vehicle k, and v_k the speed (km/h) of vehicle k, then travel times can be computed or are given as parameters. Each vehicle also has an associated emission rate e_k, representing the CO_2 emissions (ton) per kg per km of vehicle $k \in K$, allowing the model to account for environmental impact, and unit cost coefficients c_k^n for different cost components, including transport cost per hour (c_k^1), transport cost per km ($c_k^{1'}$), loading/unloading cost (c_k^2), storage cost (c_k^3), carbon tax coefficient (c_k^4), and delay penalty (c_k^5).

The key decision variables of the proposed IAG-MTPP include the routing variable x_{ij}^k, a binary routing variable equal to 1 if vehicle k is assigned to arc $(i, j) \in A$ (and 0 otherwise); the request assignment variable y_{ij}^{kr}, a binary variable equal to 1 if request r is transported by vehicle k on arc (i, j) (and 0 otherwise); the precedence variable z_{ij}^k, a binary indicator equal to 1 if station i precedes station j in the route of vehicle k (and 0 otherwise); and the transshipment variable s_{ir}^{kl}, a binary variable equal to 1 if request r is transferred at station i from vehicle k to a different vehicle l; 0 otherwise. Additionally, several time-continuous decision variables are included to schedule operations in compliance with time window constraints.

4. Mathematical formulations

This study formulates the proposed IAG-MTPP as a MILP model aimed at optimizing the integrated transport of passengers and parcels within and between urban and inter-city networks using eVTOLs, drones, and GVs. It is worth noting that the proposed model is adaptable to various contexts involving different vehicle types, request profiles, and transport networks. For instance, when drones are unavailable or parcel delivery is not required, the model can be readily adjusted by excluding the corresponding vehicles or requests. In our model, two objectives are defined in a hierarchical manner, where Objective (2) has a higher priority than Objective (3). Objective (2) (F) maximizes the number of served requests, while Objective (3) (F') minimizes the total operational cost. The objective function 3 comprises transit costs, transfer costs, storage costs, carbon emission taxes, and delay penalties. To ensure commensurability, all cost components are expressed in a unified monetary unit (CNY). The model simultaneously determines optimal vehicle routing and scheduling for both passenger and parcel requests, taking into account constraints related to spatial routing, vehicle capacity, time windows, and transfer operations.

$$\max F = \sum_{r \in R} h_r \tag{2}$$

$$\min \underbrace{F'}_{\text{Total generalized costs}} = \underbrace{\sum_{k \in K} \sum_{(i,j) \in A} \sum_{r \in R} (c_k^1 \tau_{ij}^k + c_k^{1'} d_{ij}^k) q_r y_{ij}^{kr}}_{\text{transit costs}}$$

$$+ \underbrace{\sum_{k,l \in K, k \neq l} \sum_{r \in R} \sum_{i \in T} (c_k^2 + c_l^2) q_r s_{ir}^{kl} + \sum_{k \in K} \sum_{(i,j) \in A_p} \sum_{r \in R} c_k^2 q_r y_{ij}^{kr} + \sum_{k \in K} \sum_{(i,j) \in A_d} \sum_{r \in R} c_k^2 q_r y_{ij}^{kr}}_{\text{transfer costs}}$$

$$+\underbrace{\sum_{k,l\in K,k\neq l}\sum_{r\in R}\sum_{i\in T}c_k^3 q_r s_{ir}^{kl}(t'^{lr}_i-\bar{t}_i^{kr})+\sum_{k\in K}\sum_{(i,j)\in A_p}\sum_{r\in R}c_k^3 q_r y_{ij}^{kr}(t'^{kr}_i-a_{p(r)})}_{\text{storage costs}}$$

$$+\underbrace{\sum_{k\in K}\sum_{(i,j)\in A}\sum_{r\in R}c_k^4 e_k q_r d_{ij}^k y_{ij}^{kr}}_{\text{carbon emission taxes}}+\underbrace{\sum_{r\in R}c_k^5 q_r t_r^{\text{delay}}}_{\text{delay penalty costs}} \tag{3}$$

Constraints (4)-(11) represent standard constraints in vehicle routing problems. Constraints (4) and (5) ensure vehicles begin and end their routes at designated stations. These constraints specifically govern routes for eVTOLs and GVs, whereas drone operations are treated as independent fleets due to their distinct operational characteristics. Subtour elimination constraints (6)-(8) effectively prevent infeasible routing loops, providing a compact yet rigorous representation of routing constraints. Constraints (9) and (10) guarantee that each request's demand (either parcels or passengers) is picked up and delivered at the appropriate stations, while Constraints (11) enforce the vehicle capacity limitations.

$$\sum_{j\in N}x_{\bar{o}(k)j}^k\leqslant 1\quad\forall k\in K_{\text{e\&dr}} \tag{4}$$

$$\sum_{j\in N}x_{\bar{o}(k)j}^k=\sum_{j\in N}x_{j\overline{o'}(k)}^k\quad\forall k\in K_{\text{e\&dr}} \tag{5}$$

$$x_{ij}^k\leqslant z_{ij}^k\quad\forall i,j\in N,\ \forall k\in K_{\text{e\&dr}} \tag{6}$$

$$z_{ij}^k+z_{ji}^k=1\quad\forall i,j\in N,\ \forall k\in K_{\text{e\&dr}} \tag{7}$$

$$z_{ij}^k+z_{jp}^k+z_{pi}^k\leqslant 2\quad\forall i,j,p\in N,\ \forall k\in K_{\text{e\&dr}} \tag{8}$$

$$\sum_{k\in K}\sum_{j\in N}y_{p(r)j}^{kr}=1\quad\forall r\in R \tag{9}$$

$$\sum_{k\in K}\sum_{i\in N}y_{id(r)}^{kr}=1\quad\forall r\in R \tag{10}$$

$$\sum_{r\in R}q_r y_{ij}^{kr}\leqslant u_k x_{ij}^k\quad\forall(i,j)\in A,\ \forall k\in K \tag{11}$$

Constraints (12) enforce that a transshipment operation at the same station is performed at most once within the planning horizon. Constraints (13) explicitly prohibit intra-vehicle transshipment for any vehicle k.

$$\sum_{j\in N}y_{ji}^{kr}+\sum_{j\in N}y_{ij}^{lr}\leqslant s_{ir}^{kl}+1\quad\forall r\in R,\ \forall i\in T,\ \forall k,l\in K \tag{12}$$

$$s_{ir}^{kk}=0\quad\forall r\in R,\ \forall i\in T,\ \forall k\in K \tag{13}$$

Constraints (14) ensure that a request is marked as served only if at least one vehicle transports it on some arc, while Constraints (15) ensure that no transport arc is assigned to a request that is not served.

$$h_r\leq\sum_{k\in K}\sum_{(i,j)\in A}y_{ij}^{kr},\quad\forall r\in R \tag{14}$$

$$\sum_{k\in K}\sum_{(i,j)\in A}y_{ij}^{kr}\leq Mh_r,\quad\forall r\in R \tag{15}$$

Constraints (16)–(21) are designed to enforce flow conservation principles for both vehicles and service requests. Constraint (16) ensures vehicle flow conservation, meaning that each vehicle's incoming and outgoing flows at any given station are balanced. Constraints (17)-(20) handle the flow conservation of requests. Specifically, Constraints (17) addresses flow conservation at regular stations, ensuring that parcels or passengers entering a station also leave it. Constraint (18) manages request flow conservation explicitly at transshipment stations, accounting for mode changes. Constraints (19) and (20) account for special cases where a vehicle passes through a transshipment station without transferring the request, due to handling other requests. These constraints ensure request flows are accurately maintained even when a vehicle traverses a station without conducting transfers. Constraint (21) integrates the request flow variable y_{ij}^{kr} with the vehicle flow variable x_{ij}^k, ensuring a vehicle must traverse an arc for it to carry a corresponding request on that arc.

$$\sum_{j\in N}x_{ij}^k-\sum_{j\in N}x_{ji}^k=0\quad\forall k\in K_{\text{e\&dr}},\ \forall i\in N\setminus\bar{o}(k),\overline{o'}(k) \tag{16}$$

$$\sum_{j\in N}y_{ij}^{kr}-\sum_{j\in N}y_{ji}^{kr}=0\quad\forall k\in K,\ \forall r\in R,\ \forall i\in N\setminus T,p(r),d(r) \tag{17}$$

$$\sum_{k\in K}\sum_{j\in N}y_{ij}^{kr}-\sum_{k\in K}\sum_{j\in N}y_{ji}^{kr}=0\quad\forall k\in K,\ \forall r\in R,\ \forall i\in T\setminus p(r),d(r) \tag{18}$$

$$\sum_{j\in N}y_{ij}^{kr}-\sum_{j\in N}y_{ji}^{kr}\leqslant\sum_{l\in K}s_{ir}^{lk}\quad\forall k\in K,\ \forall r\in R,\ \forall i\in T\setminus p(r),d(r) \tag{19}$$

$$\sum_{j\in N} y_{ji}^{kr} - \sum_{j\in N} y_{ij}^{kr} \leqslant \sum_{l\in K} s_{ir}^{kl} \quad \forall k \in K,\ \forall r \in R,\ \forall i \in T \setminus p(r), d(r) \tag{20}$$

$$y_{ij}^{kr} \leqslant x_{ij}^{k} \quad \forall (i,j) \in A,\ \forall k \in K,\ \forall r \in R \tag{21}$$

Constraints (22)-(24) explicitly address specific operational characteristics within the proposed transport system. Constraints (22) ensure vehicles are restricted to suitable routes; for example, drones are prohibited from routes unsuitable for aerial travel. Constraints (23) enforce predefined routes for vehicles operating on fixed schedules. Constraints (24) ensure that transshipment operations occur at the correct designated stations, thereby enabling mode-specific transfers and maintaining operational coherence within the network.

$$x_{ij}^{k} = 0 \quad \forall k \in K_w,\ \forall (i,j) \in A \setminus A_w,\ \forall w \in W \tag{22}$$

$$x_{ij}^{k} = 0 \quad \forall k \in K_{\text{fix}},\ \forall (i,j) \in A \setminus A_{\text{fix}}^{k} \tag{23}$$

$$s_{ir}^{kl} = 0 \quad \forall k \in K_{w_1},\ \forall l \in K_{w_2},\ \forall i \in T \setminus T_{w_1}^{w_2},\ \forall r \in R,\ \forall w_1, w_2 \in W \tag{24}$$

Temporal constraints are addressed by constraints (25)-(29). Specifically, Constraint (25) ensures service activities begin only after vehicle arrivals. Constraint (26) guarantees that service activities conclude according to their durations. Constraints (27) ensure departures occur after service completion, while Constraints (28) prevent vehicles from arriving later than the requests' scheduled times. Constraints (29) precisely define the last service initiation time for vehicles.

$$t_i^{kr} \leqslant t'^{kr}_i \quad \forall i \in N,\ \forall k \in K,\ \forall r \in R \tag{25}$$

$$t'^{kr}_i + t''^{kr}_i \sum_{j\in N} y_{ij}^{kr} \leqslant \bar{t}_i^{kr} \quad \forall i \in N,\ \forall k \in K_{\text{e\&dr}},\ \forall r \in R \tag{26}$$

$$\bar{t}_i^{k} \geqslant \bar{t}_i^{kr} \quad \forall i \in N,\ \forall k \in K_{\text{e\&dr}},\ \forall r \in R \tag{27}$$

$$t_i^{k} \leqslant t_i^{kr} \quad \forall i \in N,\ \forall k \in K_{\text{e\&dr}},\ \forall r \in R \tag{28}$$

$$t'^{k}_i \geqslant t'^{kr}_i \quad \forall i \in N,\ \forall k \in K_{\text{e\&dr}},\ \forall r \in R \tag{29}$$

Constraints (30)-(31) ensure temporal consistency between travel distances and vehicle speeds. Constraint (30) addresses basic travel time requirements, whereas Constraint (31) precisely handles waiting periods at stations to prevent unintended delays. Time windows at pickup and fixed stations are managed by Constraints (32) and (33), respectively.

$$\bar{t}_i^{k} + \tau_{ij}^{k} - t_j^{k} \leqslant M(1 - x_{ij}^{k}) \quad \forall (i,j) \in A,\ \forall k \in K_{\text{e\&dr}} \tag{30}$$

$$\bar{t}_i^{k} + \tau_{ij}^{k} - t_j^{k} \geqslant -M(1 - x_{ij}^{k}) \quad \forall (i,j) \in A,\ \forall k \in K_{\text{e\&dr}} \tag{31}$$

$$t'^{kr}_{p(r)} \geqslant a_{p(r)} y_{ij}^{kr},\ \bar{t}_{p(r)}^{kr} \leqslant b_{p(r)}(y_{ij}^{kr} + M(1 - y_{ij}^{kr})) \quad \forall (i,j) \in A, \forall r \in R,\ \forall k \in K \tag{32}$$

$$t_i^{kr} \geqslant a_i^k y_{ij}^{kr},\ \bar{t}_i^{kr} \leqslant b_i^k (y_{ij}^{kr} + M(1 - y_{ij}^{kr})) \quad \forall (i,j) \in A,\ \forall r \in R,\ \forall k \in K_{\text{fix}} \tag{33}$$

Constraints (34) synchronize vehicle arrivals at transfer stations, ensuring feasible transshipment operations. Waiting is permitted for vehicle l if it reaches the station before vehicle k departs, thus accommodating operational flexibility. Constraints (35) and (36) explicitly quantify waiting and delay times, respectively.

$$\bar{t}_i^{kr} - t'^{lr}_i \leqslant M(1 - s_{ir}^{kl}) \quad \forall r \in R,\ \forall i \in T,\ \forall k,l \in K,\ k \neq l \tag{34}$$

$$t_{ki}^{\text{wait}} \geqslant t'^{k}_i - t_i^{k} \quad \forall i \in N,\ \forall k \in K_{\text{e\&dr}} \tag{35}$$

$$t_r^{\text{delay}} \geqslant (\bar{t}_{d(r)}^{kr} - b_{d(r)}) \sum_{i\in N} y_{id(r)}^{kr} \quad \forall r \in R,\ \forall k \in K \tag{36}$$

Constraints (37)–(40) explicitly govern the energy feasibility of eVTOLs, drones, and electric ground vehicles in the transport system. Constraints (37)–(38) describe the evolution of energy usage along selected arcs. Constraints (39) describe instantaneous battery replenishment at designated charging stations, and Constraints (40) ensure that energy consumed remains within the available battery capacity.

$$C_{k,j} \geq C_{k,i} + \alpha_k \cdot \frac{d_{ij}^{k}}{v_k} + \beta_k\, d_{ij}^{k} - M(1 - x_{ij}^{k}) \quad \forall k \in K, \forall (i,j) \in A \tag{37}$$

$$C_{k,j} \leq C_{k,i} + \alpha_k \cdot \frac{d_{ij}^{k}}{v_k} + \beta_k\, d_{ij}^{k} + M(1 - x_{ij}^{k}) \quad \forall k \in K, \forall (i,j) \in A \tag{38}$$

$$C_{k,i} = 0 \quad \forall k \in K, \forall i \in \Gamma \tag{39}$$

$$C_{k,i} \leq B_k \quad \forall k \in K, \forall i \in N \tag{40}$$

5. Customized adaptive large neighborhood search algorithm

Integrating eVTOLs, drones, and GVs into multimodal transport planning introduces several significant challenges, particularly due to the increased problem scale and the inherent complexity of operational constraints such as vehicle capacities, time windows,

and transshipment rules. Exact solution methods are typically computationally intensive and thus become impractical for large-scale applications. To effectively address these issues, this study employs a customized ALNS algorithm, which leverages an adaptive operator-selection mechanism informed by historical performance. The ALNS approach efficiently handles large-scale problems, balancing exploration and exploitation to rapidly generate near-optimal solutions.

5.1. Algorithm framework

Algorithm 21 details the input, output, and iterative search process of the customized ALNS algorithm. The inputs comprise two primary datasets: a vehicle set, containing vehicle-specific details such as capacities, speeds, and origin-destination information; and a request set, specifying transport demands, including pickup and delivery locations, parcel weights or passenger counts, time windows, and priority levels. The algorithm's output is a near-optimal solution derived through the customized ALNS algorithm.

Initially, the algorithm employs a greedy insertion operator to generate the initial solution X_{initial}, sequentially evaluating each request to identify insertion positions that minimize the associated costs. If there remain unprocessed requests within the request pool R_{pool}, a random removal operator is activated. Requests are randomly selected for removal and subsequently reinserted via greedy insertion until the request pool is cleared.

During the iterative search phase, solutions are explored within a predefined iteration limit guided by the optimization objective. Every s iterations, operator weights are recalibrated based on historical performance, specifically reflecting improvements achieved in the objective function following operator application. Operators demonstrating superior performance are assigned higher weights. Subsequently, operators for the subsequent s iterations are selected probabilistically using a roulette-wheel strategy. Throughout the iterations, the algorithm continuously updates and refines candidate solutions, tracks the best-found solution, evaluates solution quality, and adaptively adjusts the operator-selection strategy. Collectively, these operators, guided by the adaptive mechanism, collaborate to progressively enhance solution quality. Ultimately, the algorithm outputs the best-performing solution from the solution set X.

Algorithm 1: Customized adaptive large neighborhood search algorithm.

Input: K, R, N, A, X_{current}; **Output:** X_{best} // $X_{\text{current}}/X_{\text{best}}$: current/best solution
$[K, R, N, A] \leftarrow Preprocess(K, R, N, A)$; $R_{\text{pool}} \leftarrow R$
$X_{\text{initial}} \leftarrow \emptyset$
while $R_{pool} \neq \emptyset$ **do**
    Insert requests in R_{pool} into X_{initial} with Greedy Insertion
    if $R_{pool} \neq \emptyset$ **then**
        Remove some requests from X_{initial} to R_{pool} with Random Removal
Set T_{Temp} based on $F'(X_{\text{initial}})$; $max\_iteration \leftarrow$ Preset max iteration number
$X_{\text{last}} \leftarrow X_{\text{initial}}$; $X_{\text{best}} \leftarrow X_{\text{last}}$; $iteration \leftarrow 0$;
repeat
    Refresh operator weights; Select destruction and repair operators by roulette-wheel strategy
    $X_{\text{current}} \leftarrow X_{\text{last}}$
    X_{current}, $R_{\text{pool}} \leftarrow$ Destroy X_{current} with selected operator; update R_{pool}; $flag \leftarrow$ false
    Insert requests from R_{pool} into X_{current} using selected repair operator
    X_{current}, $R_{\text{pool}} \leftarrow$ Update with Swap operator
    Update X_{last} using simulated annealing (Ropke and Pisinger, 2006)
    if $F'(X_{last}) < F'(X_{best})$ **then**
        $X_{\text{best}} \leftarrow X_{\text{last}}$
    $T_{\text{Temp}} \leftarrow T_{\text{Temp}} \times c$; $iteration \leftarrow iteration + 1$; // c is the cooling rate.
until $iteration \geq \max\_iteration$;
return X_{best}

5.2. Operator design

Operator design is pivotal for efficient resource allocation, high transport performance, and cost control. In the proposed ALNS framework, operators jointly reconfigure transport plans, enabling continual exploration of improved strategies tailored to the current demand profile. The customized insertion, removal, and swap operators are presented in Sections 5.2.1, 5.2.2, and 5.2.3, respectively.

5.2.1. Insertion operators

We adopt four insertion operators to construct and diversify solutions, including *greedy insertion*, *transshipment insertion*, *random insertion*, and *regret insertion*. The greedy insertion operator exploits the solution space by selecting the minimum-cost insertion position for each request. The transshipment insertion operator explicitly considers intermediate transfer opportunities to improve

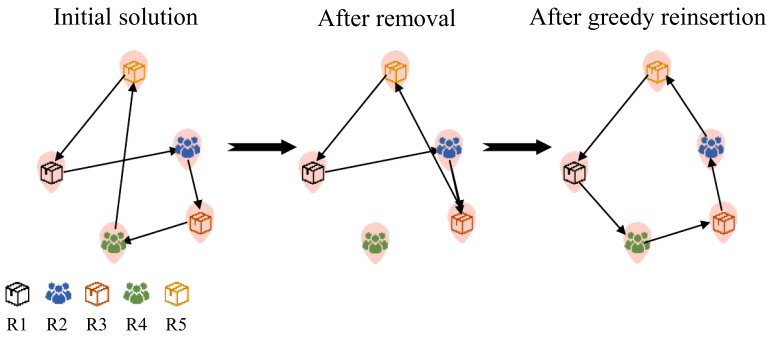


Fig. 2. Illustration of the greedy insertion operator mechanism.

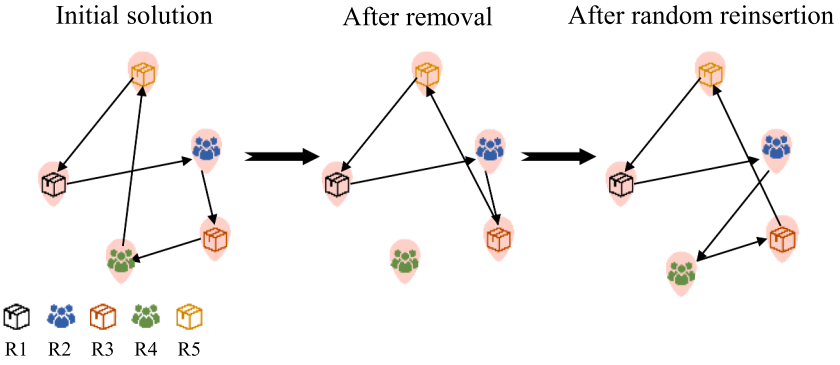


Fig. 3. The mechanism of random insertion operator.

connectivity in multimodal routes. The random insertion operator introduces diversity by assigning requests to randomly selected feasible positions. The regret insertion operator balances exploration and exploitation by inserting requests based on the maximum difference between the best and second-best insertion costs.

The *greedy insertion operator* evaluates all feasible insertion positions for each request r across the set of vehicles $k \in K$, as illustrated in Fig. 2. Let Positions(k) denote all possible insertion points in vehicle k's route. The operator selects the position (k^*, j^*) that minimizes the cost increment $\Delta F'(r, k, j)$, computed as:

$$(k^*, j^*) = \underset{k \in K, j \in \text{Positions}(k)}{\operatorname{argmin}} \Delta F'(r, k, j). \tag{41}$$

The *transshipment insertion operator* enables requests to be transferred at intermediate transshipment centers. This operator promotes inter-vehicle cooperation to reduce costs, making it particularly suitable for complex or large-scale requests. For instance, large deliveries can first be transported by a high-capacity vehicle to a transshipment point and subsequently distributed to smaller vehicles for final delivery.

As shown in Fig. 3, the *random insertion operator* randomly selects a vehicle $k \in K$ and a potential insertion position within its route. It then checks whether inserting request r at the selected position satisfies all constraints, including time-window feasibility. If the insertion is valid, the request is added to the route. This stochastic approach enhances solution space exploration, helping avoid premature convergence to local optima and fostering the discovery of diverse routing configurations.

The *regret insertion operator* prioritizes requests based on their calculated regret values. For each request r, the regret value c_r is defined as the gap between its k-th lowest insertion cost $\Delta F_r^{(k)}$ and the minimal insertion cost $\Delta F_r^{\min}$, i.e.,

$$c_r = \Delta F_r^{(k)} - \Delta F_r^{\min} \tag{42}$$

Typically, the request with the highest regret value c_r is inserted first, as it would incur the greatest additional cost if not placed in its most favorable position.

To improve computational efficiency, multiple requests are considered for insertion simultaneously. The set of requests is split into two categories: non-conflicting requests (group a), which can be inserted to separate vehicles, and conflicting requests (group b), which compete for insertion into the same vehicle. For group b, requests are ranked according to their regret values. If the top-ranked request with the highest regret value, denoted $r_{\text{regret}k}$, is associated with only one feasible vehicle, it is directly inserted. If $r_{\text{regret}k}$ is linked to multiple potential vehicles, the algorithm examines whether these vehicles are also targeted by other requests. Based on a comparative analysis of regret values, the algorithm determines whether to insert $r_{\text{regret}k}$ or, if necessary, select the next-highest regret request $r_{\text{second}k}$ for insertion. The remaining conflicting requests in group b are excluded from insertion during this iteration.

5.2.2. Removal operators

We adopt five removal operators with distinct characteristics: *worst removal*, which eliminates requests with the highest insertion cost to intensify optimization; *random removal*, which introduces exploration by randomly selecting requests for removal; *related removal*, which removes spatially or temporally related requests to encourage localized restructuring; *historical removal*, which prioritizes requests frequently removed in past iterations; and *route removal*, which clears entire routes to enable large-scale solution diversification. The fundamental operation common to all removal operators is the removal of a request's operations from the existing route, followed by recalculating the timing of the affected routes. However, they differ in the criteria used to select which requests to remove.

The *worst removal operator* selects the request r^* with the highest cost contribution using:

$$r^* = \underset{r \in R}{\operatorname{argmax}}\, C(r), \tag{43}$$

where $C(r)$ denotes the contribution of request r to the total cost F':

$$C(r) = \sum_{k \in K} \sum_{(i,j) \in A} y_{ij}^{kr} \left(c_k^1 \tau_{ij}^k + c_k^{1'} d_{ij}^k \right) q_r. \tag{44}$$

Removing the most costly request is expected to reduce the total cost in subsequent repair phases, particularly for long-distance or resource-intensive requests.

The *random removal operator* introduces diversification by randomly selecting a subset of n vehicles $K' \subseteq K$ and removing one request from each. For each vehicle $k \in K'$, let

$$R_k = \left\{ r \in R \mid y_{ij}^{kr} = 1 \right\} \tag{45}$$

be the set of requests served by vehicle k. Then, a request r_k is randomly selected from R_k. This stochastic approach enhances exploration and mitigates the risk of premature convergence to local optima.

The *related removal operator* initiates by randomly choosing request r and proceeds to eliminate a set of requests that are related to r in terms of spatial and temporal characteristics. Let $\mathcal{N}(r)$ represent the set of related requests, expressed as:

$$\mathcal{N}(r) = \left\{ r' \in R \;\middle|\; \frac{|t_{p(r)} - t_{p(r')}| + d_{p(r)p(r')}}{2} \leq \theta \right\}, \tag{46}$$

where $t_{p(r)}$ and $t_{p(r')}$ denote the pickup times, $d_{p(r)p(r')}$ is the spatial distance between pickup locations of requests r and r', and θ is a user-defined proximity threshold. By considering contextual similarity, this operator enables more strategic route restructuring and potentially yields better optimization results.

The *historical removal operator* leverages accumulated knowledge during the search process. Let H_r represent the historical lowest cost. This operator identifies the request that maximizes the ratio between the current cost and the historical lowest cost:

$$r^* = \underset{r \in R}{\operatorname{argmax}} \left(\frac{F_r}{H_r} \right). \tag{47}$$

By doing so, it guides the search toward reinserting requests into more cost-effective positions, thus enhancing solution quality.

The *route removal operator* targets entire vehicle routes with low utilization or high cost. Specifically, when the utilization rate of a vehicle k, given by

$$\frac{\sum_{r \in R_k} q_r}{u_k}, \tag{48}$$

is below a threshold (where R_k is the set of requests assigned to k), or when the route's cost is disproportionately high, the entire route is removed. All associated requests are returned to the request pool for reassignment. This operator supports global restructuring of the transport plan, aiming to reduce the total number of vehicles used and improve overall load efficiency.

5.2.3. Swap operator

The primary goal of the *swap operator* is to enhance vehicle utilization and reduce overall transport costs. By redistributing requests among vehicles, it seeks to optimize the use of available vehicle capacity, minimize idle time, and reduce unnecessary empty travel.

Let $C(k)$ denote the set of conflicting requests associated with vehicle k, expressed as:

$$C(k) = \left\{ r \in R \;\middle|\; t_{d(r)}^{kr} > b_{d(r)} \text{ or } \sum_{r \in R_k} q_r > u_k \right\}, \tag{49}$$

where $t_{d(r)}^{kr}$ is the delivery time of request r in vehicle k's current route, $b_{d(r)}$ is the upper bound of the delivery time window for request r, R_k is the set of requests currently assigned to vehicle k, and u_k is the capacity of vehicle k. A request $r \in C(k)$ is said to violate either time window or capacity constraints.

Once such a conflicting request is identified, the swap operator is invoked. It first applies the *historical removal operator* to eliminate inefficient or problematic requests that may cause constraint violations. Subsequently, it utilizes the *greedy insertion operator* to reallocate these requests to alternative vehicles in a cost-efficient and constraint-compliant manner.

Table 3
Fleet composition settings in the experiments.

Fleet compositions	Modes	Fleet composition	Modes
S1	eVTOL-drone-GV	S4	eVTOL-GV
S2	GV only	S5	GV-drone
S3	eVTOL only	S6	eVTOL-drone

For each request $r \in C(k)$, the operator identifies the best alternative vehicle $l^* \in K \setminus \{k\}$ that yields the lowest additional cost when inserting r, formulated as:

$$l^* = \underset{l \in K \setminus \{k\}}{\operatorname{argmin}} \Delta F'(r, l). \tag{50}$$

This ensures that each conflicting request is reassigned to a more suitable vehicle in a way that minimizes the total cost and satisfies all operational constraints.

By incorporating these operators, the ALNS algorithm balances exploitation of high-performing regions in the solution space with exploration of alternative configurations, thereby achieving efficient solutions for multimodal transport planning.

6. Numerical experiments

This section introduces the numerical setting (Section 6.1), benchmarking comparison with the Gurobi solver and representative heuristics (Section 6.2), the small-instance case studies (6.4), the large–instance scalability experiments (Section 6.5, and sensitivity analysis (Section 6.6). We summarize the managerial insights in Observations 1-8. The calculations of load factor and unit cost are provided in A.

6.1. Experimental design

This section outlines the numerical settings in Section 6.1.1 and the synthetic demand-generation process used in our experiments in Section 6.1.2.

6.1.1. Numerical setting

To evaluate the efficacy and practicality of the proposed IAG-MTPP for joint passenger–parcel multimodal transport planning, together with the customized ALNS algorithm, we design a series of computational experiments that approximate realistic operating conditions. The experiments are structured to reflect the key sources of complexity in multimodal transport systems-heterogeneous vehicle types, operational constraints, and spatio-temporal demand patterns-and to capture how these elements interact. The vehicle parameters are available online[1].

To account for the stochastic nature of the metaheuristic algorithms, each experimental instance was executed independently three times. The results reported in tables represent the average values over these runs, and a fixed set of random seeds was used to ensure reproducibility.

The experimental section includes six representative fleet-composition scenarios, summarized in Table 3. Each scenario corresponds to a distinct configuration of available transport modes, ranging from conventional ground vehicles to emerging air-mobility options such as drones and eVTOL aircraft.

To improve reproducibility, we summarize the key parameter settings used in our ALNS implementation in Table 4 (Ropke and Pisinger, 2006). Operator weights are initialized uniformly and updated every s iterations using a reaction-factor rule:

$$\omega_i \leftarrow (1-\rho)\omega_i + \rho\left(\pi_i / n_i\right) \tag{51}$$

where π_i is the accumulated score of operator i within the last segment and n_i is its usage count. The score calculation scheme is implemented in accordance with Ropke and Pisinger (2006).

All experiments are executed on a computer with a 5.4 GHz processor and 32.0 GB of RAM. For experiments requiring exact optimization, Gurobi 11.0.1 was employed to solve the MILP formulations. The overall algorithm was implemented in Python 3.12.

6.1.2. Synthetic demand generation

We adopt the procedure of Miyoshi et al. (2025) to generate synthetic passenger-travel and parcel-delivery requests.

Temporal dimension. Because no publicly available dataset exists for hybrid passenger-parcel services, we construct stylized time-intensity profiles that reproduce the typical peak and off-peak patterns observed in on-demand mobility systems. These synthetic profiles enable controlled and reproducible experiments and are commonly used when empirical calibration data are unavailable. Let

[1] https://github.com/cjgogo27/vehicle-parameter

Table 4
Key parameter settings of the customized ALNS.

Category	Parameter	Symbol	Value
Simulated annealing	Initial temperature	T_0	$F'(X_{\text{initial}})/10$
	Cooling rate	c	0.9
	Acceptance scaling factor	δ	5
Termination	Maximum iterations	$N_{\max}$	200
	Time limit	$t_{\max}$	48 h
Adaptive weights	Update interval	s	20
	Reaction factor	ρ	0.5

$F'(X_{\text{initial}})$ denotes the objective value of the initial solution.

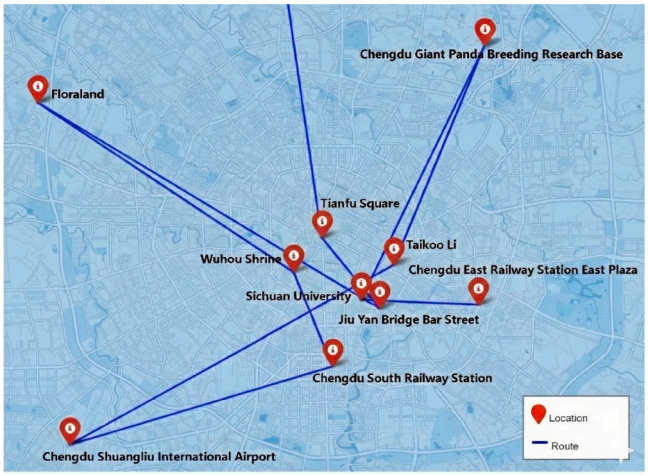


Fig. 4. Stations in Chengdu city.

$X(T)$ denote the number of requests arriving in the interval $T = [t, t + \Delta t)$. Under a non-homogeneous Poisson process, $X(T)$ follows a Poisson distribution (Lewis and Shedler, 1976) with mean $\Lambda(T) = \int_t^{t+\Delta t} \lambda(\tau)\,\mathrm{d}\tau$:

$$\Pr\big[X(T) = k\big] = \frac{\Lambda(T)^k \mathrm{e}^{-\Lambda(T)}}{k!}, \qquad k = 0, 1, 2, \ldots, \tag{52}$$

where $\lambda(\tau)$ is the time-varying arrival intensity. Passenger intensity $\lambda^{\mathrm{P}}(\tau)$ exhibits commuter peaks (08:00–10:00 and 17:00–19:00), whereas parcel intensity $\lambda^{\mathrm{C}}(\tau)$ reflects business-hour shipping waves (09:00–12:00 and 14:00–17:00). Outside these windows, lower baseline arrival rates apply.

Request attributes. Each generated request $r \in R$ is represented by the attribute tuple

$$r = \big\langle p(r), d(r), [a_{p(r)}, b_{p(r)}], [a_{d(r)}, b_{d(r)}], q_r, w_r \big\rangle,$$

where $p(r)$ and $d(r)$ denote the pickup and delivery stations; $[a_{p(r)}, b_{p(r)}]$ and $[a_{d(r)}, b_{d(r)}]$ are the admissible pickup and delivery time windows; and (q_r, w_r) is the size–weight pair that consumes vehicle capacity.

Urban joint-mobility scenario (Chengdu). The urban test bed comprises ten high-demand hubs in Chengdu (Fig. 4). Chengdu East Railway Station, Chengdu Shuangliu International Airport, and Chengdu South Railway Station are sampled most frequently as origins because of their superior connectivity and traffic volumes. All ten locations can act as transshipment nodes, reflecting the flexibility of real urban operations. To reflect urban congestion in the numerical setup, we differentiate travel conditions through the mode-specific arc feasibility A_w and the corresponding travel distances d_{ij}^k for vehicles operating on these arcs. In addition, the peak-hour arrival intensities $\lambda^{\mathrm{P}}(\tau)$ and $\lambda^{\mathrm{C}}(\tau)$ emulate demand surges during commuting and business peaks, thereby amplifying delay-related effects in urban operations.

Emergency-rescue scenario (Jiuzhaigou). To stress-test the framework under rugged terrain and time-critical conditions, we construct a second scenario centered on Jiuzhaigou County, an area prone to earthquakes and landslides. Requests originate from the two Chengdu gateways above and terminate at either Jiuzhai-Huanglong Railway Station or Jiuzhaigou Town (Fig. 5). This layout mirrors rapid-response corridors and highlights the advantages of aerial modes when ground access is disrupted. In the mountainous setting, terrain constraints are captured by restricting ground accessibility through the sparser ground arc set A_g, and by adopting longer

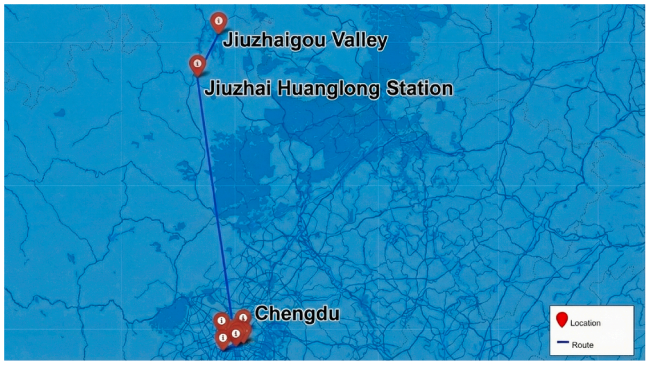


Fig. 5. Locations in the emergency transport scenario.

Table 5
Comparison of Gurobi and customized ALNS.

		Avg. cost (CNY)			Avg. CPU (s)	
$\|T\|$	$\|R\|$	Gurobi	ALNS	Gap (%)	Gurobi	ALNS
1	1	193	193	0.00	23.11	**0.02**
1	3	668	668	0.00	624.40	**0.08**
1	5	863	872	1.04	3414.43	**0.19**
1	10	1223	1280	4.66	8117.11	**1.04**
1	20	-	2789	-	10800.00*	**7.52**
3	1	193	193	0.00	689.98	**0.07**
3	3	667	668	0.15	10800.00*	**0.07**
3	5	863	872	1.04	10800.00*	**0.12**
3	10	-	1280	-	10800.00*	**1.23**
3	20	-	2576	-	10800.00*	**8.34**
5	1	193	193	0.00	1050.69	**0.03**
5	3	667	668	0.15	10800.00*	**0.08**
5	5	863	872	1.04	10800.00*	**0.13**
5	10	-	1135	-	10800.00*	**1.31**
5	20	-	2576	-	10800.00*	**8.41**

$|T|$: number of transshipment stations;
$|R|$: number of requests. $*$ time limit reached (3 hours), no optimal solution found. - no feasible solution found due to time limitation. Gap = (ALNS cost - Gurobi cost) / Gurobi cost $\times$ 100%, calculated to two decimal places.

effective ground travel distances d_{ij}^k to represent winding roads and limited road conditions. By contrast, aerial vehicles can use the arc sets A_e or A_u to follow more direct flight paths and bypass ground obstacles. When applicable, pre-determined safe corridors are represented via fixed-route arcs A_{fix}^k for vehicles in K_{fix}.

The resulting synthetic demand sets exhibit realistic temporal clustering, spatial heterogeneity, and size/weight distributions, providing a consistent basis for evaluating the proposed multimodal planning framework in both routine and emergency contexts.

6.2. Benchmarking against exact and heuristic approaches

This section benchmarks the proposed approach against both a commercial exact solver in Section 6.2.1 and several representative heuristics in Section 6.2.2.

6.2.1. Comparison with the commercial solver

Table 5 compares the commercial MILP solver GUROBI with the proposed ALNS algorithm across progressively larger instances. For the smallest cases ($|T| = 1,\ |R| \in \{1, 3\}$), both approaches return the same optimal objective values, 193 CNY and 668 CNY, respectively, confirming that the algorithm is capable of attaining global optima when the search space is limited.

As instance complexity grows, the performance of GUROBI degrades sharply. With $|T| = 1,\ |R| = 20$ and $|T| = 5,\ |R| = 10$, the solver reaches the time limitation (10800 s) without producing a feasible solution, whereas ALNS is able to efficiently find feasible schedules and does so within seconds of CPU time. For these instances where GUROBI does not return a feasible solution within the time limit, we cannot assess solution quality via an optimality gap; therefore, the results primarily demonstrate the practicality of ALNS in producing feasible solutions under tight computational budgets. The same pattern is observed for all larger instances (see the "Gap (%)" and "Avg. CPU (s)" columns in Table 5).

Table 6
Comparing the proposed ALNS with the representative heuristic algorithms.

		Total Cost (CNY)					Number of Served Requests				
$\|R\|$	$\|K\|$	2-OPT	Greedy	SA	GA	ALNS	2-OPT	Greedy	SA	GA	ALNS
5	5	14355.8	14355.8	14355.8	14288.3	**14288.3**	5	5	5	5	5
10	5	24456.8	24456.8	40692.0	24456.8	**20819.3**	8	8	8	8	8
10	10	41298.0	41298.0	41448.0	37570.5	**37533.0**	10	10	10	10	10
25	10	73981.5	74864.3	71980.5	74864.3	**68857.5**	19	19	18	19	19
20	20	88068.8	96347.3	87737.3	84297.0	**83431.5**	20	20	20	20	20
30	20	90017.3	91002.8	92714.3	91002.8	**83028.8**	30	30	30	30	30
40	20	106343.3	141206.3	113484.8	107549.3	**92629.5**	35	35	34	35	35
30	30	90017.3	91002.8	92714.3	91002.8	**82926.0**	30	30	30	30	30
40	30	124671.0	121954.5	134918.3	124659.8	**109179.0**	40	40	40	40	40
50	30	190425.8	187108.5	198735.8	175033.5	**171423.0**	50	50	50	49	50
60	30	177026.3	188271.8	188271.8	186789.8	**175785.0**	54	54	54	54	54
40	40	124671.0	121954.5	134918.3	122305.5	**109044.8**	40	40	40	40	40
50	40	190425.8	187108.5	198735.8	191767.5	**171511.5**	50	50	50	50	50
60	40	199639.5	192133.5	210894.8	200799.0	**178900.5**	60	60	60	60	60
70	40	236747.3	239283.0	259542.0	255177.8	**233290.5**	65	65	65	65	65
90	40	258059.3	276324.0	303948.0	272368.5	**252861.8**	75	75	75	75	75

$|R|$: number of requests; $|K|$: number of vehicles; 2-opt: Two-Optimal Exchange Algorithm; Greedy: Greedy Algorithm; SA: Simulated Annealing Algorithm; Genetic: Genetic Algorithm; ALNS: Adaptive Large Neighborhood Search.

Regarding computational efficiency, ALNS exhibits significant advantages. Across all tested instances, the average CPU time required by ALNS remains under ten seconds. For example, in the simplest case of $|T| = 1, |R| = 1$, the solution is obtained in just 0.02 seconds. In contrast, Gurobi's runtime increases sharply with problem size. It takes 624.40 seconds for the case of $|T| = 1, |R| = 3$, and reaches the imposed time limit of 10,800 seconds in the cases of $|T| = 3, |R| = 3$ and $|T| = 5, |R| = 3$. In emergency-rescue scenarios, such delays could compromise the timely delivery of critical supplies. Even in routine urban logistics, excessive runtime may lead to delayed dispatch decisions, reduced responsiveness, and ultimately lower customer satisfaction and operational efficiency. These results confirm that ALNS is both computationally necessary and practically effective. When GUROBI provides feasible benchmarks, ALNS attains comparable costs with small gaps (Table 5); for larger instances, its primary advantage lies in ensuring feasibility and delivering solutions within acceptable time limits.

6.2.2. Comparison with representative heuristics

We further benchmark ALNS against several representative heuristics, including 2-OPT (Englert et al., 2014), a greedy constructive heuristic (Zhao et al., 2025), Simulated Annealing (SA), and a Genetic Algorithm (GA) (Zhu et al., 2025). These algorithms are tested on a series of instances with different request volumes $|R|$ and fleet sizes $|K|$, and are evaluated in terms of total cost, number of completed requests, and computation time. The detailed numerical results are reported in Table 6.

Across all scenarios, ALNS consistently delivers the best solution quality. It achieves the lowest or jointly lowest total cost in every row of Table 6, while always matching the highest request-completion level attained by any of the benchmark algorithms. In contrast, the alternative heuristics frequently generate substantially higher costs and, in some instances, fail to serve all available requests, for example, SA at $(O, K) = (25, 10)$ and GA at $(O, K) = (50, 30)$. These gaps widen as the problem size increases.

In terms of computation time, ALNS is naturally slower than simple constructive procedures such as 2-OPT and the greedy heuristic, but it remains efficient: runtimes range from well below one second for the smallest instances to about 53 s for the largest case ($O = 90,\ K = 40$), scaling smoothly with problem size. GA and SA exhibit larger variability in runtime across instances and do not deliver comparable gains in solution quality or service completeness. Overall, ALNS offers the most robust balance among cost, service level, and computational effort, making it a suitable algorithmic choice for the IAG-MTPP.

6.3. Ablation study on ALNS operators

Table 7 reports an ablation study in which individual operators are removed from the ALNS and the resulting performance is compared with the full algorithm. The complete operator set ("Ours") is the only configuration that simultaneously serves all 20 requests ($F = 20$) and achieves the lowest total cost ($F' = 12{,}593$). When key operators such as Greedy Insertion, Worst Removal, Route Removal, Related Removal, or Random Insertion are disabled, the number of served requests drops sharply to six, rendering these variants impractical for real operations. Only the variants without Historical Removal or Random Removal still serve all 20 requests, but both incur higher total costs (12,918 and 12,904 CNY, respectively), as well as higher average cost per served request. The apparently lower cost of 7,052 CNY observed in several ablated variants is therefore misleading, since it is obtained by serving only six requests and does not reflect a realistic service level. The proposed approach also requires more computation time (about 368 s on average), because it explores a much larger feasible region to accommodate all 20 requests and refine their routes. Overall, the ablation results indicate that each operator contributes meaningfully either to service coverage or to cost minimization, and that removing any of them degrades the effectiveness of the ALNS.

Table 7
Comparison of performance between the approach without individual operators and the proposed approach.

Without (W/O) Operator	F	F'	F_1	F_2	F_3	F_4	Avg. CPU(s)
Ours	20	12593	9973	2555	80	66	367.94
W/O Greedy Insertion	6	7052	6070	953	0	29	11.30
W/O Historical Removal	20	12918	10319	2527	80	72	510.92
W/O Worst Removal	6	7052	6070	953	0	29	11.64
W/O Route Removal	6	7052	6070	953	0	29	11.18
W/O Related Removal	6	7052	6070	953	0	29	11.27
W/O Random Removal	20	12904	10322	2515	0	68	331.36
W/O Random Insertion	6	7052	6070	953	0	29	11.10

F: Number of served requests; F': Total cost; F_1: Transport cost; F_2: Transshipment cost; F_3: Storage cost; F_4: Carbon tax

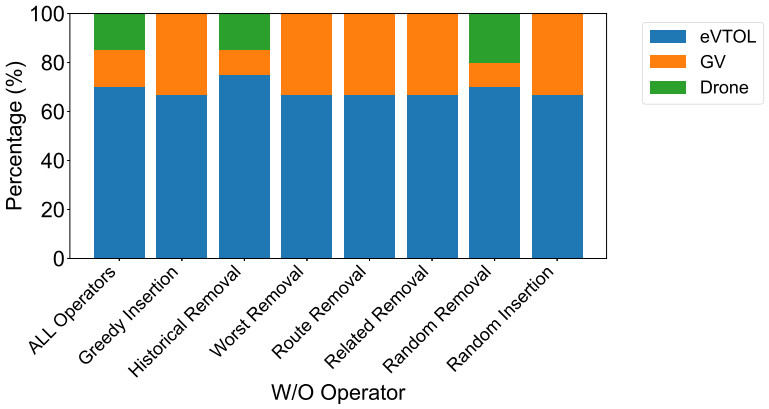


Fig. 6. Comparison of modal shares between the approach without individual operators and the proposed approach.

Fig. 6 compares the modal shares of the operator-ablated variants with those of the proposed approach. The variants that serve only six requests exhibit identical modal shares. Among the approaches that still serve all 20 requests but at a higher cost, the share of ground vehicles (GVs) is systematically lower, whereas the shares of drones and eVTOLs are higher than in the proposed approach. This pattern indicates that the proposed approach is better at exploiting low-cost GV options, while the ablated variants rely more heavily on comparatively expensive aerial modes.

6.4. Small-instance case studies

Before scaling up to larger instances, we first apply the integrated IAG–MTPP model and the customized ALNS algorithm to four computationally tractable, context-rich case studies: (i) a time-critical mountain-rescue mission in Jiuzhaigou (Section 6.4.1); (ii) an urban joint passenger–parcel service in the Chengdu metropolitan area (Section 6.4.2); (iii) an examination of priority strategies in hybrid passenger–parcel operations (Section 6.4.3); (iv) an assessment of system behavior under different weighting schemes in the weighted-sum objective (Section 6.4.4); and (v) analyses of the impact of capital cost (Section 6.4.5). These scenarios differ markedly in urgency, service priorities, request heterogeneity, and logistical complexity, thereby providing a stringent test of the framework's ability to coordinate heterogeneous fleets under contrasting operating conditions. The remainder of this subsection presents detailed results for each case, including cost structures, service reliability, modal utilization, and sensitivity to fleet composition and demand characteristics.

6.4.1. Case 1: Mountain–rescue case study in Jiuzhaigou

Fig. 7 summarizes the operational characteristics of the three modes used for emergency rescue in mountainous areas. eVTOLs provide high speed and long range but are associated with higher operating costs and emissions. Drones offer strong route flexibility and low emissions, though their payload capacity is limited. GVs are cost-effective and capable of transporting large volumes, yet they are slower and constrained by terrain and road conditions. When these complementary modes are jointly deployed, an efficient collaborative transport system can be established. Transshipment stations act as critical nodes within this system, enabling seamless handovers between transport modes. This multimodal framework has the potential to reduce response times for medical-supply delivery, increase the request-fulfillment rate, and create a vital time window for life-saving interventions.

Jiuzhaigou's steep valleys and frequent landslides impose severe accessibility constraints, limiting both rapid response and spatial coverage. Demand in this case consists of urgent medical supplies (e.g. medicines, blood packs) and rescue personnel, each with

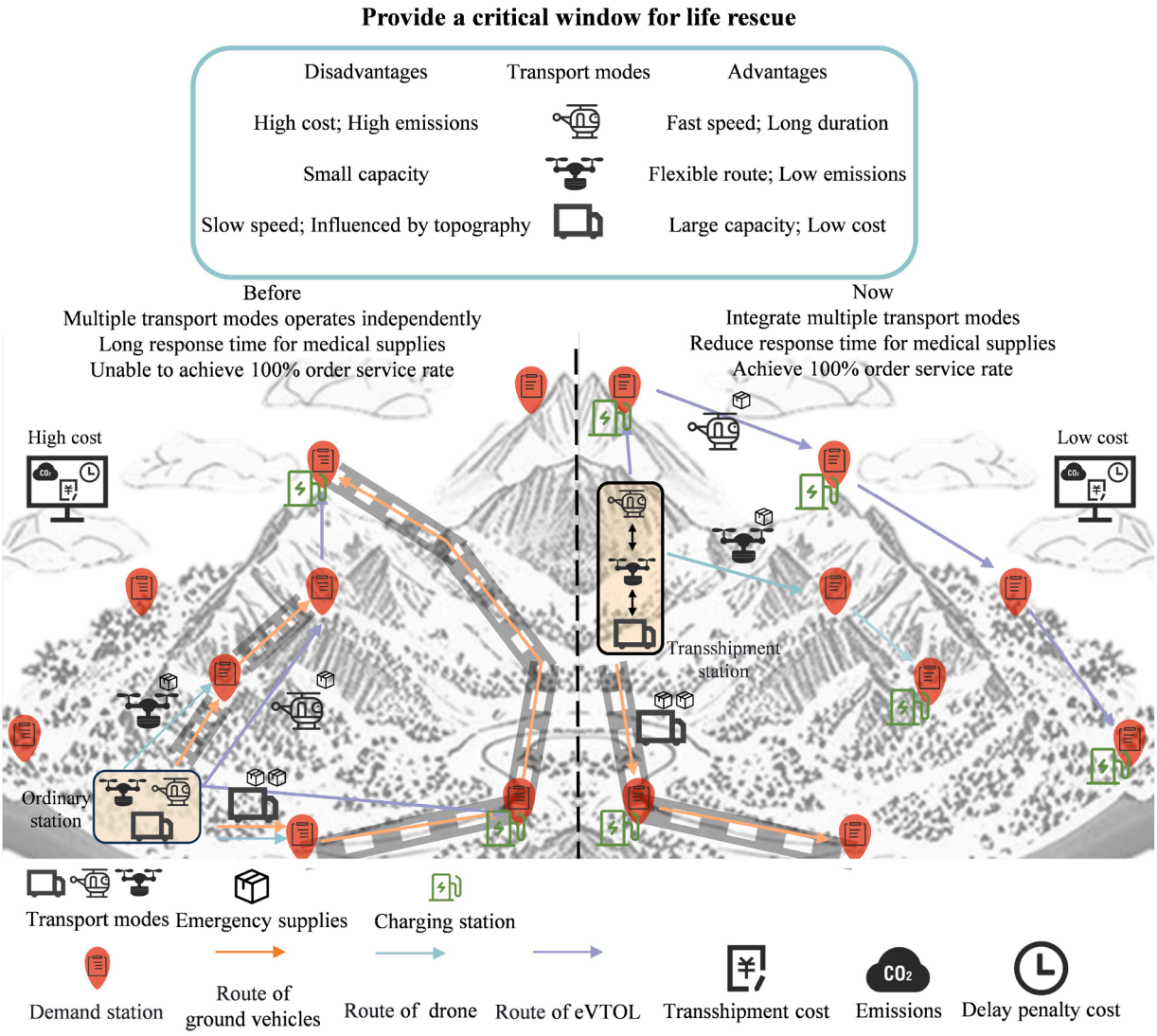


Fig. 7. Characteristics of transport modes used for emergency rescue in mountainous areas.

distinct time windows, urgency factors, and priority weights. Origins are restricted to Chengdu East Railway Station and Chengdu Shuangliu International Airport, whereas Jiuzhai–Huanglong Railway Station and Jiuzhaigou Town act as destinations and transshipment hubs (see Fig. 5).

We evaluate the six fleet compositions listed in Table 3. For each configuration, multiple demand instances are generated with a mix of passenger and parcel requests, and the IAG–MTPP is solved using the customized ALNS algorithm.

Table 8 reports the aggregate performance metrics. The fully integrated eVTOL–drone–GV system attains a *100% fulfillment rate* (10/10 demand points) with the *lowest unit cost* of 0.0098 CNY, outperforming all single- and dual-mode alternatives. GVs handle bulk cargo (40% of total volume), eVTOLs cover medium-to-long aerial legs (40%), and drones deliver lightweight, high-priority items to dispersed sites (20%). This coordinated division of labor, enabled by the Jiuzhai–Huanglong transfer hub, reduces the average delay penalty to 0.3295, an 18.7% improvement relative to the GV-only benchmark. By contrast, relying solely on eVTOLs satisfies only six requests, and the Drone + GV combination raises fulfillment to 37.5% of total demand but still falls short of complete coverage. Crucially, the integrated configuration compresses delivery times for critical medical supplies to under three hours, a target unattainable with any single- or dual-mode alternative.

Observation 1. *In the Jiuzhaigou emergency scenario, the integrated eVTOL–drone–GV fleet achieves full demand coverage (10/10 requests served) at the lowest unit transport cost (0.0098 CNY), outperforming all single- and dual-mode baselines. By assigning long aerial legs to eVTOLs, bulk movements to GVs, and dispersed high-priority deliveries to drones, the system exploits strong speed–capacity complementarity and reduces average delay penalties by about 18.7% relative to a GV-only benchmark. These results suggest that, in rugged and disaster-prone*

Table 8
Performance comparison of multi-modal transport systems in rescue scenarios.

Scenario	Unit cost (CNY)	Load Factor (%)	Unit delay penalty (CNY)	Served Requests	Used vehicles	Served portion (%) eVTOL	GV	Drone
S1	0.0140	57.32	0.4051	7	7	0	100	0
S2	0.0126	56.88	0.0000	6	5	100	0	0
S3	0.0089	60.26	0.3372	9	10	60	40	0
S4	0.0159	71.74	0.3941	8	8	0	62.5	37.5
S5	0.0187	68.94	0.0000	6	6	50	0	50
S6	0.0098	68.84	0.3295	10	10	40	40	20

Table 9
Performance comparison of multi-modal transport systems in joint passenger-parcel scenarios.

Passenger-parcel Ratio	Unit cost (CNY)	Load Factor (%)	Served Requests	Emission Cost (CNY)	Used Vehicles	Served Portion (%) eVTOL	GV	Drone
All passengers	0.063	0.69	10.00	55.32	10	50	50	0
9:1	0.061	0.65	10.00	51.01	10	40	50	10
8:2	0.067	0.63	10.00	47.31	10	50	40	10
7:3	0.068	0.63	10.00	38.52	10	50	30	20
6:4	0.069	0.58	10.00	36.46	10	50	30	20
5:5	0.069	0.56	10.00	34.10	10	50	30	20
4:6	0.069	0.49	10.00	30.67	10	50	30	20
3:7	0.068	0.50	10.00	26.64	10	50	20	30
2:8	0.069	0.49	10.00	26.14	10	50	20	30
1:9	0.066	0.49	10.00	17.85	10	40	10	50
All parcels	0.068	0.45	10.00	14.06	10	30	10	60

regions, investing in a heterogeneous fleet and well-located transshipment hubs can simultaneously improve response times, service coverage, and cost efficiency for time-critical rescue logistics.

6.4.2. Case 2: Joint passenger–parcel operations in Chengdu

Rapid urbanisation in Chengdu has intensified congestion, increased last–mile logistics costs, and expanded the city's carbon footprint. To examine whether the proposed multimodal framework can alleviate these pressures, we construct ten demand sets whose passenger–to–parcel ratios range from *all–passenger* (10:0) to *all–parcel* (0:10). Table 9 summarizes the resulting system performance.

Across all demand compositions, the integrated eVTOL–drone–GV fleet achieves a 100% request fulfilment rate (10/10 requests) while keeping the unit cost tightly bounded between 0.061 CNY and 0.069 CNY. As the parcel share rises from 0 to 100%, two clear trends emerge. First, drone utilisation increases from 0 to 60% of served volume, leveraging drones' low–emission profile and short–haul efficiency. The associated decline in emission cost, from 55.32 CNY to 14.06 CNY, represents a 74.6% reduction relative to the all–passenger baseline. Second, the average vehicle load factor naturally falls (0.69 → 0.45) because parcels are lighter and smaller than passengers; yet operating cost remains stable, indicating that flexible routing and mode choice effectively offset the apparent loss in volumetric utilisation. The system's responsiveness is particularly evident at the 1:9 passenger–to–parcel ratio: drones absorb 50% of demand, eVTOLs 40%, and GVs only 10%, while the unit cost increases by just 0.005 CNY compared with the 9:1 scenario, demonstrating strong cost–containment capability under large shifts in demand mix.

During commuting peaks (all–passenger case), eVTOLs shoulder up to 50% of trips, with the remainder carried by GVs. In contrast, drones dominate e–commerce deliveries when parcels account for most requests (up to 60% of served volume in the all–parcel case). At a balanced 5:5 passenger–parcel split, a modal mix with 50% eVTOL, 30% GV, and 20% drone usage achieves a favourable cost–emission trade–off, providing practitioners with quantitative guidance for policy–aligned fleet management and low–carbon urban operations.

Observation 2. *In the Chengdu case, the integrated eVTOL–drone–GV fleet maintains a 100% request fulfilment rate across all passenger–parcel ratios while keeping unit cost within a narrow band of 0.061–0.069 CNY (Table 9). As the parcel share increases and drones take over up to 60% of served volume, emission costs fall by about 74.6% (from 55.32 to 14.06 CNY) despite a reduction in average load factor from 0.69 to 0.45, showing that flexible modal reallocation can simultaneously contain operating costs and substantially improve environmental performance. These results suggest that urban operators can accommodate large shifts between passenger and parcel demand primarily by adjusting modal shares within a fixed eVTOL–drone–GV fleet, without resizing the fleet or sacrificing service quality.*

6.4.3. Case 3: Priority strategies in hybrid passenger–parcel transport

In hybrid passenger–parcel transport, priority rules play a crucial role in determining operational performance. Depending on service policies, passengers may be given precedence over parcels, or both may be treated with equal priority. To assess the impact of these priority strategies, we conduct experiments under the two settings. Fig. 9 shows the routes under these two strategies, and Fig. 8 compares system performance under Passenger High Priority and Equal Priority strategies in hybrid passenger–parcel transport.

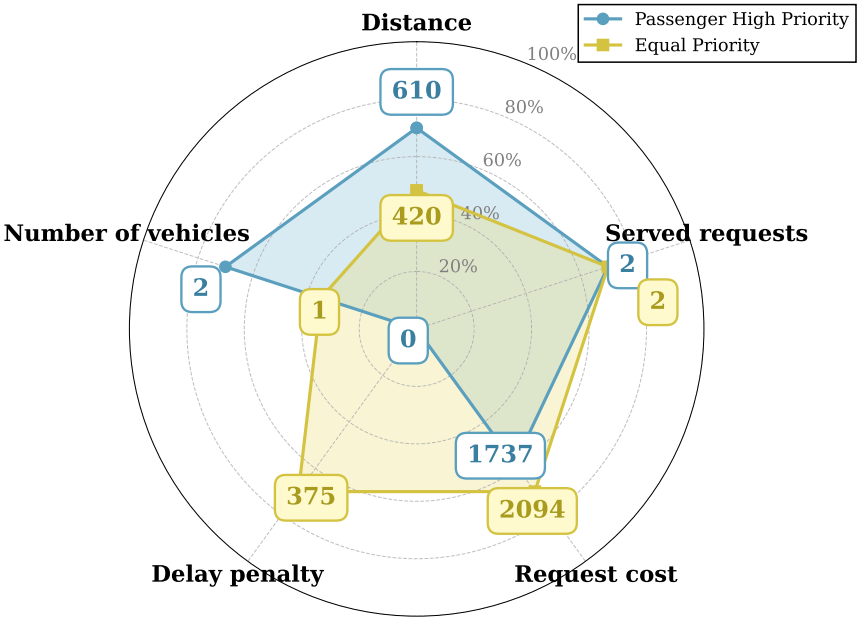


Fig. 8. Analysis of priority strategies in hybrid passenger–parcel transport.

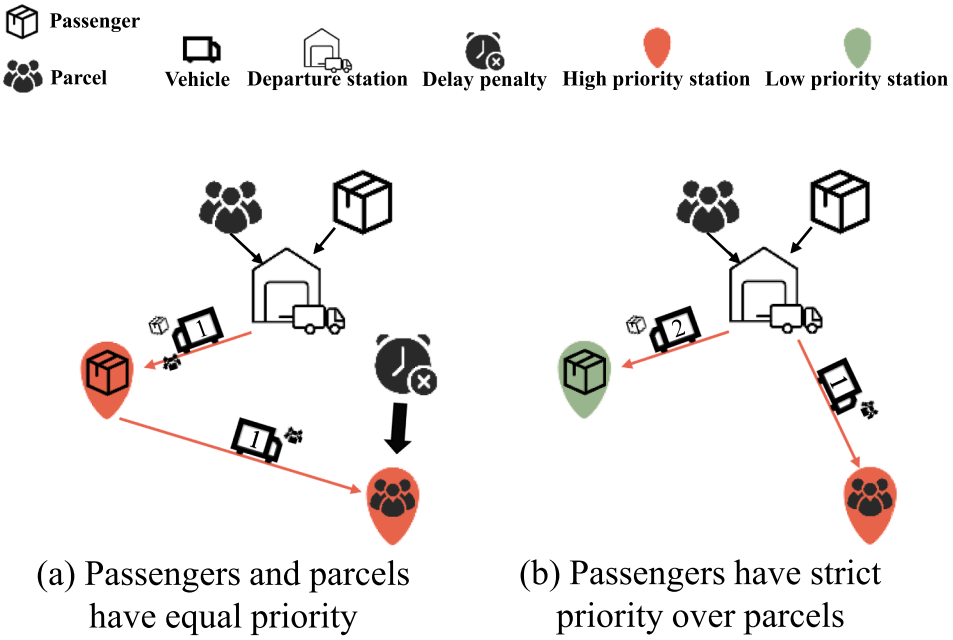


Fig. 9. Routes of vehicles under different priority strategies in hybrid passenger–parcel transport.

Both strategies serve the same number of requests, but their operational outcomes differ substantially. When passengers are given priority, the system deploys more vehicles (2 vs. 1) and travels a longer distance (610 vs. 420) to eliminate delay, resulting in a zero delay penalty. This strategy also achieves a lower total request cost (1737 vs. 2094), indicating that avoiding passenger lateness yields overall efficiency gains despite increased fleet usage. In contrast, the Equal Priority strategy consolidates service onto fewer vehicles, but this leads to a high delay penalty (375), reflecting the reduced service quality experienced by passengers when parcel and passenger requests are treated uniformly. These results reveal a trade-off: prioritizing passengers improves timeliness and reduces cost, but requires more intensive fleet deployment, whereas equal treatment lowers resource use at the expense of higher delays and reduced overall performance.

Table 10
Experimental results under different weight configurations.

Scale	Distance	Total service costs (CNY)	Total carbon tax (CNY)	Total delay penalty (CNY)	Total costs (CNY)	Avg. CPU (s)
1:10:10	2343	47902	360	90	48352	64.19
1:3:1	2335	47902	362	90	48354	61.17
1:3:3	2590	47982	361	90	48433	75.64
1:10:1	3248	45893	296	112	46301	348.27
3:1:1	3698	43287	333	1349	44969	42.86

Scale: Weight ratio of three objectives in the transport system (format: $w_1 : w_2 : w_3$, where w_1 represents weight for transport cost, w_2 represents weight for carbon tax, and w_3 represents weight for delay penalty); F_1: Transport cost; F_2: Transshipment cost; F_3: Storage cost; F_4: Total carbon tax; F_5: Total delay penalty; $F_1 + F_2 + F_3$: Total service cost.

Table 11
Capital costs by service type and vehicle category.

Vehicle Type	Capital cost per day (CNY)		
	Multifunction (Passenger + Parcel)	Passenger-Only	Parcel-Only
eVTOL	1200	800	400
GV	500	300	150
Drone	800	600	300

Note: In practical calculations, the capital cost is converted into an hourly cost, which is then multiplied by the actual transport time to determine the total capital cost.

Observation 3. *In hybrid passenger–parcel operations, giving passengers higher priority leads the system to use more vehicles and travel longer distances, but it eliminates delay penalties and yields a lower total request cost. By contrast, an equal-priority policy reduces fleet usage but incurs substantial delay penalties, degrading passenger service quality and increasing overall cost. This suggests that, when passenger punctuality is valued, prioritizing passenger requests is economically justified even if it requires more intensive fleet deployment.*

6.4.4. Case 4: Results under varying objective weights

The results in Table 10 show that different weight configurations lead to clear shifts in the distribution of transport, carbon, delay, and transshipment costs. When higher weights are assigned to delay penalties and the carbon tax, the model prioritizes time reliability and emission reduction, thereby significantly suppressing delay and carbon-related costs. A notable example appears when the carbon-tax weight is set to 10: carbon tax drops sharply, but the system incurs higher transshipment cost and longer computation time due to more complex routing.

Conversely, when transport-cost weight dominates, transport cost decreases, but the resulting routing choices increase transshipment activities and lead to noticeably higher delay penalties. These results highlight that adjusting weights directly shapes the operational trade-offs. Managers should therefore tune the weights according to real priorities, for example, assigning a higher delay-penalty weight when service timeliness is essential, or increasing the carbon-tax weight when environmental objectives are emphasized.

Observation 4. *The modest savings in pure transport cost (e.g. configuration 3:1:1) can trigger an order-of-magnitude increase in delay penalties and storage cost, whereas placing more emphasis on delay and carbon (e.g. 1:10:10 or 1:3:3) keeps F_4 and F_5 low at the price of slightly longer routes and computation times. This indicates that weight selection should be driven by service and policy priorities: when timeliness and emissions matter, assigning higher weights to delay and carbon is preferable to minimizing transport cost.Table 10*

6.4.5. Case 5: Capital cost analysis

We evaluate whether multifunction vehicles remain advantageous once explicit vehicle capital costs are accounted for. Table 11 shows the capital costs by service type and vehicle category (Lobo Leasing, 2021; DiDi, 2025; Dajiang, 2025). The eVTOL capital cost estimates are informed by expert consultation. Fig. 10 compares a single-function fleet, where passenger and parcel tasks are operated separately, with a multifunction fleet that jointly serves both request types. Under both demand levels ($|R| = 10$ and $|R| = 20$), the single-function configuration requires between 9 and 13 vehicles, whereas the multifunction configuration needs only four to five vehicles. Although multifunction vehicles have higher per-vehicle capital cost than their passenger-only or parcel-only counterparts (Table 11), the fleet-wide capital cost contribution is lower: when $|R| = 10$, the total capital cost decreases from 238 to 200 CNY, and when $|R| = 20$, from 296 to 250 CNY. These savings are driven by substantially higher load factors (from 0.591 to 0.978) and more requests handled per vehicle (from 1.11 to 2.50), which in turn reduce unnecessary routing, shorten effective travel distances, and lower associated service costs (e.g., from 3 398 to 3 394 CNY at $|R| = 10$ and from 3 842 to 3 760 CNY at $|R| = 20$).

The above results rely on the baseline daily capital-cost estimates in Table 11. To assess sensitivity to changes in capital costs, we conduct a sensitivity analysis by varying only the daily capital costs of eVTOLs. Table 12 reports the results for $|R| = 10$ under several alternative eVTOL cost configurations. When eVTOL daily capital costs increase moderately (configurations M1–M3 in Table 12), the

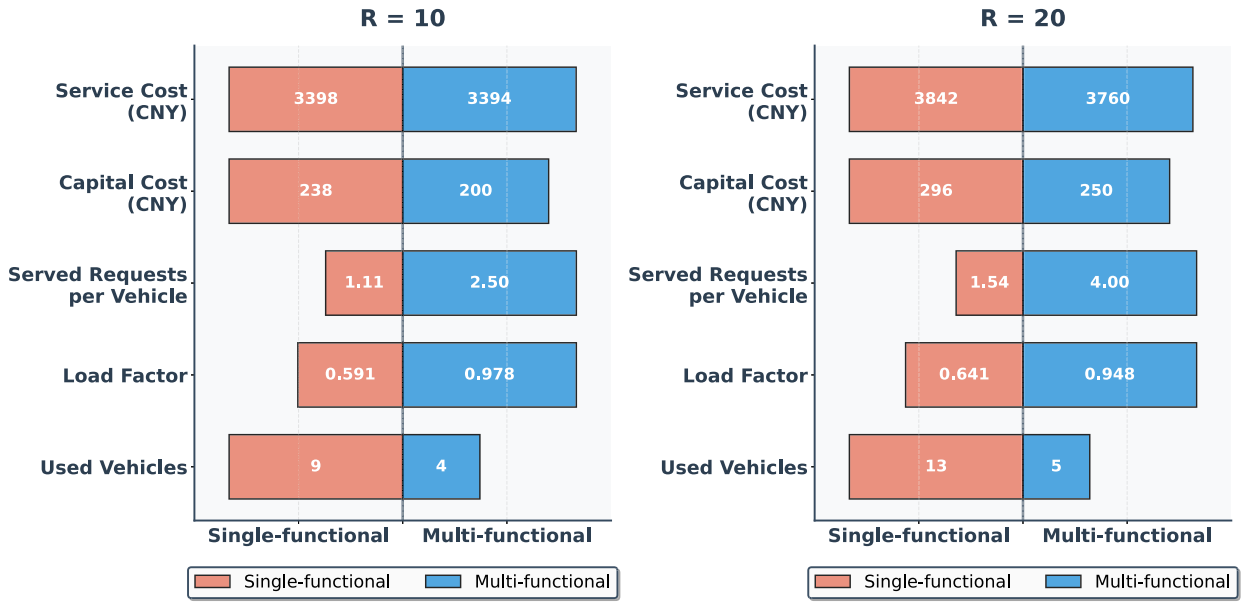


Fig. 10. Indicator trade-off between single and mixed passenger-parcel transport.

Table 12
Sensitivity analysis results under different eVTOL daily capital cost configurations ($|R| = 10$).

Mode	Multi/Pas/Par	Fleet type	Service cost (CNY)	Capital cost (CNY)	Used vehicles	Served req. per vehicle	Load factor
M1	1200/800/400	Specialized	3398	238	9	1.11	0.591
		Multifunction	3394	200	4	2.50	0.978
M2	1400/1000/600	Specialized	3398	270	9	1.11	0.591
		Multifunction	3394	220	4	2.50	0.978
M3	1500/900/500	Specialized	3398	255	9	1.11	0.591
		Multifunction	3394	230	4	2.50	0.978
M4	2000/1500/800	Specialized	3415	300	10	1.00	0.570
		Multifunction	3420	265	5	2.00	0.880
M5	5000/1000/600	Specialized	3398	270	9	1.11	0.591
		Multifunction	3505	340	7	1.43	0.700

Multi, Pas, Par: Daily capital cost configuration (CNY/day) for multifunction, passenger-only, and parcel-only eVTOLs, respectively. Specialized fleet: fleet composed of passenger-only and parcel-only eVTOLs. Multifunction fleet: fleet composed of multifunction eVTOLs capable of serving both passenger and parcel requests.

optimal fleet structure remains unchanged and the multifunction fleet continues to achieve higher utilization with fewer vehicles, thereby maintaining a lower or comparable system cost. However, when the daily capital cost of the multifunction eVTOL increases sharply relative to passenger-only and parcel-only eVTOLs (configuration M5 in Table 12), the solution structure adjusts: the multifunction system deploys more vehicles and its utilization declines, which increases both the capital-cost contribution and the service cost. In this case, the multifunction configuration can lose its cost advantage. This extreme setting corresponds to scenarios in which multifunction eVTOLs face substantially higher acquisition or maintenance burdens than single-purpose eVTOLs.

Overall, the results highlight a clear trade-off between per-vehicle capital cost and fleet-wide efficiency. When demand density and temporal alignment allow for high load consolidation and route sharing, and multifunction-vehicle capital costs are within a reasonable range, reductions in fleet size and improvements in utilization can offset higher unit capital costs, yielding better system-level performance than separate passenger-only and parcel-only fleets. From an operational standpoint, fleet planning should not be based solely on capital cost; decisions should incorporate utilization efficiency and the potential for pooling heterogeneous requests. In scenarios with low demand density, significant spatiotemporal mismatch, or very expensive multifunction eVTOLs, dedicated fleets may become economically preferable. Conversely, in contexts where passenger and parcel time windows align or where weight profiles are complementary, multifunction vehicles can deliver both economic benefits and more effective resource deployment.

Fig. 11. Vehicle load factor characteristics.

Observation 5. *Pooling passenger and parcel requests can make multifunction vehicles economically attractive when the resulting utilization improvements are substantial and the capital cost of multifunction eVTOLs does not become excessively higher than that of passenger-only or parcel-only eVTOLs. Managers should therefore evaluate fleet options jointly in terms of pooling potential, utilization, and capital cost, rather than relying on unit capital cost alone.*

6.5. Large–instance scalability experiments

The following subsections present the results from large-instance scalability experiments, including analyses of fleet composition performance in Section 6.5.1 and load factor behavior under different request distributions in Section 6.5.2.

6.5.1. Results analysis of different fleet compositions

To evaluate scalability, we expand the demand set to $|R| = 200$ and solve all six fleet compositions (S1–S6; see Table 3). The main performance indicators are summarized in Table 13. For the largest instances, the fully integrated fleet (S1: eVTOL + Drone + GV) clearly stands out: at $|R| = 200$ it serves 135 requests (67.5% of all demand) at a unit cost of 0.152 CNY. In comparison, the GV-only baseline (S2) serves only 58 requests at a much higher unit cost of 1.415 CNY, and the GV + Drone mix (S5) serves 95 requests at 0.879 CNY. The eVTOL-only (S3), eVTOL + GV (S4), and eVTOL + Drone (S6) fleets achieve somewhat lower unit costs (between 0.136 CNY and 0.253 CNY), but at the expense of significantly lower service rates (66–112 requests). Overall, S1 offers the most attractive combination of high service coverage and low unit cost among all configurations.

This advantage is due to the mode-specific task allocation. In S1, eVTOLs absorb medium- and long-haul commuter trips, drones handle short-range, high-frequency parcel drops, and GVs move bulky or heavy freight. The ALNS algorithm continually rebalances these roles, assigning each request to the mode that best matches its distance, payload, and time-window characteristics. This coordinated use of all three modes improves fleet utilization and keeps delay and carbon components of the objective (F_4, F_5) under control, even as the problem size grows.

From a computational standpoint, the customized ALNS remains tractable at large scales. For small and medium instances ($|R| \leq 50$), all scenarios are solved within a few tens of seconds of CPU time. When the request volume increases to $|R| = 100$, runtimes rise into the order of several minutes (e.g., about 1,209 s for S1), reflecting the enlarged search space. At $|R| = 200$, the S1 instance requires roughly 4,966 s on average, while the other scenarios fall in the 600–4,700 s range (Table 13). Although the growth in runtime is nonlinear, it remains acceptable for tactical planning problems of this scale and is still far below the effort required by exact MILP solvers, which quickly hit time or memory limits for significantly smaller instances (see Section 6.2.1).

For low-altitude-economy operators, S1 is also attractive from an infrastructure perspective. By co-locating vertiports and logistics hubs, the integrated system can share facilities between passenger and freight flows, reducing the number of dedicated sites and lowering estimated infrastructure costs by about 34% in our numerical setting, while improving operational agility by roughly 22%. As the request volume increases from 10 to 200, the unit costs of all scenarios rise, but S1 exhibits the slowest rate of increase. This

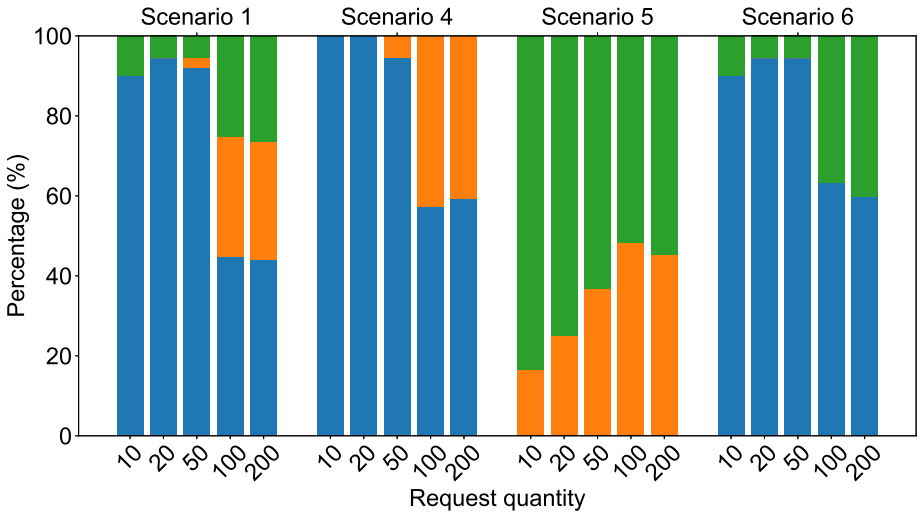


Fig. 12. Modal shares under varying numbers of requests (green: drone, blue: eVTOL, orange: GV).

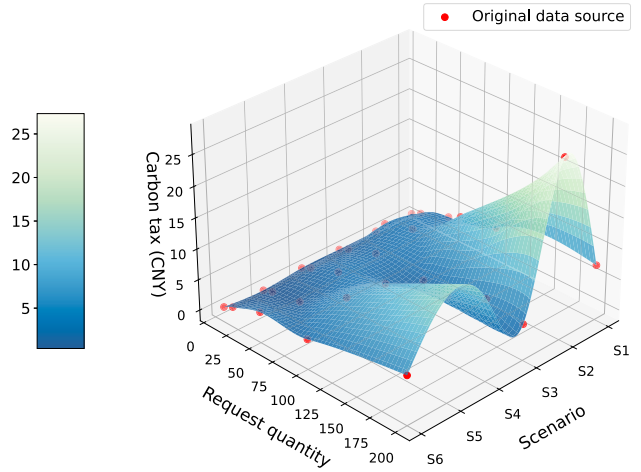


Fig. 13. Carbon emission analysis across operational modes.

efficiency gain is primarily due to (i) dual-use facilities that spread fixed investments over a larger number of services (Choi and Park, 2022), and (ii) a high-frequency transshipment strategy that reduces empty mileage and improves vehicle utilization.

As shown in Fig. 11, in small-scale scenarios ($|R| = 10$) S1 achieves a load factor of 48.93%, with eVTOLs handling 90% of services and drones covering the remaining 10%. This reflects a centralized deployment pattern that leverages eVTOLs' flexibility for medium- to long-haul operations. As demand scales up to $|R| = 200$, the system gradually shifts toward a more decentralized structure: transshipment becomes more frequent, the average load factor declines slightly, and a larger share of work is shifted to drones and GVs. This transition enhances responsiveness and adaptability in the face of spatially dispersed requests.

The modal shares in Scenario S1 are depicted in Fig. 12. At low request volumes, eVTOLs dominate the network due to their versatility and speed. As the scale increases, drones and GVs progressively take on a greater share of tasks, especially for short- and mid-range movements. In medium- and large-scale settings, this shift intensifies: more transshipment operations and vehicle deployments are required, but the resulting substitution of long-haul eVTOL segments with shorter drone and GV legs yields substantial cost savings while maintaining high service levels.

From an environmental perspective, S1 also outperforms the alternatives. Fig. 13 shows that at $|R| = 200$ the integrated system incurs a carbon-tax cost of only 7.19 CNY/ton, compared with 27.31 CNY/ton for the GV-only scenario S2. This reduction is achieved through a combination of measures: replacing short-haul GV routes with drones, exploiting advanced route optimization, and assigning about 26.47% of logistics tasks to drones in parcel-dominant situations. These adjustments enable S1 to decouple growth in demand from growth in emissions more effectively than any partial-mode fleet.

Observation 6. *Across all demand scales up to $|R| = 200$, the fully integrated fleet (Scenario S1) provides the best overall balance between service coverage, unit cost, and emissions: for the largest instance it serves 135 requests at a unit cost of 0.152 CNY while keeping carbon tax*

Table 13
Comparison of results in different fleet composition.

Fleet Composition	R	Unit cost (CNY)	Served requests	F' (CNY)	F_1 (CNY)	F_2 (CNY)	F_3 (CNY)	F_4 (CNY)	Initial cost	Number of vehicles	Iteration	Avg. CPU (s)
S1	10	0.052	10	3934	2336	1090	489	20	3934	10	0	0.40
S2	10	0.287	2	2170	1508	270	385	7	2170	2	0	0.40
S3	10	0.049	10	3970	2339	1101	511	20	3970	10	0	0.60
S4	10	0.049	10	3970	2339	1101	511	20	3970	10	0	0.60
S5	10	0.410	6	2175	1609	268	290	8	2175	6	0	0.70
S6	10	0.052	10	3934	2336	1090	489	20	3934	10	0	0.60
S1	20	0.054	18	8665	5257	2050	1314	44	8699	18	8	26.30
S2	20	0.238	6	6161	4636	768	735	22	6161	6	0	1.30
S3	20	0.052	18	8701	5260	2061	1336	44	8735	18	3	13.20
S4	20	0.052	18	8701	5260	2061	1336	44	8735	18	5	20.50
S5	20	0.336	12	6186	4763	760	641	22	6186	12	0	2.00
S6	20	0.054	18	8665	5257	2050	1314	44	8699	18	5	22.30
S1	50	0.064	37	21940	13949	4709	3170	112	21973	37	1	27.30
S2	50	0.321	10	13140	9521	1911	1663	44	13140	11	0	5.00
S3	50	0.060	35	20196	12537	4362	3192	105	20229	35	3	46.30
S4	50	0.062	37	22135	14041	4770	3212	112	22169	37	1	48.10
S5	50	0.325	18	13199	9720	1885	1549	46	13199	19	0	7.80
S6	50	0.061	36	20295	12610	4376	3204	105	20295	36	0	18.40
S1	100	0.089	86	63670	45822	10556	7035	256	62640	79	17	1209.10
S2	100	0.409	42	47721	35921	6654	5010	136	47721	33	0	51.00
S3	100	0.066	44	22236	13456	4680	3997	102	22196	35	5	278.60
S4	100	0.082	75	63819	45366	10731	7475	247	63855	66	5	495.80
S5	100	0.287	60	46783	36566	6587	3470	160	48923	50	12	824.00
S6	100	0.069	60	23275	14966	4744	3450	115	23539	55	5	417.80
S1	200	0.152	135	79907	57730	13076	8839	262	77946	96	16	4966.30
S2	200	1.415	58	54144	41907	7476	4626	134	55733	33	3	601.10
S3	200	0.136	66	26802	16172	5649	4890	90	26500	37	18	2560.30
S4	200	0.253	112	81738	57842	13496	10151	249	76926	69	5	1541.20
S5	200	0.879	95	59978	46061	8073	5675	169	57191	65	15	4247.20
S6	200	0.189	97	29531	19317	5899	4197	118	29390	67	17	4735.60

F': Total cost; F_1: Transport cost; F_2: Transshipment cost; F_3: Storage cost; F_4: Carbon tax

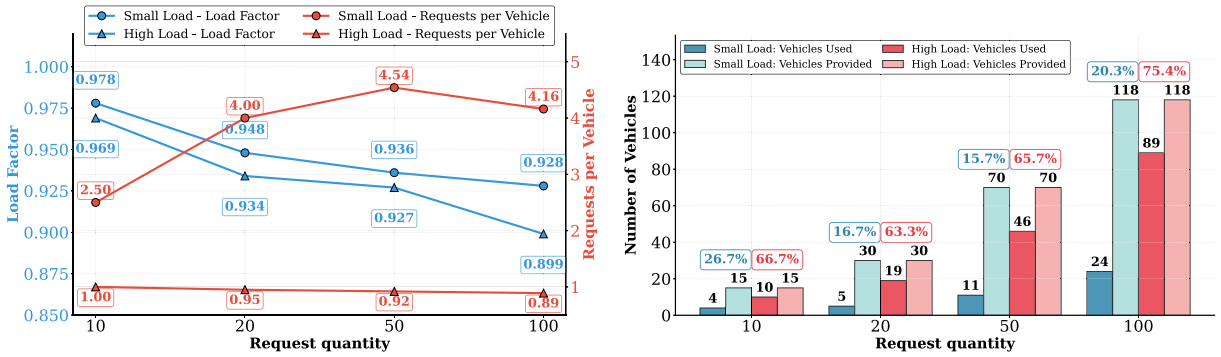
far below single- and dual-mode benchmarks (Table 13, Fig. 13). As demand grows, S1 leverages shared facilities, frequent transshipment, and adaptive reallocation between eVTOLs, drones, and GVs (Figs. 11–12) to slow the growth of both unit cost and carbon tax, indicating that heterogeneous eVTOL–drone–GV integration is particularly well suited for scaling low-altitude logistics with controlled costs and emissions.

6.5.2. Load factor analysis under different request-distribution scenarios

The load factor is highly sensitive to the spatial distribution, temporal structure, and weight profile of requests. The large–scale experiments above rely on randomly generated demand, which can naturally produce instances with relatively low average load factors even when vehicles are used efficiently. To examine this issue more systematically, we design two stylized request–distribution scenarios that isolate the main drivers of vehicle utilization: a small–load scenario, dominated by time–window overlap, and a high–load scenario, dominated by payload constraints.

The small–load scenario represents situations where passenger and freight requests share similar or overlapping time windows and jointly respect vehicle capacity limits. As shown in Figs. 14(a) and 14(b), this temporal alignment enables the algorithm to consolidate multiple heterogeneous requests onto a small number of vehicles, resulting in highly efficient fleet utilization. Even with only four vehicles, the system can serve ten requests, and as demand rises to 25, 50, and 100 requests, each vehicle consistently handles more than four requests on average. This scenario highlights a key mechanism: temporal compatibility across requests greatly expands pooling opportunities and avoids unnecessary dispatches. From an operational perspective, this suggests that operators can enhance consolidation efficiency by encouraging time–window alignment, for example, through coordinated scheduling, incentive schemes, or slightly flexible submission windows.

The high–load scenario captures cases in which either passenger weight or freight weight pushes vehicle payload toward its limit. In these settings, high load factors can still be achieved, but the number of requests that each vehicle can serve drops markedly. As shown in Fig. 14a, the average number of served requests per vehicle is substantially lower than in the small–load scenario. This reflects a structural constraint: payload–heavy requests reduce combinability, forcing the system to deploy more vehicles to satisfy the same total demand. Fig. 14b further shows that occupancy remains high, but this is driven by payload saturation rather than by efficient aggregation of multiple small requests. Managerially, this underscores the importance of monitoring weight distributions and adapting operations accordingly, for instance, by deploying higher–capacity vehicles during freight–heavy periods, separating weight–intensive jobs, or designing station networks that shorten overly long or heavy trips.



(a) Load factor and served requests per vehicle (b) Vehicle occupancy

Fig. 14. Comparison of load factor and served requests in high-load and small-load scenarios.

Observation 7. *When passenger and freight requests have overlapping time windows (small-load scenario), temporal compatibility enables strong consolidation: the system pools heterogeneous requests on shared tours, maintaining high load factors and serving more than four requests per vehicle on average as demand grows. By contrast, in high-load scenario, the binding constraint shifts from timing to capacity, which reduces request combinability and lowers the average number of requests served per vehicle, even though load factors can remain high due to payload saturation.*

6.6. Sensitivity analysis

To better understand the behavior of the proposed IAG-MTPP under different operating conditions, we conduct a series of sensitivity analyses. These experiments examine how changes in key parameters, i.e., fleet flexibility, vehicle speed and unit–kilometer cost, time–window width, and battery capacity, affect transport cost, service efficiency, and vehicle utilization.

6.6.1. Impact of vehicle flexibility

Fleet flexibility is examined by distinguishing six vehicle categories: flexible eVTOLs, fixed eVTOLs, flexible GVs, fixed GVs, flexible drones, and fixed drones. By varying the proportion of fixed vehicles in the fleet, we evaluate the impact on (i) total and unit transport cost, (ii) computation time, (iii) number of active vehicles, (iv) utilization rates, and (v) modal shares. These indicators help determine how much routing flexibility is worth paying for in practice.

As summarized in Table 14, full demand fulfillment is attainable whenever either the entire fleet is flexible or only the eVTOL subgroup retains routing freedom. As the proportion of fixed vehicles increases, the number of served requests declines monotonically. Although higher flexibility sometimes appears to be associated with higher unit costs, a comparison at the same service level (e.g., $|R| = 100$) shows that more flexible fleets achieve lower total cost, reflecting more efficient assignment of requests to vehicles. A pronounced side effect of reduced flexibility is the sharp rise in storage costs: fixed vehicles are tied to rigid departure and arrival times, so early-arriving passengers and parcels generate waiting time at stations, which is accounted for as storage cost and idle time.

Vehicle usage and modal shares also vary with both the flexibility level and demand scale. At lower demand ($|R| = 50$), fleet capacity is ample, and most requests are assigned to eVTOLs. In this regime, using drones more intensively does not reduce cost; instead, their tendency to operate in formations and to require more frequent loading, unloading, and transshipment increases F_2, and the algorithm therefore keeps drone usage minimal. Under higher demand ($|R| = 100$), increasing flexibility enables more requests to be shifted from eVTOLs to drones and GVs. When flexibility is low and demand is high, eVTOLs and GVs alone cannot satisfy all requests, and drones become essential to relieve pressure on the system, especially for parcel transport.

Algorithmic behavior mirrors this change in problem complexity. Table 14 reports average CPU time and iteration counts for $|R| \in \{50, 100\}$. As flexibility decreases, both runtime and the number of iterations required to reach the best solution fall systematically, because the solution space shrinks with fewer feasible routing combinations. As the number of requests doubles from 50 to 100, the computational burden increases sharply, reflecting the greater complexity of large–scale instances. A notable case arises at $|R| = 50$ with a fixed–vehicle proportion of 0.72, where the algorithm finds the best solution in the initial iteration, indicating that, under highly constrained flexibility and moderate demand, the solution space can be simple enough for the search to converge almost immediately.

6.6.2. Sensitivity to vehicle speed and unit–kilometer cost

Fig. 15 reports the performance of nine operating scenarios obtained from the Cartesian product of three travel speeds (80, 100, and 120 $\mathrm{km\,h^{-1}}$) and three unit–kilometer cost levels (baseline, −20%, and +20%). For each scenario, the ALNS algorithm computes near–optimal joint routes for aerial and ground segments. Five indicators are normalized to the $[0, 1]$ interval and plotted on radar axes: total cost, transit cost, unloading cost, the iteration at which the best solution first appears (Best iters), and the number of vehicles required to serve all requests (Vehicles count). Larger values correspond to lower costs, fewer vehicles, or faster convergence, enabling

Table 14
Comparison of different vehicle flexibility levels.

R	proportion	served requests	unit cost (CNY)	F' (CNY)	F1 (CNY)	F2 (CNY)	F3 (CNY)	F4 (CNY)	F5 (CNY)	vehicles count	Served Portion (%) eVTOL	GV	Drone	iteration	Avg. CPU(s)
50	0	50	0.043	44466	35468	8732	0	267	0	31	72.0	28.0	0.0	11	366.99
50	0.3	50	0.040	44468	35452	8732	4	280	0	36	72.0	28.0	0.0	14	347.41
50	0.72	15	0.148	23650	19620	3681	253	96	0	15	33.3	66.7	0.0	0	39.41
50	1	37	0.064	21940	13949	4709	3170	112	0	37	91.9	2.7	5.4	9	153.25
100	0	100	0.044	78886	62945	15485	9	442	6	54	73.0	27.0	0.0	17	2646.65
100	0.3	100	0.046	79516	63232	15485	218	434	146	53	75.0	25.0	0.0	6	1125.42
100	0.72	100	0.056	87549	69071	14935	3138	405	0	79	37.0	46.0	17.0	14	1710.56
100	1	85	0.088	62640	45300	10435	6647	257	0	79	44.7	29.4	25.9	3	373.46

F': Total cost; F1: Transport cost; F2: Transshipment cost; F3: Storage cost; F4: Carbon tax

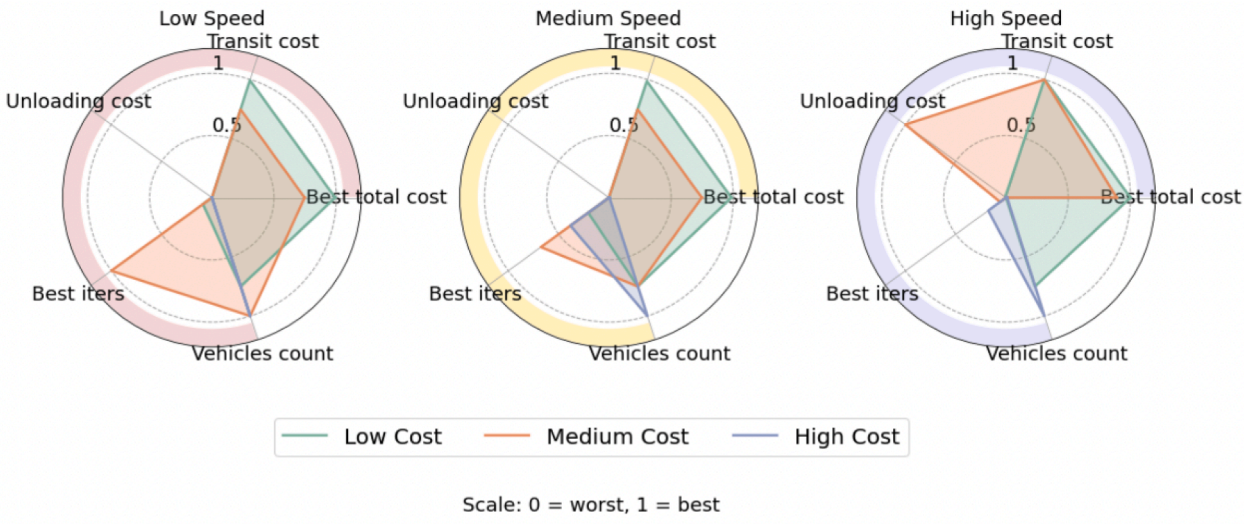


Fig. 15. Impact of aircraft speed and the unit kilometer cost.

a compact visual comparison of how speed and marginal operating cost propagate through system performance and algorithmic behavior.

Two clear patterns emerge. First, increasing vehicle speed reduces both total and transit costs: moving from 80 to 120 $\mathrm{km\,h^{-1}}$ lowers overall cost by about 18% and cuts transit expenses by roughly 20%, while leaving unloading cost essentially unchanged because handling fees are independent of speed. Second, lowering the unit–kilometer cost amplifies these gains, pushing the normalized total–cost score from 0.82 to 0.92 at medium speed and to 1.00 at high speed. These cheaper, faster configurations encourage the algorithm to schedule slightly more trips, which is reflected in a modest deterioration on the "Vehicles count" axis, but the increase in fleet size is small relative to the cost savings. Convergence behavior follows the same logic: high–speed, low–cost settings enlarge the solution space and therefore require a few additional iterations before the algorithm settles on the incumbent optimum, but the extra computation remains modest.

Practical implementation requires balancing the tradeoff between speed and cost, which reduces total transport costs by 40%. Experiment results in Chengdu demonstrate that this strategy enables drones to complete 60% of short-distance deliveries, reducing delay penalties by 18.7%. Operational constraints, including payload capacity (≥ 60 kg) and battery endurance, must be taken into account. For instance, in Jiuzhaigou, drones could complete only 37.5% of delivery tasks due to payload limitations. These findings suggest that operators should calibrate fleet composition, ensuring that the integration of drones, eVTOLs, and GVs aligns with regulatory payload limits.

6.6.3. Sensitivity to request time–window width

Fig. 16 compares two experiments in which every request's pickup–and–delivery window is either tightened by 50% or relaxed by 50% relative to the baseline schedule. When time windows are halved, scheduling flexibility shrinks and the share of feasible requests drops by almost one–tenth. To meet the tighter deadlines, the ALNS reallocates tasks toward the fastest transport modes: drone utilization increases sharply, roughly tripling in relative terms. Because many drone sorties substitute multi–leg eVTOL + GV routes, the total number of active vehicles rises only slightly. Transit and transfer costs decrease modestly, as shorter windows suppress detours and intermediate handling, but the stricter timing also lowers the average load factor and reduces vehicle productivity. The

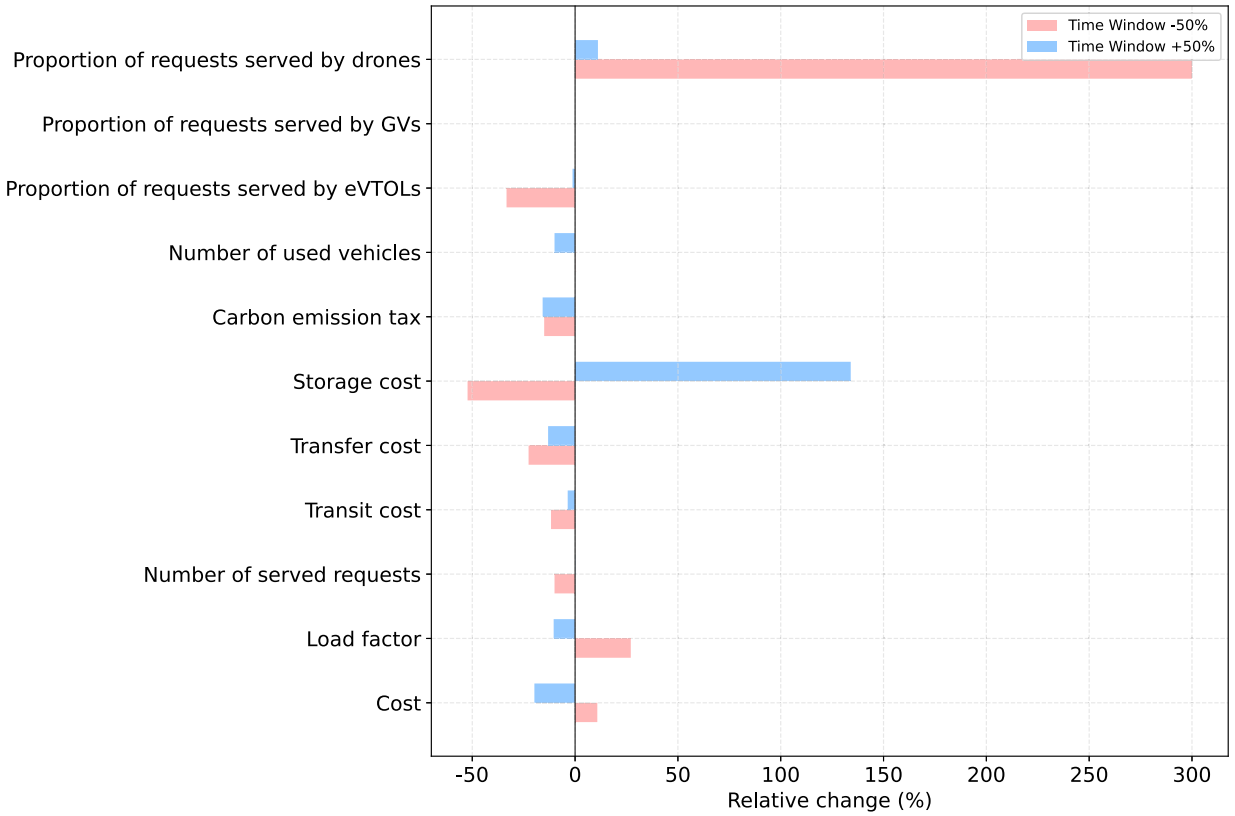


Fig. 16. Relative change in key indicators when all time windows are tightened (−50%) or relaxed (+50%).

dominant effect is a 50% reduction in storage cost, driven by shorter dwell times for passengers and parcels, which more than offsets the modest efficiency loss and leaves the overall transport bill slightly below the baseline.

Conversely, widening time windows by 50% restores nearly all previously infeasible requests and allows the algorithm to pack more trips into each tour, improving the load factor by roughly 12%. The relaxed schedule favors slower but higher–capacity eVTOLs, driving drone usage down to negligible levels. However, this flexibility comes at a substantial cost: consignments spend longer in intermediate depots, pushing storage cost up by about 140% and increasing the total system cost above the baseline despite savings in transit and transfer components. These opposing effects imply that moderate time–window widths offer the best compromise between feasibility, consolidation, and cost: extremely narrow windows erode service coverage, while very wide windows inflate storage cost and weaken environmental performance.

6.6.4. Sensitivity analysis of battery capacity

Because most aerial vehicles (eVTOLs and drones) and an increasing share of ground vehicles can be electric, battery capacity is a critical determinant of network reachability and service feasibility. Specific energy consumption parameters and battery specifications (Grepow, 2025; All About Industries, 2025; EV. guide, 2025; Zhan et al., 2025; Sripad and Viswanathan, 2021) are detailed in Table 15. In the first experiment, all vehicles start with a full battery, and charging is allowed only at four designated stations: Chengdu East Railway Station East Plaza, Chengdu Shuangliu International Airport, Chengdu South Railway Station, and Jiuzhai–Huanglong Station. We then scale battery capacity from 50% to 200% and observe how many requests can be served and how the cost structure changes.

Table 16 shows that battery capacity directly constrains the number of feasible trips. With $|R| = 10$, vehicles operating at only 50–80% capacity can serve six requests, whereas full or enlarged capacity (100–200%) allows all ten requests to be completed. At $|R| = 20$, the number of served requests rises from 11 at 50–80% capacity to 18 once capacity reaches 100% or more. Beyond this point, additional capacity no longer increases coverage in our instances, because two long and isolated trips remain unreachable under all settings; these infeasibilities are driven by spatial layout and time windows rather than energy alone. Total cost and its components (F_1–F_4) increase with demand but remain stable across the 100–200% range, highlighting that simply adding energy storage is not a universal remedy, and effective design also requires well–positioned depots, vertiports, and charging points, and assignments that respect both distance and energy constraints.

To evaluate the suitability of replacing conventional ground vehicles with electric trucks, this study conducts a battery-capacity sensitivity experiment for electric ground vehicles and the results are shown in Table 17. The results reveal a clear and consistent

Table 15
Vehicle energy characteristics and battery specifications.

Vehicle Type	B_k (kWh)	α_k (kWh/h)	β_k (kWh/km)
eVTOL	100	30	0.15
Ground Vehicle (GV)	50	10	0.20
Drone	10	18	0.30

Table 16
Performance under different battery-capacity levels and numbers of requests.

Battery level (%)	R	F	F'	F_1	F_2	F_3	F_4
50	10	6	2099	1461	626	262	12
80	10	6	2099	1461	626	262	12
100	10	10	3459	2339	1101	511	20
150	10	10	3459	2339	1101	511	20
200	10	10	3459	2339	1101	511	20
50	20	11	5193	3850	1311	829	32
80	20	11	5193	3850	1311	829	32
100	20	18	7365	5260	2061	1336	44
150	20	18	7365	5260	2061	1336	44
200	20	18	7365	5260	2061	1336	44

F: Number of served requests; F': Total cost; F_1: Transport cost; F_2: Transshipment cost; F_3: Storage cost; F_4: Carbon tax.

Table 17
Results of electric truck with varying battery capacities.

Battery level (%)	F	F'	F_1	F_2	F_3	F_4	Distance	Avg. CPU (s)
50	12	5125	3371	575	0	46	366	23.38
80	13	6341	4120	763	55	60	392	19.90
100	15	6601	4464	454	0	54	384	16.53
150	18	9312	6334	798	751	74	428	46.59
200	20	13073	10384	0	42	74	476	11.17

F: Number of served requests; F': Total cost ; F_1: Transport cost ; F_2: Transshipment cost; F_3: Storage cost; F_4: Carbon tax;

pattern: as battery capacity increases, the number of serviceable requests rises significantly. When the available battery capacity reaches 200% of the capacity, the system can fulfill all planned requests without violating energy constraints. This improvement is accompanied by a gradual increase in the total system cost, yet a crucial structural change is observed, transshipment cost drops to zero when the battery capacity becomes sufficiently large. This indicates that electric trucks can directly complete all required transport routes without additional transshipment links, thereby eliminating unnecessary transport mode-switching operations.

These results confirm that when ground transport is electrified, and battery capacity is appropriately configured, operators can simultaneously expand service coverage, reduce transhipment complexity, and improve environmental performance. Rather than maximizing battery capacity indiscriminately, however, planners should identify capacity levels that are just sufficient to eliminate critical energy bottlenecks and unnecessary transfers, and then co–design charging infrastructure and routing policies around those thresholds.

Observation 8. *The sensitivity analyses show that system performance is jointly shaped by routing flexibility, vehicle speed and unit–kilometer cost, time–window width, and battery capacity for both aerial vehicles and electric trucks (Tables 14, 16, and 17, Figs. 15 and 16). Maintaining at least one flexible mode and deploying faster, lower–cost vehicles reduces total cost and delay, while excessively tight time windows or low battery capacity sharply shrink feasibility and leave some requests unserved; conversely, overly relaxed windows and very large batteries can inflate storage or transport costs even as coverage improves.*

7. Discussions and policy implications

This section presents key operational insights derived from our results in Section 7.1 and outlines corresponding policy recommendations in Section 7.2.

7.1. Operational insights

The numerical experiments in Section 6 show clear operational benefits from integrating eVTOLs, drones, and GVs into a single fleet. Across both the Jiuzhaigou emergency rescue and Chengdu joint passenger–parcel cases (Sections 6.4.1–6.4.2), the fully integrated configuration (Scenario S1) consistently achieves full or near-full request fulfillment at the lowest unit cost, while single- and

dual-mode fleets either leave high-priority demand unserved or incur higher costs (Tables 8, 9, and 13). These results confirm that assigning medium- and long-haul trips to eVTOLs, short-haul and last-mile tasks to drones, and bulk flows to GVs allows each mode to specialize, yielding system performance that exceeds any subset of modes.

Routing flexibility emerges as a second core driver of performance. The sensitivity analysis in Section 6.6 (Table 14) shows that fleets with at least one flexible mode can maintain high service rates, whereas fully fixed-route operations generate unmet demand and high storage costs. This supports a hybrid scheduling strategy in which a fixed-route backbone on dense corridors is complemented by a flexible subfleet that absorbs off-peak, irregular, or spillover demand. From a computational standpoint, the tailored ALNS heuristic is essential to make such integrated planning tractable. In the large-instance experiments (Section 6.5), ALNS finds high-quality solutions for up to $|R| = 200$ while exact MILP formulations quickly hit time or memory limits (Section 6.2.1, Table 5), indicating that advanced heuristics are necessary for city-scale multimodal routing.

The experiments also highlight several design levers for practitioners. The speed–cost analysis in Section 6.6.2 (Fig. 15) shows that higher eVTOL speeds and lower unit-kilometer costs substantially reduce total and transit costs with only marginal changes in fleet size, underscoring the value of investing in faster and more efficient vehicles. Battery-capacity experiments (Table 16) confirm that limited energy restricts reachable trips and service coverage, and that energy upgrades must be paired with appropriate siting of depots and vertiports to fully extend network reach. Finally, the time-window experiments in Fig. 16 reveal that extremely tight windows reduce feasibility and force heavy reliance on fast modes, whereas overly wide windows increase storage time and cost; in practice, moderate time-window widths offer the best compromise between feasibility, consolidation, and overall cost.

7.2. Policy recommendations

Our findings suggest that policymakers and operators move beyond single-mode designs and adopt integrated eVTOL–drone–GV fleets. Scenario S1 serves more requests at lower unit cost than any partial fleet (Table 13, Sections 6.4.1–6.4.2), while also improving resilience when one mode is disrupted. Rather than choosing between fixed and on-demand services, operators should implement hybrid routing and scheduling: a fixed-route backbone for recurrent flows, complemented by a flexible subfleet to handle irregular and peak demand, as the flexibility analysis in Section 6.6 shows that such combinations maintain high fulfillment and curb storage costs. Service policies should also calibrate time windows and priority rules. Our sensitivity results (Fig. 16) suggest that excessively tight windows erode feasibility, whereas overly relaxed windows inflate storage costs; moderate windows therefore provide the best compromise between service coverage and cost efficiency. Priority settings, e.g., for passengers versus parcels, should ensure that high-priority tasks are expedited without destabilizing overall routing. To unlock the full benefits of multimodal integration, policymakers need to invest in infrastructure, such as well-located transfer hubs, vertiports, dedicated drone corridors, and charging or battery-swapping facilities (Sections 6.4.1 and 6.5, Tables 16–17), that support seamless air–ground transfers and mitigate energy constraints.

At the same time, targeted support for technological innovation and fleet modernization is essential. R&D programs, pilot projects, and incentives for advanced eVTOLs, drones, and electric GVs can help realize the “high-speed & low-cost” configurations identified in Section 6.6.2 and Fig. 15. Finally, regulatory and governance frameworks should be aligned with multimodal, mixed passenger–parcel operations by updating airspace, traffic, and logistics regulations to accommodate routine drone and eVTOL services, establishing certification standards for integrated passenger–parcel use, and setting carbon-related targets that explicitly recognize the emission benefits of integrated fleets (e.g., Scenario S1 in Table 13). Coordinated governance structures and city- or region-level pilots will be crucial to test, refine, and scale these solutions in practice.

8. Conclusions

This paper investigates an Integrated Air–Ground Multimodal Transport Planning Problem (IAG-MTPP) for joint passenger mobility and parcel delivery in the context of the emerging low-altitude economy. We consider a heterogeneous fleet composed of eVTOLs, drones, and GVs, operating over a multimodal network with pickup, delivery, and transshipment nodes. The problem is formulated as a Mixed-Integer Linear Programming (MILP) model with a hierarchical objective: first maximizing the number of served requests, and then minimizing a generalized cost function that aggregates transit, loading/unloading, storage, carbon-tax, and delay-penalty components. The formulation accommodates both fixed-route and flexible-route services, supports joint passenger–parcel operations, and can seamlessly specialize to parcel-only, passenger-only, or mixed-use regimes without altering the underlying model structure. Energy-consumption constraints are explicitly incorporated for electric vehicles (air and ground), capturing battery capacity, charging decisions, and energy usage along routes. To overcome the computational challenges posed by realistically sized instances, we develop a customized ALNS algorithm, which exploits problem-specific insertion (greedy, transshipment-based, random, regret-based), removal (worst, random, related, historical, route-based), and swap operators, combined with an adaptive weight-update scheme and simulated annealing–type acceptance criterion. Benchmarking against the commercial solver Gurobi shows that, on small instances, ALNS matches the optimal MILP solutions, and on larger instances, the commercial solver quickly hits the time limit while the proposed ALNS continues to return high-quality solutions within practical runtimes. Additional comparisons with representative heuristics, i.e., 2-OPT, greedy constructive heuristics, simulated annealing, and genetic algorithms, demonstrate that ALNS consistently achieves lower total costs and higher request-completion rates. An ablation study confirms that each operator class contributes meaningfully either to service coverage or cost reduction; removing key operators leads to fewer served requests, higher generalized costs, and less efficient modal usage.

Extensive numerical experiments on both small- and large-scale instances provide several substantive findings. In the Jiuzhaigou emergency-rescue case, the fully integrated eVTOL–drone–GV fleet achieves 100% demand fulfillment at the lowest unit cost among all fleet compositions, underscoring the benefit of combining fast aerial modes with high-capacity ground transport in rugged, time-critical environments; in the Chengdu urban case, the same integrated system maintains a 100% service rate over a wide range of passenger–parcel demand splits while keeping unit costs in a narrow band and substantially reducing emission-related costs as the parcel share grows and drones are used more intensively. Priority-strategy experiments for hybrid passenger–parcel operations show that giving passengers higher priority eliminates delay penalties and lowers total request costs, at the expense of slightly higher fleet usage and travel distance, whereas equal-priority policies reduce fleet deployment but generate substantial delay penalties and degraded passenger service. The framework scales to large instances with up to 200 requests and multiple fleet compositions, and across all scales the integrated eVTOL–drone–GV fleet provides the best overall balance between unit cost, coverage, and emissions: it serves more requests at lower cost than any single- or dual-mode alternative while keeping carbon-tax costs significantly lower. Load-factor analysis reveals that when request time windows are aligned and payloads are moderate, vehicles can pool multiple heterogeneous requests, achieving high utilization and serving more than four requests per vehicle on average; when requests are payload-heavy, capacity constraints rather than temporal compatibility bind, so each vehicle serves fewer requests even though load factors remain high. A capital-cost analysis further shows that, despite higher per-vehicle capital cost, multifunction vehicles that jointly carry passengers and parcels require fewer vehicles overall, achieve higher utilization, and yield lower total capital and service costs than separate passenger-only and parcel-only fleets. Sensitivity analyses indicate that maintaining at least one flexible mode, e.g., flexible eVTOLs or drones, is crucial for high service rates and cost control, that increasing vehicle speed and reducing unit-kilometer cost for aerial modes jointly lower total and transit costs with only modest increases in fleet size or iterations, and that insufficient battery capacity for eVTOLs, drones, and electric trucks severely limits reachable trips and service coverage; once capacity passes certain thresholds all planned requests become feasible and transshipment can drop sharply, but further capacity increases yield diminishing returns unless supported by appropriate charging infrastructure and network design.

The proposed modeling framework is not specific to Chengdu or Jiuzhaigou, in China. By calibrating the input data and constraints, e.g., vehicle and drone performance, network distances, and regulatory restrictions, to reflect local conditions, the model is adaptable to different cities or countries. For instance, global guidelines from aviation authorities define vertiport and flight-zone specifications that could be incorporated as constraints. Despite these promising results, a key limitation is the assumption of a pre-existing network of vertiports and transshipment hubs. In practice, infrastructure availability, siting feasibility, e.g., zoning, land, noise, environmental approval, and grid access, and regulatory processes are binding constraints that shape what routes are even feasible. To relax this assumption, future work should endogenize infrastructure planning within the optimization via an integrated location–routing extension that chooses a subset of candidate sites under budget and policy limits, attaches fixed and expansion costs to opening or upgrading facilities, and links route feasibility to whether sites are actually available. A dynamic space allocation mechanism for passengers and parcels can also be further explored, which could consider real-time demand and passengers' willingness to share, and an optimization approach can be studied to achieve flexible adaptation and dynamic adjustment of passengers and parcels in different vehicles. Site-level capacity and service quality can be captured through throughput limits, pad/berth counts, charger power, and handling times, with simple queueing or piecewise-linear approximations to reflect congestion; airspace management and regulation can be represented by no-fly zones, altitude/corridor rules, separation minima, and curfews that restrict candidate locations and arcs. Staged deployment and uncertain approvals can be modeled in a multi-period or robust setting that explores scenario-based investment paths and contingency measures. Alongside infrastructure, operator acceptance remains crucial; incentive-compatible profit-sharing and supportive policy instruments can encourage participation. More realistic operations should also model UTM procedures, battery state-of-charge and recharge requirements, stochastic disruptions (weather, reliability, demand volatility), and real-time data feeds for dynamic re-routing and decentralized, multi-agent coordination. These extensions would further enhance the realism, robustness, and policy relevance of air–ground multimodal transport planning in the low-altitude economy.

CRediT authorship contribution statement

Yimeng Zhang: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization; **Chenjie Yang:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization; **Haoning Xi:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization; **Songhan Peng:** Writing – review & editing, Visualization, Software, Investigation; **Junjie Yang:** Writing – review & editing, Software, Methodology, Formal analysis; **Mi Gan:** Writing – review & editing, Validation, Investigation, Conceptualization; **Xiaobo Liu:** Writing – review & editing, Validation, Conceptualization; **Ruixue Ai:** Writing – review & editing, Visualization, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

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

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 52402521), 2026 UON CHSF Pilot Research Scheme (No. G2600052), and Scientific and Technological Research and Development Program of China State Railway Group Co., Ltd. (No. N2025X022-B(JB)). This work is also partly co-funded by Sichuan Science and Technology Program (No. 2025NSFSC1969), Fundamental Research Funds for the Central Universities of China (No. 2682025CX056), National Natural Science Foundation of China (No. U2568219, 52372306, 52232011), and Sichuan Science and Technology Program "PIANJI" Project (No. 2025HJPJ0011).

Appendix A. Calculation of load factor and unit cost

This section details the computation logic for the load factor and the cost per unit kg-kilometer. The total load (q_{total}), serving as the basis for these calculations, is the sum of the loads of all served requests:

$$q_{\text{total}} = \sum_{\substack{r \in R \\ k \in K, (i,j) \in A}} q_r y_{ij}^{kr} \tag{A.1}$$

where q_r is the load of request r, and y_{ij}^{kr} is the binary decision variable for assignment.

Load Factor

To calculate the load factor, the weighted vehicle capacity (Λ) is first determined based on the vehicle types and usage proportions:

$$\Lambda = n_k \times \left(u_1 p_1 + u_2 p_2 + u_3 p_3\right) \tag{A.2}$$

where n_k represents the number of used vehicles, u_1, u_2, u_3 denote the capacity of eVTOL, GV, and Drone, respectively, and p_1, p_2, p_3 represent the proportion of requests served by eVTOL, GV, and Drone, respectively.

The load factor (Ω), measuring the utilization of theoretical capacity, is calculated as:

$$\Omega = \frac{q_{\text{total}}}{\Lambda} \tag{A.3}$$

Unit Cost per kg-km

The unit cost (c_{dq}) represents the average transport cost per kilogram per kilometer. It involves the total transport distance (d_{total}):

$$d_{\text{total}} = \sum_{k \in K, (i,j) \in A} d_{ij}^k x_{ij}^k \tag{A.4}$$

where d_{ij}^k represents the distance of arc (i, j) traveled by vehicle k and x_{ij}^k is the route assignment variable.

Finally, c_{dq} is computed as:

$$c_{dq} = \frac{F'}{d_{\text{total}}} \times q_{\text{total}} \tag{A.5}$$

where F' denotes the total transport cost.

References

Agatz, N., Bouman, P., Schmidt, M., 2018. Optimization approaches for the traveling salesman problem with drone. Transp. Sci. 52 (4), 965–981.

All About Industries, 2025. How much power does an eVTOL actually need. https://www.all-about-industries.com/how-much-power-does-an-evtol-actually-need-a-deff7888cd04eadaa93a5b9ca9a845d4. Accessed: 2025-12-02.

Athanasopoulos, K., Chatziioannou, I., Boutsi, A.-M., Tsingenopoulos, G., Soile, S., Chliverou, R., Petrakou, Z., Papanikolaou, E., Karolemeas, C., Kourmpa, E., et al., 2024. Integrating cargo bikes and drones into last-mile deliveries: Insights from pilot deliveries in five greek cities. Sustainability 16 (3), 1060.

Babaee Tirkolaee, E., Cakmak, E., Karadayi-Usta, S., 2025. Traveling salesman problem with drone and bicycle: multimodal last-mile e-mobility. Int. Transac. Oper. Res. 32 (6), 3232–3258.

Bridgelall, R., 2022. Reducing risks by transporting dangerous cargo in drones. Sustainability 14 (20), 13044.

Choi, J.H., Park, Y., 2022. Exploring economic feasibility for airport shuttle service of urban air mobility (UAM). Transp. Res. Part A: Policy Prac. 162, 267–281.

Dajiang, 2025. Latest DJI Drone Price List. https://www.dji.com/. Accessed: 2026-02-10.

DiDi, 2025. DiDi transportation platform. https://www.didiglobal.com/. Accessed: 2026-02-10.

Doo, J., Pavel, M., Didey, A., Hange, C., Diller, N., Tsairides, M., Smith, M., Bennet, E., Bromfield, M., Mooberry, J., 2021. NASA electric vertical takeoff and landing (eVTOL) aircraft technology for public services–a white paper: Nasa transformative vertical flight working group 4 (tvf4). Technical Report. NASA.

Englert, M., Röglin, H., Vöcking, B., 2014. Worst case and probabilistic analysis of the 2-opt algorithm for the TSP. Algorithmica 68 (1), 190–264.

EV. guide, 2025. EV battery capacity & estimating range. https://www.ev.guide/lesson-articles/ev-battery-capacity-and-estimating-range. Accessed: 2025-12-02.

Gonzalez-R, P.L., Canca, D., Andrade-Pineda, J.L., Calle, M., Leon-Blanco, J.M., 2020. Truck-drone team logistics: A heuristic approach to multi-drop route planning. Transp. Res. Part C: Emerg. Technol. 114, 657–680.

Grepow, 2025. What is an eVTOL and how to choose batteries for it. https://www.grepow.com/blog/what-is-an-evtol-and-how-to-choose-batteries-for-it.html. Accessed: 2025-12-02.

He, Q., Liu, W., Xi, H., 2025. Dynamic electric vehicle fleets management problem for multi-service platforms with integrated ride-hailing, on-time delivery, and vehicle-to-grid services. Transp. Res. Part B: Methodol. 199, 103281.

Huang, H., Su, J., Wang, F.-Y., 2024. The potential of low-altitude airspace: The future of urban air transportation. IEEE Transac. Intell. Veh. 9(8), 5250–5254.

Kellermann, R., Biehle, T., Fischer, L., 2020. Drones for parcel and passenger transportation: A literature review. Transp. Res. Interdisc. Perspect. 4, 100088.

Lewis, P. A.W., Shedler, G.S., 1976. Simulation of nonhomogeneous poisson processes with log linear rate function. Biometrika 63 (3), 501–505.
Li, W., Hu, H., Zhou, J., 2022. [retracted] design method of intelligent touchpoint: Intelligent auto-loading cargo transport vehicle for automobile passenger transportation. Mobile Info. Syst. 2022 (1), 4563307.
Lobo Leasing, 2021. eVTOL Financing. Technical Report. Lobo Leasing. White paper. https://cdn.prod.website-files.com/679cb2fe14bbc38f3ee387d9/679cb2fe14bbc38f3ee389d6_eVTOL%20Financing%20(Lobo%20Leasing)_June%202021.pdf.
Lu, C.-C., Diabat, A., Li, Y.-T., Yang, Y.-M., 2022. Combined passenger and parcel transportation using a mixed fleet of electric and gasoline vehicles. Transp. Res. Part E: Logist. Transp. Rev. 157, 102546.
Luo, Y., Zhang, Y., Huang, J., Yang, H., 2021. Multi-route planning of multimodal transportation for oversize and heavyweight cargo based on reconstruction. Comp. Oper. Res. 128, 105172.
Meng, S., Guo, X., Li, D., Liu, G., 2023. The multi-visit drone routing problem for pickup and delivery services. Transp. Res. Part E: Logist. Transp. Rev. 169, 102990.
Michigan Economic Development Corporation, 2024. Lt. governor gilchrist announces over $6 million to four projects to scale advanced aerial mobility infrastructure. https://www.michiganbusiness.org/press-releases/2024/07/advanced-aerial-mobility-funding/.
Miyoshi, H., Zhang, Y., Azadeh, S.S., Cats, O., 2025. Dynamic fleet management of waterborne vessels with mixed passenger and parcel services. npj Sustain. Mobil. Transp. 2 (1), 16.
Mohamed, A., Mohamed, M., 2025. Unmanned aerial vehicles in last-mile parcel delivery: A state-of-the-art review. Drones 9 (6), 413.
Mulumba, T., Najy, W., Diabat, A., 2024. The drone-assisted pickup and delivery problem: An adaptive large neighborhood search metaheuristic. Comp. Oper. Res. 161, 106435.
Murray, C.C., Chu, A.G., 2015. The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. Transp. Res. Part C: Emerg. Technol. 54, 86–109.
Petalite, 2025. Fflip demonstrates scalable charging infrastructure for evtol & electric transport. https://www.aaminternational.com/2025/04/fflip-demonstrates-scalable-charging-infrastructure-for-evtol-electric-transport/.
Ren, X., Wang, J., 2025. Symbiotic evolution mechanism of urban air mobility industrial innovation ecosystem: Evidence from low altitude air mobility in shenzhen. Journal of Air Transport Management 124, 102750.
Roa, J., 2022. Opportunities and challenges using hybrid and electric aircraft for passenger and cargo operations. In: International Conference on Transportation and Development 2022, pp. 190–200.
Ropke, S., Pisinger, D., 2006. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transp. Sci. 40 (4), 455–472.
Sacramento, D., Pisinger, D., Ropke, S., 2019. An adaptive large neighborhood search metaheuristic for the vehicle routing problem with drones. Transp. Res. Part C: Emerg. Technol. 102, 289–315.
Schermer, D., Moeini, M., Wendt, O., 2019. A hybrid VNS/tabu search algorithm for solving the vehicle routing problem with drones and en route operations. Comp. Oper. Res. 109, 134–158.
SkyDrive Inc., 2025. Skydrive, 7a drones, and ITRI partner to pioneer eVTOL medical transport in penghu, taiwan. https://en.skydrive2020.com/archives/15479.
Sripad, S., Viswanathan, V., 2021. The promise of energy-efficient battery-powered urban aircraft. Proc. National Acad. Sci. 118 (45), e2111164118.
Su, E., Qin, H., Li, J., Zhang, R., 2025. The freight multimodal transport problem with buses and drones: An integrated approach for last-mile delivery. arXiv preprint arXiv:2506.10311 .
Thomas, T., Srinivas, S., Rajendran, C., 2024. Collaborative truck multi-drone delivery system considering drone scheduling and en route operations. Annal. Oper. Res. 339 (1), 693–739.
UK Government, 2025. Over £20 million to help drones and flying taxis take to UK skies. https://www.gov.uk/government/news/over-20-million-to-help-drones-and-flying-taxis-take-to-uk-skies.
Wang, Z., Sheu, J.-B., 2019. Vehicle routing problem with drones. Transp. Res. Part B: Methodol. 122, 350–364.
Xi, H., Tang, Y., Waller, S.T., Shalaby, A., 2023. Modeling, equilibrium, and demand management for mobility and delivery services in mobility-as-a-service ecosystems. Computer-Aided Civil Infrast. Eng. 38 (11), 1403–1423.
Xiang, S., Xie, A., Ye, M., Yan, X., Han, X., Niu, H., Li, Q., Huang, H., 2024. Autonomous eVTOL: A summary of researches and challenges. Green Energy Intell. Transp. 3 (1), 100140.
Yan, Y., Wang, K., Qu, X., 2024. Urban air mobility (UAM) and ground transportation integration: A survey. Front. Eng. Manag. , 1–25.
Zandieh, F., Ghannadpour, S.F., Mazdeh, M.M., 2023. Integrated ground vehicle and drone routing with simultaneous surveillance coverage for evading intentional disruption. Transport. Res. Part E: Logist. Transport. Rev. 178, 103266.
Xi, H., Shao, Z., Hensher, D.A., Nelson, J.D., Chen, H., Wijayaratna, K., 2025. A multi-task Transformer with mixture-of-experts for personalized periodic predictions of individual travel behavior in multimodal public transport. Transportation Research Part C: Emerging Technologies 179, 105287.
Zhai, R., Mei, Y., Guo, T., Du, W., 2024. A collaborative drone-truck delivery system with memetic computing optimization. IEEE Transact. Syst., Man, Cybern.: Syst..
Zhan, W., Liao, Y., Deng, J., Wang, Z., Yeh, S., 2025. Large-scale empirical study of electric vehicle usage patterns and charging infrastructure needs. npj Sustain. Mobil. Trans. 2 (1), 9.
Zhan, X., Szeto, W.Y., Wang, Y., 2023. The ride-hailing sharing problem with parcel transportation. Transport. Res. Part E: Logist. Transport. Rev. 172, 103073.
Zhao, Y., Peng, P., Zhou, J., Wang, Y., 2025. Heuristic algorithm for integrated ship scheduling, routing and stowage problem in multi-vessel roll-on/roll-off shipping. J. Heuris. 31 (1), 15.
Hörsting, L., Cleophas, C., 2023. Scheduling shared passenger and freight transport on a fixed infrastructure. European Journal of Operational Research 306 (3), 1158–1169.
Yang, X., Kong, X.T., Huang, G.Q., 2024. Synchronizing crowdsourced co-modality between passenger and freight transportation services. Transportation Research Part E: Logistics and Transportation Review 184, 103490.
Zheng, S, Wang, K., Jiang, C., 2024. Converting passenger aircraft into cargo planes under volatile market demand. Transport. Res. Part A: Pol. Pract. 181, 104013.
Zhu, Z., Zhang, S., Cao, Z., 2025. Exact and heuristic methods for the integrated optimisation of vessel deployment and liner shipping routing schedule. Intern. J. Syst. Sci.: Operat. Logist. 12 (1), 2505920.