mathematics

MDPI

Article

Multi-UAV Emergency Scheduling Under Demand Accumulation and Payload-Dependent Energy

Haifei Yang [1], Ganwei Cai [2,4,5,*], Junjie Gong [1,3] and Yu Long [5]

[1] School of Business, Guilin University of Electronic Technology, Guilin 541004, China; yanghf@mails.guet.edu.cn (H.Y.); gongjj@guet.edu.cn (J.G.)
[2] School of Mechanical Engineering, Anhui Institute of Information Technology, Wuhu 241000, China
[3] Key Laboratory of Intelligence Integrated Automation in Guangxi Universities, Guilin University of Electronic Technology, Guilin 541004, China
[4] School of Artificial Intelligence, Wuchang University of Technology, Wuhan 430223, China
[5] School of Mechanical Engineering, Guangxi University, Nanning 530004, China; longyu@gxu.edu.cn
* Correspondence: cganwei@gxu.edu.cn

Abstract

Emergency delivery by multiple unmanned aerial vehicles (UAVs) requires scheduling decisions that account for demand accumulation, payload-dependent energy consumption, and the feasibility of successive sorties. To address this coupled problem, this paper proposes an energy-aware discrete black-winged kite algorithm with local search and insertion screening (EDBKA-LS-I). The model minimizes cumulative unmet demand exposure subject to payload, battery, return energy, and mission time constraints. Payload-dependent energy is calibrated from public flight records, and a sufficient speed dominance condition supports preprocessing under no-wait landing delivery. Geometric and random screening select insertion candidates for complete evaluation, while bounded local search refines UAV assignments, the sortie order, and customer sequences. Nine synthetic and public layout-derived instances were tested with twenty seeds per method under matched evaluation budgets. Three instances were additionally tested with five seeds per method under a six-second search budget. Relative to adaptive large neighborhood search with identical screening, EDBKA-LS-I reduced the mean cumulative unmet demand exposure by 7.74% and 7.93% under the respective budgets. EDBKA-LS-I also obtained lower mean exposure than variable neighborhood search on seven of nine instances under the evaluation budget. EDBKA-LS-I achieved complete service in all main comparisons. The energy model attained a mean absolute prediction error of 0.993 Wh on forty held-out delivery flights.

Keywords: multi-UAV emergency scheduling; demand accumulation; payload-dependent energy; local search; insertion screening

MSC: 90B06; 90C27; 90C59

Check for updates

Academic Editor: Renato Bruni

Received: 3 October 2026
Revised: 19 October 2026
Accepted: 20 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The timely delivery of essential supplies is required when emergency response is disrupted by damaged infrastructure and geographically dispersed demand [1,2]. Unmanned aerial vehicles (UAVs) can provide aerial access to affected communities. Attenni et al. [3] reviewed route planning for drone-only delivery, while Duan et al. [4] synthesized solution approaches for coordinated truck and drone routing. Chen et al. [5] proposed adaptive ant colony search for UAV routing with lateness and safety objectives. Within disaster response,

Khalaj Rahimi and Rahmani [6] incorporated deprivation penalties to represent the urgency of assistance. To accommodate flexible relief distribution, Mahmoudi and Eshghi [7] formulated a split-delivery model with en route drone launches and rendezvous. Calamoneri et al. [8] minimized the total completion time in UAV-based post-disaster assessment. Tureci-Isik et al. [1] formulated a two-stage stochastic model linking pre-disaster depot locations with post-disaster parallel truck–drone routing for humanitarian aid. Zhao et al. [2] addressed pharmaceutical delivery after road disruptions through coordinated mother-end drones, using distributionally robust optimization for uncertain demand and flight times. In delivery operations with accumulating demand, however, a service delay also increases the parcel mass required at the customer. Energy checks must therefore use payloads calculated at the scheduled service completion time. A later depot return also delays subsequent sorties on the same UAV and increases the loads required for such deliveries. Route modifications must consequently be evaluated together with UAV assignments and the sortie order. This dependence makes the choice of candidates for complete evaluation important regarding computational efficiency.

Route feasibility depends on the energy required to carry the assigned load over each flight leg. Rodrigues et al. [9] quantified the delivery energy from measured flights. Wu et al. [10] incorporated the payload and speed into energy-efficient drone routing, bringing these effects into route evaluation. Guo et al. [11] incorporated wind and payload effects into energy-aware vehicle–UAV route planning. Shangguan et al. [12] modeled UAV–rider dispatching under stochastic wind, coupling nonlinear energy consumption with scenario-dependent flight plans. For coordinated truck and drone delivery, Cicek et al. [13] used learned energy estimates, while Ren et al. [14] developed planning based on energy predictions for drone services. When such estimates are used to evaluate routes, the cruise speed must be considered together with energy consumption and service timing. Tamke and Buscher [15] investigated speed selection and dominated-speed preprocessing for truck and drone routing. For multi-trip UAV energy minimization, Büyüksan et al. [16] combined surrogate power modeling with a battery-constrained split decoder. Addressing the interaction between speed and coordinated service, Ji et al. [17] jointly optimized speed, routing, and task allocation under nonlinear energy constraints. At the fleet level, Kong et al. [18] coordinated visits to multiple customers within each flight with ground-vehicle operations. Jiang et al. [19] formulated flexible docking decisions, while Jiang et al. [20] incorporated heterogeneous UAV capabilities and nonlinear energy consumption into multi-visit routing. To connect replenishment with time-sensitive service, Bhuiyan et al. [21] integrated battery replacement into fleet deployment. Under uncertain demand, Roytvand Ghiasvand et al. [22] formulated a data-driven robust truck and drone routing model. Xing et al. [23] addressed travel time uncertainty through reliable dynamic synchronization, and Chagas et al. [24] incorporated joint demand and energy uncertainty into dynamic drone routing. Sun et al. [25] formulated robust truck–drone routing for post-disaster delivery and surveillance under travel time uncertainty. Energy estimation, vehicle assignment, and temporal coordination are therefore closely connected within these formulations. In the present setting, where demand grows until service completion, an earlier route change can alter the payload and energy feasibility of subsequent sorties. Candidate evaluation must therefore propagate changes in completion time, payload, and energy consumption through the affected UAV schedule.

Neighborhood search provides several ways to generate candidate schedules. Vidal [26] introduced the SWAP* customer exchange neighborhood within hybrid genetic search, and Wouda et al. [27] implemented hybrid genetic search in PyVRP. For coordinated truck and drone delivery, Madani et al. [28] organized routing, launch, and retrieval decisions into categories within adaptive variable neighborhood search. Removal and repair allow more

extensive route changes. Wouda and Lan [29] presented an implementation of adaptive large neighborhood search, abbreviated as ALNS, and Voigt [30] compared its operators to examine the effects of different removal and insertion choices. Wu and Bao [31] combined ALNS with local search for two-echelon routing with clustered customers. Lorenz et al. [32] used structured subproblems in a very large-scale neighborhood search for energy replenishment routing. Applying these methods to accumulating demand requires each candidate to be evaluated through its effects on completion times, parcel masses, and return energy feasibility. A small insertion detour can still postpone later sorties and increase demand exposure. The search problem is therefore to identify promising insertion positions while retaining the ability to change UAV assignments, the sortie order, and customer sequences within the available evaluation budget.

To address this problem, an energy-aware discrete black-winged kite algorithm with local search and insertion screening is proposed and denoted as EDBKA-LS-I. Wang et al. [33] proposed the black-winged kite algorithm, abbreviated as BKA, whose population framework is adapted here to discrete delivery schedules. This study aims to demonstrate that EDBKA-LS-I can reduce the mean cumulative unmet demand exposure in multi-UAV emergency delivery while satisfying payload and energy constraints under limited computational budgets. The main contributions are as follows. First, the service-dependent demand is coupled with the payload and phase-based energy calibrated from public flight records, and a sufficient speed dominance condition is derived for no-wait landing delivery. Second, bounded local search is combined with geometric and random insertion screening to refine UAV assignments, the sortie order, and customer sequences through complete schedule evaluation. Third, the contributions of these mechanisms are assessed through component ablations and comparisons with an ALNS implementation using the same evaluator and screening. Held-out flight validation assesses the energy prediction accuracy, while routing experiments under matched evaluation and time budgets assess scheduling performance.

The rest of this paper is organized as follows. The delivery setting and mathematical model are presented in Section 2. The proposed algorithm is described in Section 3. The compared methods, experimental results, and analysis are presented in Section 4. Conclusions are presented in Section 5.

2. Problem Definition and Mathematical Formulation

2.1. Delivery Setting and Modeling Assumptions

A set of n customers, $\mathcal{N} = \{1, \ldots, n\}$, is served by K homogeneous UAVs from a common depot, denoted by node 0. Customer locations, initial demands, and priority weights are specified at the planning instant. Demand accumulation rates are specified as nominal forecasts and held fixed during schedule evaluation. A deterministic schedule is constructed over a horizon H. Each UAV may perform several successive sorties, and each sortie starts and ends at the depot. The decisions comprise customer allocation and the visiting order within each sortie, UAV assignment, and the sortie order on each UAV. A cruise speed is also selected from a finite set $\mathcal{V}$.

Landing delivery comprises a climb, horizontal cruise, and descent on every leg, including customer-to-customer travel and the final depot return. A fixed ground-service duration is assigned to each customer, and parcels are unloaded at service completion. Flight legs are represented by straight horizontal segments at a common altitude. Wind, obstacles, inter-UAV separation, and airspace conflicts are excluded from this mission scheduling and routing model.

At the depot, a fixed turnaround duration represents preparation and battery replacement. Sufficient replacement batteries are assumed, and every sortie departs with the same available energy. This fixed-turnaround representation abstracts the battery inventory,

charger capacity, and charging queues. Joint battery replacement scheduling [34] offers a formulation for explicitly modeling these resource decisions.

The model represents a short first-response delivery round to isolated aid posts or relief collection points. Requests for compact supplies, such as dressing materials or water purification tablets, remain outstanding until handover. Each served site receives one consolidated batch in a single visit, limiting repeated landings and handling. This policy applies when the batch satisfies the UAV payload and energy limits. Requests arising after handover belong to subsequent delivery rounds. Partial plans are allowed, with omitted customers assigned the horizon in the objective and counted as unserved. The principal scheduling and energy symbols are summarized in Table 1.

Table 1. Principal notation for the scheduling and energy model.

Symbol	Definition
ℓ_{ij}	Horizontal distance from node i to node j in m
d_i^0, λ_i, w_i	Initial demand in kg, accumulation rate in kg/s, and priority weight
H, s_i, τ	Planning horizon, ground service duration, and depot turnaround in s
π_r, k_r, v_r	Customer sequence, assigned UAV, and cruise speed of sortie r
D_r, R_r, C_i	Sortie departure, depot return, and customer service completion times in s
h, u_a, u_d	Cruise altitude in m and climb and descent speeds in m/s
$q_i, Q_{r,j}$	Delivered mass and onboard payload before leg j in kg
$Q_{\max}, B, B_{\text{res}}$	Payload limit in kg and starting and reserve energy in Wh
$\mu, E_r, E_{r,j}^{\text{abort}}$	Relative energy margin and sortie and abort return energy in Wh
J	Normalized cumulative unmet demand exposure

2.2. Time Recursion and Service-Dependent Payload

A plan with R sorties is represented as an ordered collection $\mathcal{P} = \{(\pi_r, k_r, v_r)\}_{r=1}^{R}$. For sortie r, π_r is the customer sequence, k_r identifies the assigned UAV, and v_r is its cruise speed. The nonempty sequence $\pi_r = (i_{r,1}, \ldots, i_{r,m_r})$ contains m_r customers, with $i_{r,j}$ denoting the customer visited at position j. Each customer occurs in at most one sequence. The relative order of entries assigned to the same UAV determines its sortie order. Different UAVs may operate concurrently. Let r^- be the preceding sortie assigned to UAV k_r, and let a_k be the initial availability of UAV k. Departure is determined by

$$D_r = \begin{cases} a_{k_r}, & \text{if no preceding sortie exists,} \\ R_{r^-} + \tau, & \text{otherwise.} \end{cases} \tag{1}$$

The depot endpoints are represented by $i_{r,0} = i_{r,m_r+1} = 0$. The duration $t_{r,j}^{\text{fly}}$ of leg j is

$$t_{r,j}^{\text{fly}} = t_a + \frac{\ell_{i_{r,j-1}, i_{r,j}}}{v_r} + t_d, \qquad t_a = \frac{h}{u_a}, \quad t_d = \frac{h}{u_d}, \tag{2}$$

where, t_a and t_d are the climb and descent durations in seconds. Travel and service durations are independent of the payload within the supported operating range. With no earliest-service waiting, let $T_{r,j}$ denote the service completion time at position j of sortie r, initialized by $T_{r,0} = D_r$ at departure. Customer completion and depot return times are evaluated as

$$C_{i_{r,j}} = T_{r,j} = T_{r,j-1} + t_{r,j}^{\text{fly}} + s_{i_{r,j}}, \qquad j = 1, \ldots, m_r, \tag{3}$$

$$R_r = T_{r,m_r} + t_{r,m_r+1}^{\text{fly}}. \tag{4}$$

These recursions include the effect of an earlier sortie on every subsequent sortie assigned to the same UAV. In particular, changing a UAV assignment may alter the completion times even when all customer sequences remain unchanged.

The delivered mass is $q_i = d_i^0 + \lambda_i C_i$, where $d_i^0 \geq 0$ is the initial outstanding quantity in kilograms and $\lambda_i \geq 0$ is the forecast accumulation rate in kilograms per second. The constant λ_i approximates a short horizon with stable request arrival rates and an average requested mass. Both demand parameters may vary across customers. Supplies become available to the customer at service completion. For each candidate schedule, C_i is predicted before dispatch and q_i is loaded at the depot, including forecast accumulation during travel and ground service. The payload before each customer leg and on the final return is

$$Q_{r,j} = \sum_{u=j}^{m_r} q_{i_{r,u}} \quad (j = 1, \ldots, m_r), \qquad Q_{r,m_r+1} = 0. \tag{5}$$

Earlier completion reduces the required package mass, while postponing a sortie can increase its payload and those of later sorties. Evaluation therefore proceeds in two passes: completion times are calculated first, followed by package masses, onboard payloads, and energy.

Let ϵ_i denote the relative deviation of the actual demand accumulation rate from its forecast, so that $\lambda_i^{\text{act}} = (1 + \epsilon_i)\lambda_i$. At an unchanged completion time C_i, the parcel mass adjustment required to meet the actual demand is $\Delta q_i = \epsilon_i \lambda_i C_i$. Accommodating positive deviations increases the required payload and the corresponding energy consumption. The adjusted loads must therefore be checked against payload, sortie energy, and post-service return energy constraints before departure.

2.3. Phase-Based Energy Model and Calibration

An empirical phase model is used to connect the payload and flight duration to energy. Public flight records [9] are used to estimate the coefficients independently of the routing experiments. For onboard payload q in kilograms and cruise speed v in meters per second, the cruise power $P_c(q, v)$ in watts is represented by

$$P_c(q, v) = a + b\frac{q}{q_*} + c\left(\frac{v - v_c}{v_*}\right)^2 + d\frac{q}{q_*}\frac{v}{v_{**}}, \tag{6}$$

where, a is the cruise power intercept, b weights the normalized payload, c weights the quadratic speed term, and d weights the interaction between the payload and speed. All four coefficients have units of watts. The parameter v_c centers the quadratic speed term and is measured in meters per second. The fixed scales $q_* = 0.5$ kg, $v_* = 8$ m/s, and $v_{**} = 12$ m/s normalize the payload, the quadratic speed term, and the speed in the interaction term, respectively. The ascent and descent powers are $P_a(q) = a_a + b_a q/q_*$ and $P_d(q) = a_d + b_d q/q_*$, respectively, where a_a, a_d are the phase intercepts and b_a, b_d are the corresponding normalized payload coefficients, all in watts. Ground service uses constant idle power P_0. The model combines a nondecreasing payload response with a quadratic speed term over the supported interval.

The measured discharge power is reconstructed from the battery voltage and signed discharge current. The power is set to zero for samples with a current opposite to the discharge sign convention. Duplicate timestamps are removed, and flights that fail timestamp, payload, or power range checks are excluded. Flight phases are identified from the altitude, vertical rate, and horizontal velocity. The whole-flight energy is obtained by the trapezoidal integration of the measured power. Training weights equal the inverse square root of the retained sample count in each group defined by flight and phase, moderating the influence of longer records.

Coefficients are fitted by nonnegative least squares with positive intercepts enforced. The fitted coefficients are $(a_a, b_a) = (546.0696, 105.9168)$ and $(a_d, b_d) = (457.5305, 100.2350)$ W, with $P_0 = 1$ W. The ground-idle estimate reached the imposed 1 W lower bound, which was used

for the simulated ground-service energy. The speed center parameter v_c is selected from $\{4, 6, 8, 10, 12\}$ m/s using delivery flights in the validation partition. Training and validation determine the coefficients, cruise model, and energy margin. Test flights are used only for final evaluation. A multiplicative margin μ is obtained from the nonnegative part of the empirical 95th-percentile validation residual $(E^{\text{obs}} - \hat{E})/\hat{E}$, where E^{obs} and $\hat{E}$ are the observed and predicted energies of the same validation flight, expressed in watt-hours. For multi-stop scheduling, the calibrated margin is used as an empirical planning allowance under the assumption of comparable relative energy errors within the modeled operating range. It is applied to the total sortie energy and each post-service abort return estimate in the feasibility checks, so the absolute energy allowance scales with the predicted consumption.

For landing delivery, the flight energy $e_{r,j}$ on leg j of sortie r comprises climb, cruise, and descent. The total sortie energy E_r additionally includes ground service. Both energies are expressed in watt-hours:

$$e_{r,j} = \frac{P_a(Q_{r,j})t_a + P_c(Q_{r,j}, v_r)\ell_{i_{r,j-1},i_{r,j}}/v_r + P_d(Q_{r,j})t_d}{3600}, \tag{7}$$

$$E_r = \sum_{j=1}^{m_r+1} e_{r,j} + \frac{P_0}{3600}\sum_{j=1}^{m_r} s_{i_{r,j}}. \tag{8}$$

The factor 3600 converts watt-seconds to watt-hours in the energy calculation. Every customer stop includes a vertical cycle under the landing delivery assumption. Sortie merging therefore trades reduced depot travel against the energy required to carry the combined payload over longer segments.

2.4. Feasibility Constraints and Optimization Objective

The operating support imposes $v_r \in \mathcal{V}$, h within the calibrated altitude range, and $Q_{r,1} \le Q_{\max}$, where $Q_{\max}$ does not exceed the model's supported payload. Since parcels are only unloaded, checking the initial sortie payload bounds all later payloads. Complete-sortie feasibility requires

$$(1+\mu)E_r \le B - B_{\text{res}}, \qquad R_r \le H. \tag{9}$$

Let $U_{r,j}$ denote the nominal energy consumed from the departure of sortie r through service completion at customer $i_{r,j}$. After unloading, the remaining payload is $Q_{r,j+1}$. If subsequent deliveries are canceled at this point, the total nominal energy consumed from sortie departure through direct return to the depot is

$$E_{r,j}^{\text{abort}} = U_{r,j} + \frac{P_a(Q_{r,j+1})t_a + P_c(Q_{r,j+1}, v_r)\ell_{i_{r,j},0}/v_r + P_d(Q_{r,j+1})t_d}{3600}, \tag{10}$$

$$(1+\mu)E_{r,j}^{\text{abort}} \le B - B_{\text{res}}, \qquad j = 1, \dots, m_r. \tag{11}$$

Equations (10) and (11) apply to direct return after customer service and unloading. They do not guarantee abort return feasibility before unloading or during flight. These situations require separate evaluation using the onboard payload, position, altitude, and accumulated energy at the interruption state.

Priority weights are nonnegative, and the total weighted demand exposure over the planning horizon is positive. At elapsed mission time $t \in [0, H]$, measured in seconds from the planning instant, the outstanding demand at an unserved customer i is $d_i^0 + \lambda_i t$. Set $\widetilde{C}_i = C_i$ for a served customer and $\widetilde{C}_i = H$ otherwise. The priority-weighted waiting burden is $w_i \int_0^{\widetilde{C}_i} (d_i^0 + \lambda_i t)\, \mathrm{d}t$, measured in kilogram-seconds for dimensionless priority weights w_i. Integrating the outstanding quantity over time gives the routing objective

$$\min_{\mathcal{D} \in \mathcal{F}} J(\mathcal{D}) = \frac{\sum_{i \in \mathcal{N}} w_i \left(d_i^0 \widetilde{C}_i + \frac{1}{2}\lambda_i \widetilde{C}_i^2 \right)}{\sum_{i \in \mathcal{N}} w_i \left(d_i^0 H + \frac{1}{2}\lambda_i H^2 \right)}, \tag{12}$$

where $\mathcal{F}$ is the set of feasible plans, including the empty plan. Each nonempty plan comprises nonempty sortie sequences with no repeated customers within or across sorties and satisfies Equations (1) through (11) and the assignment restrictions above. The term $d_i^0 \widetilde{C}_i$ is the mass-weighted waiting time of the initial backlog, while $\lambda_i \widetilde{C}_i^2/2$ is the contribution from requests accumulated during the waiting interval. The denominator is the exposure if every customer remains unserved until H. This fixed positive baseline preserves solution rankings and gives $0 \le J \le 1$, with $J = 1$ for the empty plan. A lower J indicates less waiting exposure. Omitted customers contribute over the full horizon, so additional deliveries are weighed against delays to other customers. Optional service applies when the dispatch policy prioritizes aggregate weighted demand exposure within the current response horizon. A partial schedule may therefore yield lower exposure even when full coverage is feasible. Energy constrains feasibility and breaks ties in J within 10^{-12}, using an energy tolerance of 10^{-9} Wh.

3. Proposed EDBKA-LS-I Algorithm

3.1. Discrete Representation and Evaluation Architecture

EDBKA-LS-I draws on the population search, leader guidance, and greedy candidate acceptance used in BKA [33]. Candidate schedules form the population, and the best feasible schedule provides leader guidance. Candidate generation is formulated directly through route and scheduling operations. The original continuous attack and migration equations and Cauchy mutation are not retained. The components introduced for this paper comprise energy-aware candidate selection, bounded local search over UAV assignments, the sortie order, and customer sequences, as well as insertion screening that combines geometric ranking with random sampling. Each member stores the ordered sortie list defined in Section 2, and empty sorties are removed after route operations. Entropy-triggered recovery is evaluated separately as a component study configuration in Section 4.1.

Construction, perturbation, local search, and all comparison methods share a deterministic evaluator. UAV availability, completion-dependent parcel masses, and energy constraints are recomputed for each candidate. A historical incumbent retains the best feasible plan found by any operator.

Let F denote completed evaluations, with cap $F_{\max}$, and let t_{search} denote the elapsed wall-clock search time, with limit $t_{\max}$. Both times are measured in seconds. Every completed scoring call, including rejected candidates and initial construction, consumes the common evaluation budget. Evaluations begin while both limits permit further search, and an active call is allowed to finish. Consequently, completion of the last call may extend slightly beyond the time limit. The final plan is reevaluated and independently checked after the timed search.

3.2. Conditional Speed Dominance Preprocessing

Eliminating dominated speeds before route search has been considered in drone routing research [15]. Here, a sufficient condition is derived for Equation (6) and the landing assumptions. For a fixed payload, differentiation gives

$$\frac{\partial (P_c(q,v)/v)}{\partial v} = \frac{c}{v_*^2} - \frac{a + bq/q_* + cv_c^2/v_*^2}{v^2}. \tag{13}$$

The interaction between the payload and speed contributes the constant $dq/(q_* v_{**})$ to the energy per distance and therefore vanishes from this derivative. Let $v_{\max} = \max \mathcal{V}$ be

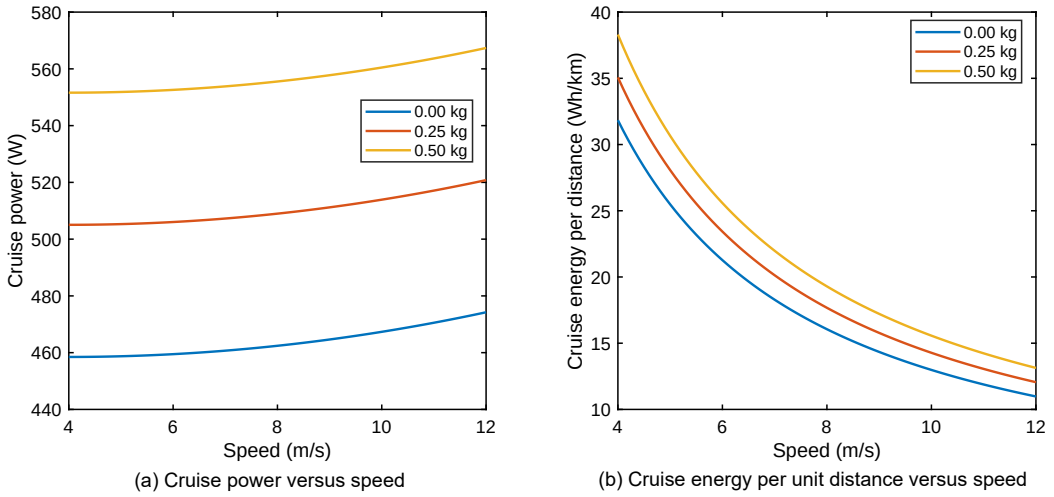
the largest permitted cruise speed in meters per second. With nonnegative coefficients and $q \geq 0$, a sufficient condition over $0 < v \leq v_{\max}$ is

$$\frac{c}{v_*^2} - \frac{a + cv_c^2/v_*^2}{v_{\max}^2} \leq 0. \tag{14}$$

Dominance statement. Under nonnegative demand accumulation, fixed service durations, landing delivery without earliest-time waiting, and nondecreasing phase power with payload, the fastest permitted speed weakly dominates slower choices if Equation (14) holds throughout the supported speed interval.

Justification. Increasing the cruise speed cannot increase the leg duration. Equations (1)–(3) therefore yield no later completion times, first within a sortie and then across successive sorties on each UAV. The required package masses and all onboard payloads cannot increase. Equation (14) makes the cruise energy per distance nonincreasing with speed at a fixed payload. Nonnegative payload coefficients preserve this relation when the payload also decreases. The ascent and descent energies cannot increase, and the ground service energy is unchanged. The same argument applies to each loaded abort return. Thus, the capacity, energy, and horizon feasibility are preserved, and Equation (12) cannot worsen.

Coefficient signs, delivery and waiting assumptions, demand signs, and calibration support are checked before speed reduction. The original speed set is retained whenever a required condition fails. For the fitted law, $(a, b, c, d) = (458.4854, 93.1245, 15.7478, 0)$ and $v_c = 4$ m/s. The derivative upper bound is -2.9652, reducing $\{4, 6, 8, 10, 12\}$ m/s to 12 m/s. Figure 1 shows the corresponding power and energy-per-distance responses used in preprocessing, which is shared among all compared methods.



(a) Cruise power versus speed (b) Cruise energy per unit distance versus speed

Figure 1. Calibrated cruise power in panel (**a**) and energy per unit distance in panel (**b**) at payloads of 0, 0.25, and 0.50 kg. Increasing the speed from 4 to 12 m/s raises the power but reduces the energy per unit distance, supporting the energy condition used in speed preprocessing.

3.3. Initial Construction and Insertion Screening

Construction starts from an empty plan and a random customer order. Existing-sortie insertion positions and one new-sortie option per UAV are evaluated, subject to the insertion allowance. A new sortie is appended to the UAV's schedule. The best feasible insertion is accepted. If no feasible insertion is found within the allowance, the customer remains unserved for later reconsideration. The constructed plan is copied into twelve population slots, with diversity subsequently generated by perturbation.

When the number of positions exceeds the insertion allowance, a geometric shortlist is formed. Inserting customer i between nodes u and w has detour

$$\Delta\ell(u, i, w) = \ell_{ui} + \ell_{iw} - \ell_{uw}. \tag{15}$$

For an insertion limit $L = 18$, K choices are reserved for new sorties. With $M = L - K$ remaining positions, $\lceil 2M/3 \rceil$ positions with the smallest detour are retained, and the remaining positions are sampled without replacement from the other candidates. For the three-UAV experiments, ten geometric positions and five random positions are retained. All existing positions are evaluated when their number does not exceed M. Random sampling preserves alternatives outside the low-detour set. The scheduling evaluator ranks the retained feasible candidates by demand exposure. Figure 2 illustrates this screening procedure.



Figure 2. Insertion screening for three UAVs combines ten low-detour positions, five random alternatives, and three new-sortie options to concentrate evaluations while preserving alternative placements. Labels a, b, c, d denote existing customers, i denotes the customer to insert, and 0 denotes the depot.

An exposure-based ranking was introduced for comparison with geometric screening. Define $f_j(t) = w_j(d_j^0 t + \lambda_j t^2/2)$ as the weighted demand exposure of customer j up to time t. For the insertion of an unserved customer i at position p in sortie r, the ranking score in kilogram-seconds is

$$\widehat{\Delta A}_{i,p} = f_i(\widehat{C}_i^p) - f_i(H) + \sum_{j \in \mathcal{A}_{i,p}} \left[f_j(C_j + \delta_{i,p}) - f_j(C_j) \right], \tag{16}$$

where $\widehat{C}_i^p$ is the predicted completion time of the inserted customer and $\mathcal{A}_{i,p}$ contains the customers following position p in sortie r and all customers in later sorties assigned to the same UAV. Under no-wait landing delivery, the additional time is $\delta_{i,p} = \Delta\ell(u, i, w)/v_r + t_a + t_d + s_i$, where u and w are the nodes adjacent to insertion position p in sortie r. The first term gives the change in the inserted customer's exposure, and the sum accounts for delays to existing customers. Lower scores receive priority. Completion times and cumulative exposure coefficients are cached to calculate these scores without energy evaluation. Both screening rules retain the same random-candidate proportion and new-sortie options. All retained candidates undergo the complete, budget-counted evaluation.

3.4. Population Perturbation and Energy-Aware Selection

Population members are visited cyclically. At each ordinary iteration, three candidates are generated from the selected member, using an energy-aware perturbation with probability 0.65 or a generic discrete operator otherwise. The best feasible candidate or unchanged member is retained. Any improving scoring call also updates the historical incumbent.

Standard routing operators include customer swaps, route reversal, relocation, sortie splitting and merging, UAV reassignment, and sortie order changes. These operators are supplemented by bounded multiple relocations and leader adjacency guidance. Multiple

relocations use a heavy-tailed step clipped to one through five moves. Leader guidance samples a consecutive customer pair from the best feasible schedule and attempts to place the second customer immediately after the first in the selected population member. The leader's information is thereby transferred through customer adjacency.

Energy-aware selection gives a higher sampling probability to sorties with smaller energy slack. The sampling stress score is $0.05 + \max\{0, 1 - \sigma_r\}$, where $\sigma_r = [B - B_{\text{res}} - (1 + \mu)E_r]/(B - B_{\text{res}})$ is the dimensionless energy slack after applying the planning margin, normalized by the usable battery energy. Within a selected sortie, a customer's sampling weight is proportional to $0.001 + q_i L_i$. Here, L_i is the horizontal distance over which the parcel has been carried before arrival. A selected customer may be moved earlier, transferred to another sortie, or used to define a split. Merge and speed moves are also available. This heuristic, based on the carried mass and distance, selects candidates for complete schedule and energy reevaluation. A move is accepted according to demand exposure and the feasibility constraints.

3.5. Bounded Route and Scheduling Improvement

Every fourth iteration applies local search to the incumbent. Three neighborhood families are applied in rotation. The first changes the UAV assignments and same-UAV sortie order, the second relocates or reverses customers within a sortie, and the third relocates customers across sorties. The last family examines all insertion positions in the three closest destination sorties, ranked according to the minimum customer-to-customer distance. Figure 3 shows how reassignment can advance a sortie without changing its customer order. Reassignment changes UAV availability, resequencing changes the sortie departure times, and customer moves change the travel and unloading order. A cross-sortie relocation and its evaluation chain are shown in Figure 4. The completion times, package masses, payloads, and energy are recomputed, including subsequent sorties on the affected UAVs. These neighborhoods adapt routing intensification ideas to the coupling between service time and payload [26,32].

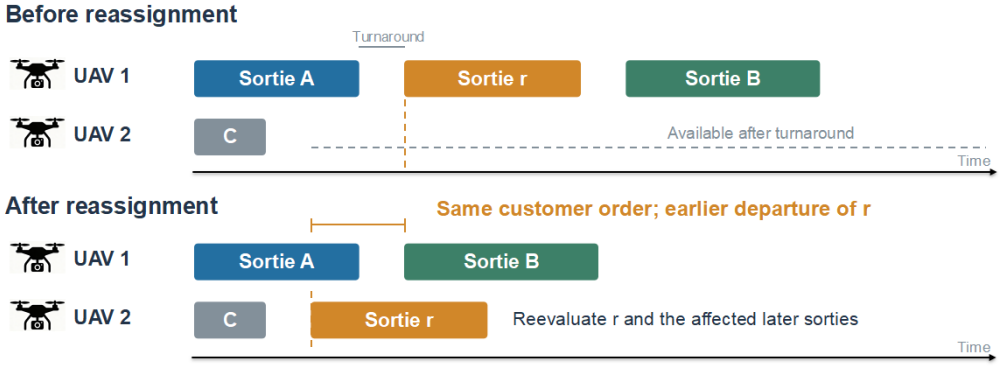


Figure 3. Reassigning sortie r advances departure and service while preserving the customer order and sortie duration. Completion times, payloads, and energy are reevaluated for r and affected later sorties. Labels A, B, and C denote other scheduled sorties.

Sorties are ranked by current demand exposure, with the starting position rotated between calls. Each call stops after the first feasible improvement or 64 complete evaluations, subject to the global budget. The best-known feasible solution is updated whenever a better effective candidate is found. The solution returned by local search replaces the worst population member if it improves upon that member.

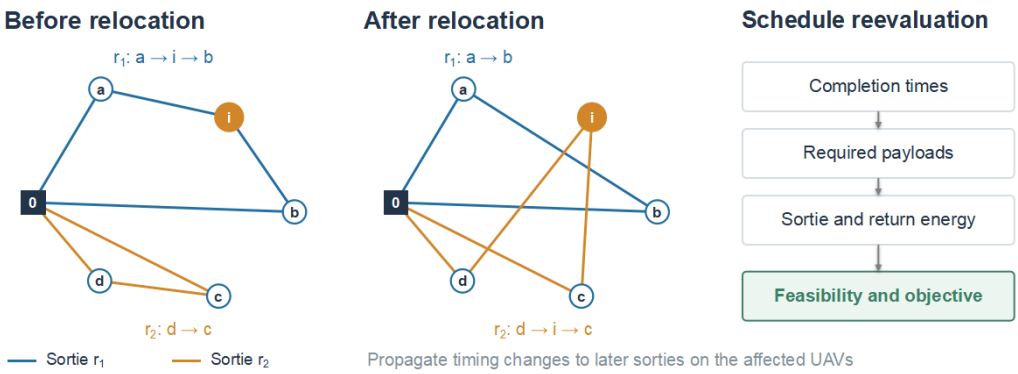


Figure 4. Relocating customer i from sortie r_1 to r_2 changes both routes and propagates timing changes to later sorties. Payloads, sortie energy, and post-service return energy are recomputed to assess the resulting schedule. Labels a, b, c, d denote other customers, and 0 denotes the depot.

3.6. Complete Procedure and Computational Cost

Algorithm 1 summarizes the procedure. Only a better feasible plan replaces the incumbent. Unserved customers are periodically offered reinsertion opportunities, and any remaining omissions are reflected in the service rate.

Algorithm 1 EDBKA-LS-I with a shared evaluation budget.

Require: Instance, energy model, population size P, limits $F_{\max}, t_{\max}$
Ensure: Best feasible plan $\mathcal{P}^*$ and objective history $\mathcal{L}$

$\mathcal{S}$: population; $\mathcal{U}(\mathcal{P})$: unserved customers in $\mathcal{P}$
F: evaluation count; t: elapsed time
Initialize $F, t, z \leftarrow 0$, $\mathcal{P}^* \leftarrow \emptyset$, and $\mathcal{L} \leftarrow \emptyset$
Initialize budgeted evaluator $\mathscr{E}$
Each $\mathscr{E}$ call checks both budgets before scoring
and updates F, t, $\mathcal{P}^*$, and $\mathcal{L}$ after scoring
$\mathcal{U} \leftarrow \textsc{SpeedPruning}(\mathcal{U})$
$\mathcal{P}_0 \leftarrow \textsc{ScreenedInsert}(\emptyset, \text{random customer order}, \mathscr{E})$
$\mathcal{S}_p \leftarrow \mathcal{P}_0, \quad p = 1, \ldots, P$
while $F < F_{\max} \ \wedge \ t < t_{\max}$ **do**
    $z \leftarrow z + 1$
    if $z \bmod 4 = 0 \ \wedge \ \mathcal{P}^* \neq \emptyset$ **then**
        $Q \leftarrow \textsc{LocalSearch}(\mathcal{P}^*, 64, \mathscr{E})$
        $\mathcal{S} \leftarrow \textsc{ReplaceWorstIfBetter}(\mathcal{S}, Q)$
    end if
    $p \leftarrow 1 + ((z-1) \bmod P)$
    $C \leftarrow \textsc{GenerateCandidates}(\mathcal{S}_p, \mathcal{P}^*, 3)$
    Evaluate C through $\mathscr{E}$
    $\mathcal{S}_p \leftarrow \textsc{BestFeasible}(\{\mathcal{S}_p\} \cup C)$
    if $z \bmod P = 0 \ \wedge \ \mathcal{U}(\mathcal{S}_p) \neq \emptyset$ **then**
        $Q \leftarrow \textsc{ScreenedInsert}(\mathcal{S}_p, \mathcal{U}(\mathcal{S}_p), \mathscr{E})$
        $\mathcal{S}_p \leftarrow \textsc{BestFeasible}(\{\mathcal{S}_p, Q\})$
    end if
end while
$\textsc{VerifyConstraints}(\mathcal{P}^*)$
return $\mathcal{P}^*, \mathcal{L}$

With precomputed pairwise distances, physical scheduling calculations for n_{served} served customers and R sorties require $O(n_{\text{served}} + R)$ arithmetic operations. Repeated customer uniqueness checks introduce additional implementation overhead, so the evaluation costs vary with the plan structure. For m insertion positions, detour computation and

sorting carry cost $O(m \log m)$. Pairwise distances require $O(n^2)$ storage, and population routes require $O(Pn)$.

Each customer insertion stage uses at most L full evaluations. With a fixed L, construction requires at most $nL + 2$ scoring calls when completed within the budget. For 100 customers, the upper bound is 1802 calls out of the 2500-call budget. A local search call uses at most 64 evaluations. An ordinary iteration uses at most three perturbation evaluations plus counted reinsertion. Evaluation count comparisons quantify calls to the shared scorer, while wall-clock comparisons account for cost variation with the plan size and feasibility rejection points.

4. Experimental Results and Analysis

4.1. Compared Methods

Table 2 defines the seven compared configurations. EDBKA-LS-I is abbreviated as LS-I in these comparisons. Base-SP retains the discrete population framework, energy-aware perturbations, and speed preprocessing. LS-NR adds local improvement, and LS-SR further adds substring recovery. LS-I adds insertion screening to LS-NR, with recovery disabled. ALNS provides a reference based on removal and repair outside the population framework. ALNS-SP samples insertion positions randomly, whereas ALNS-I uses the same insertion screening as LS-I, including in regret-2 repair.

Table 2. Compared search configurations. Speed preprocessing and the physical evaluator are common to all methods.

Method	Search Framework	Local Improvement	Insertion Selection/Recovery
Base-SP	Discrete population	No	Random/none
LS-NR	Discrete population	Yes	Random/none
LS-SR	Discrete population	Yes	Random/substring
LS-I	Discrete population	Yes	Screened/none
ALNS-SP	Adaptive removal/repair	No	Random/not applicable
ALNS-I	Adaptive removal/repair	No	Screened/not applicable
VNS	Single solution	Yes	Screened/not applicable

The ALNS implementation uses random removal, related customer removal, and removal based on exposure and energy stress. Greedy and regret-2 insertion provide two repair operators, yielding six adaptively weighted operator pairs. The removal count is $\min\{n_{\text{served}}, 8, \lceil 0.2 n_{\text{served}} \rceil\}$, where n_{served} denotes the number of currently served customers. Operator weights use a reaction factor of 0.15 and outcome-dependent rewards. Rewards are 8 for a new incumbent, 4 for current plan improvement, 1 for another accepted candidate, and 0.1 for an unaccepted outcome. Simulated annealing acceptance uses $\exp[\min\{0, (J_{\text{current}} - J_{\text{trial}})/\theta\}]$, where J_{current} and J_{trial} are the objective values of the current accepted plan and the trial plan. The annealing temperature is $\theta = 0.02 \max\{0.01, 1 - \rho\}$, where $\rho = \min\{1, \max(F/F_{\max}, t_{\text{search}}/t_{\max})\}$ is the consumed budget fraction. An inactive limit contributes zero.

An adapted variable neighborhood search comparator, abbreviated as VNS, maintained a single feasible current schedule, initialized using the same screened insertion procedure as EDBKA-LS-I. Shaking applied one to three random moves involving customer relocation, exchange, reversal, UAV reassignment, sortie ordering, splitting, or merging. Feasible shaken schedules underwent bounded local search through the three neighborhood families described in Section 3.5, returning to the first family after each improvement. Each neighborhood call used at most 64 complete evaluations, and each descent episode used at most 192. Each reinsertion call for unserved customers used at most 36 evaluations. An improving schedule replaced the current schedule and reset the shaking strength to one.

Otherwise, the shaking strength increased cyclically. Initialization, shaking, local search, and reinsertion shared the global evaluation budget.

For LS-SR, customer successor frequencies are monitored, including depot and unserved states. Successor entropy is inspected after at least 24 further evaluations. Recovery requires normalized entropy below 0.30, at least 120 evaluations since the last incumbent improvement, and a cooldown of 100 evaluations. A contiguous substring of one to four customers is removed from a selected route and repaired by regret-2 insertion. The worst population member is targeted, and the historical incumbent is preserved. A repaired member is accepted if its objective improves or if degradation is at most 0.03 and successor entropy increases. The policy combines substring removal and regret-2 repair within the neighborhood families reviewed in [30]. All recovery evaluations are charged to the common global evaluation budget. These additional recovery steps are separate from the ordinary removal and repair operations in ALNS.

4.2. Instances, Budgets, and Evaluation Protocol

Six synthetic instances use $n \in \{25, 50, 100\}$ and instance seeds 1801 and 1802. Coordinates are sampled from an 1800×1800 m square, and initial demands are drawn from $[0.035, 0.100]$ kg. Customer weights are one or three, with probability 0.2 for the latter. Three further cases use the first fifty customer coordinates and demands from C101, R101, and RC101 in the public Solomon collection. Coordinates are scaled by 20 m/unit and demands by 0.002 kg/unit, and customer weights are set to one. The original service durations and time windows are replaced with the UAV service durations and no-wait landing assumptions. These transformed layouts provide clustered, random, and mixed cases for comparison within the UAV model. Published best-known values for the original vehicle routing problem with time windows refer to a different optimization problem. Source links are provided in the Data Availability Statement.

The baseline routing configuration uses three UAVs, $H = 7200$ s, $\lambda_i = 0.5 d_i^0 / H$, landing delivery, and a payload limit of 0.50 kg. The altitude is 25 m, ground service lasts 12 s, and depot turnaround takes 60 s. The climb and descent speeds are 2.5 and 1.5 m/s, respectively. All UAVs are available at time zero, with $a_k = 0$ for every UAV. The initial energy and reserve are set to 90 Wh and 18 Wh per sortie as routing scenario parameters. The no-wait landing assumption, together with the fitted energy law, satisfies the speed dominance condition. All algorithms use the same energy model throughout the routing comparisons.

The component study compares Base-SP, ALNS-SP, LS-NR, and LS-SR over twenty seeds numbered 1001 to 1020 per instance, with 2500 complete evaluations including construction. The representation, evaluator, construction rule, and speed preprocessing are shared. Every run reaches the evaluation cap before the secondary 120 s limit. Screening is applied to LS-NR and ALNS-SP to obtain LS-I and ALNS-I. Both are evaluated on all nine cases with the same twenty seeds and evaluation cap. LS-NR serves as the common control for the component and screening comparisons, yielding 1080 distinct equal-evaluation runs in these component and screening comparisons.

An additional wall-clock comparison uses six seconds and five seeds on three prespecified cases. These are the two 100-customer synthetic cases and the RC101-derived case. The five configurations Base-SP, ALNS-SP, LS-NR, ALNS-I, and LS-I yielded 75 runs in total.

The experiments use MATLAB R2021b on Windows with an Intel Core i9-12900HX processor. In the component and screening experiments, each method was run before timing to reduce initialization effects, and the method order was randomized within each block defined by instance and seed. The timed interval covers construction and search, including screening.

The primary metric is J in Equation (12), with lower values indicating less pre-service demand exposure. Secondary metrics are the nominal predicted energy $\sum_r E_r$, the cus-

tomer service rate, the elapsed search time, and the distribution of evaluation calls across construction, local improvement, and recovery. Results are averaged over seeds within each case and then equally across cases. For a target method A and a reference method denoted by ref, the aggregate improvement is

$$G_{A:\mathrm{ref}} = 100\frac{\bar{J}_{\mathrm{ref}} - \bar{J}_A}{\bar{J}_{\mathrm{ref}}}, \qquad \bar{J}_A = \frac{1}{N_I}\sum_{\nu=1}^{N_I}\frac{1}{N_S}\sum_{s=1}^{N_S} J_{A,\nu,s}, \tag{17}$$

where N_I and N_S are the numbers of instances and optimization seeds, and $J_{A,\nu,s}$ is the final objective value of method A on instance ν with seed s. The mean $\bar{J}_{\mathrm{ref}}$ is calculated in the same way for the reference method. The reported gain compares aggregate means, with negative values indicating a higher target mean. Energy differences are calculated separately. All final-run outcomes are retained in the aggregates, including feasible schedules with unserved customers.

Paired inference uses these case-level means as the analysis units. Exact two-sided signed-rank tests use midranks for ties and omit zero differences. Holm adjustment is applied separately within the three component comparisons and the two refinement comparisons. Case-bootstrap intervals use 10,000 paired resamples under an assumption of case exchangeability. A sensitivity analysis groups synthetic cases by their shared generation seed and treats each public layout as a separate group. This yields five source groups, with a minimum attainable two-sided exact signed-rank p of 0.0625.

4.3. Validation of the Payload-Dependent Energy Model

Public DJI Matrice 100 records [9] provide the basis for whole-flight energy validation. The 205 retained records are partitioned into 122 training, 39 validation, and 44 test flights. Forty test delivery flights form the reported energy test set. Whole flights are assigned to partitions, preserving each flight's samples as one unit. For telemetry replay at speeds below the modeled interval, the power is estimated by blending the hovering power with the 4 m/s prediction. Higher telemetry speeds are clipped at the supported endpoint for replay. Routing uses the supported cruise speed set, while the hovering power is used for telemetry replay. The validation target is the whole-flight energy, and the routing model represents landing delivery.

The calibrated phase model is labeled Full. Its payload-free alternative, NoPayload, is refitted by removing payload features from all affected phase models and repeating validation-based cruise model selection. ConstantPower multiplies the flight duration by the arithmetic mean of the flight-average powers of training delivery flights. Distance is a nonnegative linear regression of the whole-flight energy on the horizontal distance, with an intercept, fitted to the same training delivery flights. Distance is obtained by integrating the measured horizontal speed.

Table 3 summarizes the energy prediction accuracy using three metrics. The mean absolute error is denoted as MAE, the mean absolute percentage error is denoted as MAPE, and bias measures the mean signed error. MAPE averages each flight's absolute prediction error divided by its observed energy and expresses the result as a percentage.

Table 3. Energy prediction on forty held-out delivery flights. MAE denotes mean absolute error, and MAPE denotes mean absolute percentage error. Bias is mean prediction minus observation.

Model	MAE/Wh	MAPE/%	Bias/Wh
Full	0.993	4.90	−0.284
NoPayload	1.827	8.43	−0.348
ConstantPower	2.141	10.20	0.186
Distance	4.295	21.12	−0.373

The full model achieved an MAE of 0.993 Wh, compared with 1.827 Wh for the payload-free alternative, a reduction of 45.68%. Figure 5 compares the full-model predictions with the observed flight energy and shows the residuals for the full and payload-free models. The validation-derived margin was approximately 9.39%, with 95% empirical upper coverage on the forty held-out delivery flights.

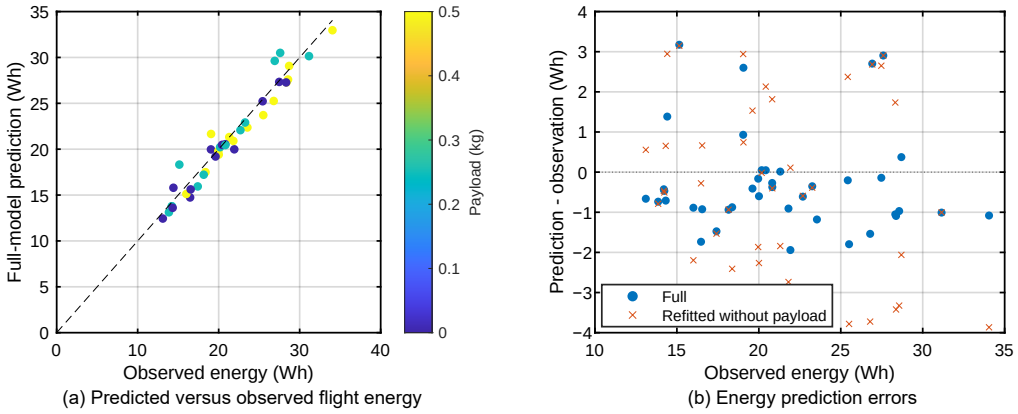


Figure 5. Energy validation on forty held-out delivery flights. Panel (**a**) compares full-model predictions with observations, colored by payload, with the dashed line marking exact agreement. Panel (**b**) compares the prediction errors, showing a lower mean absolute error for the full model than for the refitted payload-free model.

Phase energies were calculated by integrating the measured and predicted power over the recorded phase durations of each held-out flight. Trapezoidal sample weights apportioned an interval spanning two phase labels between its endpoints, so the phase contributions sum to the whole-flight energy. The test set contained 13 flights with a 0 kg payload, 13 with 0.25 kg, and 14 with 0.50 kg. The whole-flight MAE was 0.801, 1.171, and 1.005 Wh in these groups, respectively, with corresponding MAPE values of 4.47%, 5.86%, and 4.39%. Table 4 reports the phase energy prediction errors across the three payload groups.

Table 4. Phase energy prediction errors on forty held-out delivery flights. Phase MAE is calculated from the integrated energy of each phase within a flight. Bias is the mean predicted minus observed energy across all forty flights.

	MAE/Wh				Bias/Wh
Phase	**0 kg**	**0.25 kg**	**0.50 kg**	**All**	**All**
Ground/idle	0.589	0.627	0.748	0.657	−0.657
Ascent	0.121	0.089	0.121	0.111	0.001
Cruise	0.274	0.248	0.275	0.266	−0.053
Descent	0.102	0.095	0.115	0.104	0.015
Hover	0.327	0.977	0.326	0.538	0.412

For each phase within each held-out flight, coverage required the observed energy to be no greater than $(1+\mu)$ times its prediction. With the 9.39% margin, coverage was 40/40 for ascent, cruise, and descent individually, 33/40 for hover, and 0/40 for ground/idle. The ground/idle underprediction and hover overprediction indicate that the phase composition can affect the total mission error. Ground/idle samples were identified by an altitude threshold and did not isolate customer-side ground handling.

4.4. Local Search and Recovery Contributions

Table 5 reports the means and within-instance seed standard deviations for the component comparison. The equal-case mean objective was 0.09551 for LS-NR, 0.10396 for Base-SP, and 0.10676 for ALNS-SP. LS-NR reduced the mean objective by 8.13% relative to Base-SP and 10.54% relative to ALNS-SP. Both comparisons shared the same speed preprocessing, construction, and evaluation cap. The Base-SP comparison quantified the budgeted contribution of local improvement under the same recovery setting. Table 6 further shows that LS-NR achieved lower case-level means than Base-SP and ALNS-SP on all nine instances, with positive rank-biserial effects in all three paired comparisons. LS-SR attained mean $J = 0.09607$, compared with 0.09551 for LS-NR. Recovery averaged 2.33 events per run and consumed 6.95% of evaluations. The lower mean objective and the absence of recovery expenditure supported the selection of LS-NR as the basis of LS-I. This comparison evaluates the complete recovery policy at the component level.

Table 5. Component comparison: mean $J\pm$ seed standard deviation, twenty runs per cell and 2500 full evaluations per run. Sn-1/2 denote n-customer synthetic cases with instance seeds 1801/1802.

Case	Base-SP	ALNS-SP	LS-NR	LS-SR
S25-1	0.0390 ± 0.0027	0.0367 ± 0.0023	0.0353 ± 0.0020	0.0357 ± 0.0022
S25-2	0.0371 ± 0.0020	0.0362 ± 0.0020	0.0349 ± 0.0020	0.0355 ± 0.0020
S50-1	0.0899 ± 0.0056	0.0925 ± 0.0093	0.0777 ± 0.0039	0.0785 ± 0.0042
S50-2	0.0888 ± 0.0057	0.0918 ± 0.0085	0.0807 ± 0.0032	0.0812 ± 0.0034
S100-1	0.2237 ± 0.0166	0.2308 ± 0.0178	0.2092 ± 0.0114	0.2091 ± 0.0113
S100-2	0.2237 ± 0.0110	0.2309 ± 0.0133	0.2101 ± 0.0081	0.2101 ± 0.0081
C101	0.0628 ± 0.0026	0.0658 ± 0.0044	0.0599 ± 0.0028	0.0603 ± 0.0027
R101	0.0795 ± 0.0041	0.0821 ± 0.0047	0.0699 ± 0.0035	0.0720 ± 0.0037
RC101	0.0909 ± 0.0061	0.0940 ± 0.0066	0.0820 ± 0.0051	0.0824 ± 0.0048

Table 6. Paired component comparisons, with LS-NR as the target. Positive gain indicates lower target J. r_{rb} is the signed rank-biserial effect. W/T/L are case-level wins/ties/losses. p_{Holm} is the Holm-adjusted two-sided signed-rank p-value.

Reference	Gain/%	W/T/L	r_{rb}	p_{Holm}
Base-SP	8.13	9/0/0	1.000	0.0117
ALNS-SP	10.54	9/0/0	1.000	0.0117
LS-SR	0.59	7/1/1	0.944	0.0156

4.5. Insertion Screening and Comparative Performance

A separate 27-run pilot comparison yielded mean $J = 0.11657$ without screening and 0.09860 with screening. Screening parameters were fixed before the main comparisons. Table 7 compares LS-I with unscreened LS-NR and matched ALNS-I under the same evaluation cap. LS-I reduced the mean demand exposure by 7.44% relative to unscreened LS-NR. Screening also reduced the ALNS mean J from 0.10676 to 0.09582. These comparisons quantify the contribution of insertion screening within both search frameworks.

Table 7. Refinement comparison over nine cases with twenty runs per method and case. Runtime is measured at the fixed evaluation cap.

Method	Mean J	Energy/Wh	Service/%	Time/s
LS-NR	0.09551	689.82	100.00	2.100
ALNS-I	0.09582	631.47	99.97	1.853
LS-I	0.08840	629.07	100.00	1.917

With insertion screening shared by both methods, LS-I attained mean $J = 0.08840$, reducing demand exposure by 7.74% relative to ALNS-I under the same evaluation cap. LS-I and ALNS-I differed by 0.38% in average energy. The improvement therefore primarily concerned service timing at similar energy consumption. At the fixed evaluation cap, LS-I took 1.917 s versus 1.853 s for ALNS-I, approximately 3.44% longer. Table 8 reports the paired gains, intervals, and multiplicity-adjusted tests under the protocol in Section 4.2.

Table 8. Paired refinement comparisons with LS-I as the target. Confidence intervals are case-bootstrap intervals for the gain of means. W/T/L are case-level wins/ties/losses. p_{Holm} is the two-sided signed-rank p-value adjusted across these two comparisons using the Holm procedure.

Reference	Gain/%	95% Interval	W/T/L	p_{Holm}
LS-NR	7.44	[2.42, 10.25]	7/0/2	0.0391
ALNS-I	7.74	[5.50, 9.65]	7/0/2	0.0391

Figure 6 shows how the performance varies across instances. Relative to ALNS-I, LS-I reduced the mean J by 14.14% and 10.10% on the two 50-customer synthetic cases. The reductions were 7.41% and 7.42% on the two 100-customer cases. The C101-, R101-, and RC101-derived layouts showed reductions of 3.92%, 9.16%, and 9.29%, respectively. On the two 25-customer cases, LS-I produced objective values that were 1.62% and 1.21% higher than those of ALNS-I. Thus, the gains were concentrated in the larger synthetic instances and the public layout-derived cases.

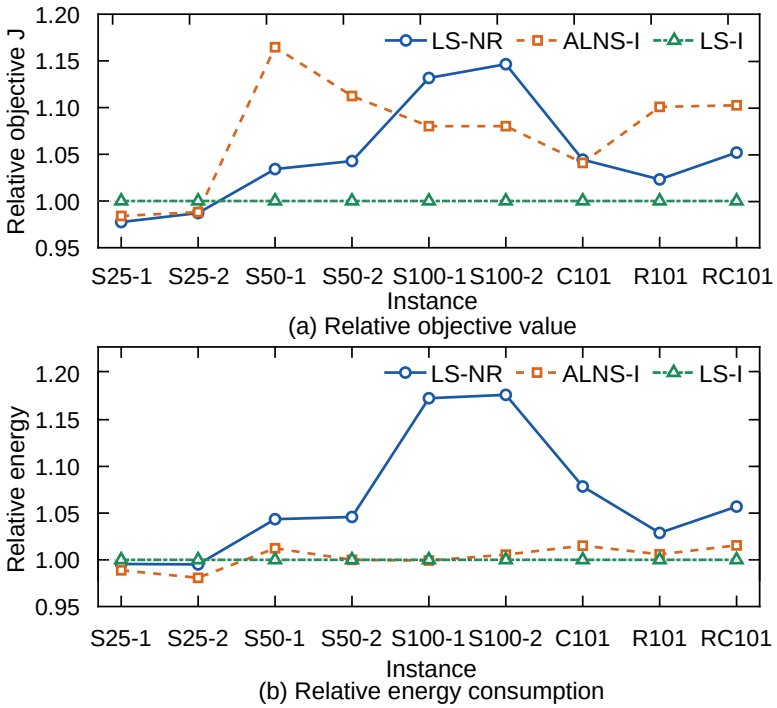


Figure 6. Mean demand exposure in panel (**a**) and energy consumption in panel (**b**) over twenty runs with 2500 evaluations per run, normalized by the corresponding LS-I mean for each instance. Sn-1 and Sn-2 ($n = 25, 50, 100$) denote synthetic instances with n customers and instance seeds 1801 and 1802, respectively. C101, R101, and RC101 denote 50-customer instances derived from the corresponding Solomon layouts. Ratios above one favor LS-I. Lower exposure on larger instances is achieved at energy consumption comparable to ALNS-I.

Figure 7 shows one LS-I schedule for the S50-1 instance. With optimization seed 1001 and a budget of 2500 full evaluations, all 50 customers were served by three UAVs in 13 sorties. The plan attained $J = 0.08430$ and consumed 668.55 Wh across the sorties. The

route and time panels show the spatial assignment and successive use of each UAV. The margin-adjusted energy remained within the 72 Wh usable battery limit for every sortie, and the initial payloads remained below 0.50 kg.

Geometric screening was compared with the ranking in Equation (16) on S25-1, S100-1, and RC101. The tested limits were $L = 6, 12, 18, 24, 30$. Each configuration used twenty seeds and 2500 complete evaluations, including construction. All other search settings were unchanged. Means were calculated over seeds within each instance and then equally across the three instances. The runtime includes construction, screening, and search. Full-service counts are reported out of sixty runs per configuration. Construction and local search shares are percentages of the complete evaluation budget.

Table 9 shows that geometric screening achieved a lower aggregate mean objective at every tested limit. At $L = 18$, its mean J was 0.09963, compared with 0.10838 for exposure-based screening, a reduction of 8.07%. Geometric screening performed better on S100-1 and RC101 at all five limits. Exposure-based screening reduced the mean J by 1.26% on S25-1 at $L = 18$.

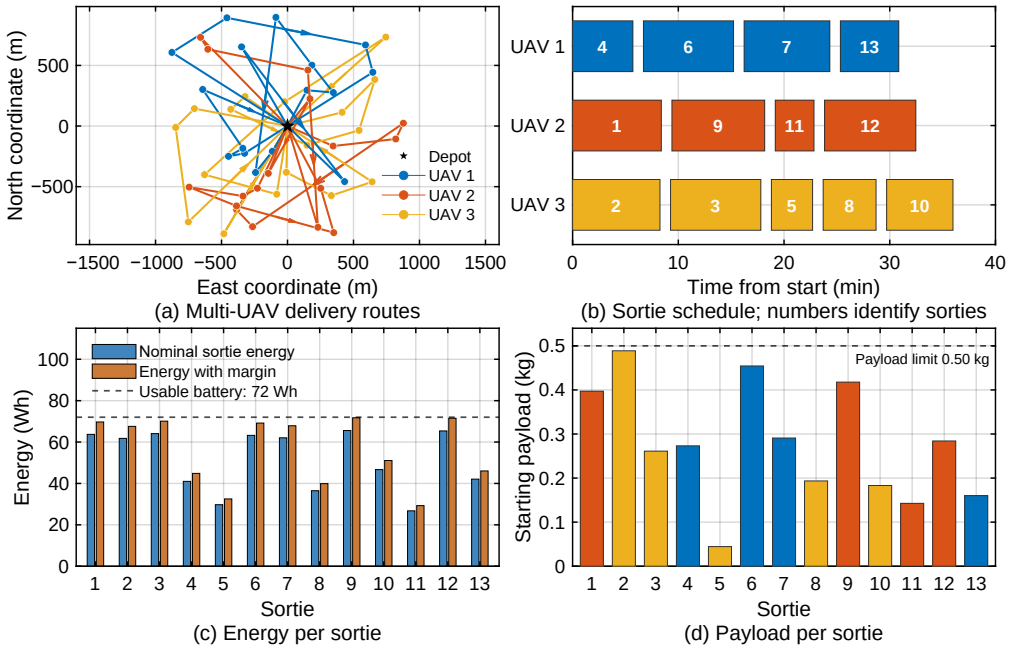


Figure 7. EDBKA-LS-I schedule for S50-1 with optimization seed 1001. Panels (**a**,**b**) show routes and timing for three UAVs serving fifty customers in thirteen sorties. Panels (**c**,**d**) show the margin-adjusted energy and starting payload within the 72 Wh and 0.50 kg limits. Matching indices connect sorties across panels (**b**–**d**).

Table 9. Effects of the screening rule and insertion limit on solution quality and evaluation allocation.

L	Rule	Mean J	Full Service	Construction /%	Local Search /%	Time /s
6	Geometric	0.09100	60/60	13.92	61.89	2.563
6	Exposure	0.10259	60/60	13.92	55.98	3.006
12	Geometric	0.09563	60/60	27.00	53.80	2.358
12	Exposure	0.10492	60/60	27.00	50.31	2.465
18	Geometric	0.09963	60/60	38.72	45.37	2.157
18	Exposure	0.10838	60/60	38.72	43.46	2.050
24	Geometric	0.10484	60/60	49.16	37.54	1.978
24	Exposure	0.11304	60/60	49.16	37.35	1.872
30	Geometric	0.12330	40/60	55.70	33.11	1.781
30	Exposure	0.12828	40/60	55.70	32.98	1.700

For both rules, the lowest aggregate mean objective occurred at $L = 6$. With geometric screening, reducing L from 18 to 6 lowered the mean J by 8.66%, from 0.09963 to 0.09100. The construction share decreased from 38.72% to 13.92%, while the local search share increased from 45.37% to 61.89%. The mean runtime increased from 2.157 to 2.563 s. Both rules achieved complete service in every run at $L = 6, 12, 18, 24$. At $L = 30$, construction consumed all 2500 evaluations on S100-1, leaving no budget for subsequent search. Full service was achieved in forty of sixty runs for each rule. All 600 schedules satisfied the model constraints. The results favor a smaller insertion limit for final solution quality under this evaluation budget.

VNS was evaluated on the same nine instances using twenty seeds and 2500 complete evaluations per run. The energy model, constraints, initialization, speed preprocessing, and insertion limit $L = 18$ were shared with EDBKA-LS-I. Table 10 reports the means and standard deviations over seeds. Aggregate means give equal weight to each instance. A positive gain denotes the percentage reduction in the mean J for EDBKA-LS-I relative to VNS. EDBKA-LS-I attained a lower mean J on seven of nine instances, with reductions of 1.18% to 5.88%. Its aggregate mean was 0.08840, compared with 0.08923 for VNS, a reduction of 0.93%. On S100-1 and S100-2, its mean J was 3.65% and 3.58% higher than that of VNS, respectively. The two-sided exact signed-rank test on the nine instance means gave $p = 0.5703$, with a rank-biserial correlation of 0.2444. Each method achieved complete service in all 180 runs, and all final schedules passed independent feasibility checks.

Table 10. Comparison with VNS under 2500 complete evaluations per run. EDBKA-LS-I obtains lower mean demand exposure on seven instances, while VNS performs better on the two 100-customer instances.

Instance	EDBKA-LS-I J	VNS J	Gain/%
S25-1	0.03608 ± 0.00254	0.03815 ± 0.00205	5.42
S25-2	0.03535 ± 0.00191	0.03654 ± 0.00191	3.26
S50-1	0.07512 ± 0.00489	0.07950 ± 0.00478	5.52
S50-2	0.07737 ± 0.00308	0.08110 ± 0.00390	4.59
S100-1	0.18485 ± 0.01129	0.17834 ± 0.00995	−3.65
S100-2	0.18328 ± 0.01041	0.17694 ± 0.00700	−3.58
C101	0.05733 ± 0.00293	0.05801 ± 0.00333	1.18
R101	0.06829 ± 0.00414	0.07171 ± 0.00336	4.77
RC101	0.07795 ± 0.00636	0.08282 ± 0.00688	5.88
Aggregate	0.08840	0.08923	0.93

4.6. Fixed-Time Comparison and Computational Expenditure

Table 11 reports the six-second comparison over three prespecified cases, with five seeds per method. Under the six-second budget, LS-I achieved a mean objective that was 7.93% lower than that of ALNS-I. Including construction and screening in the timed search provided a complementary comparison to the fixed-evaluation study.

Table 11. Six-second wall-clock comparison: equal weighting of three case means, each based on five runs.

Method	Mean J	Energy/Wh	Service/%
LS-NR	0.14733	1042.18	100.00
ALNS-I	0.14673	949.79	100.00
LS-I	0.13510	948.41	100.00

Figure 8 shows that LS-I continues to improve after construction. The smaller changes toward the end of the search reflect diminishing gains as the remaining evaluation budget is used.

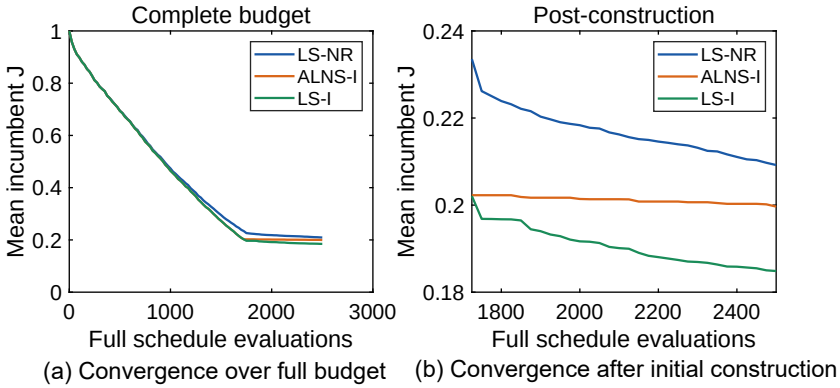


Figure 8. Mean best-so-far demand exposure over twenty runs per method on S100-1 with instance seed 1801. Panel (**a**) covers all 2500 evaluations, including construction. Panel (**b**) retains the cumulative evaluation axis after all runs complete construction, revealing continued improvement by LS-I.

In the unscreened component study, initial construction used 14.72%, 32.72%, and 68.72% of the LS-NR evaluation budget on instances with 25, 50, and 100 customers, respectively, leaving progressively fewer evaluations for refinement. For Figure 9, each phase count was divided by the 2500-evaluation budget. The resulting shares were averaged over twenty runs per instance and then equally across the nine instances. These comprise two cases with 25 customers, five with 50, and two with 100. Weighting the three size-group construction means by these instance counts gives the overall share of 36.72% shown for LS-NR.

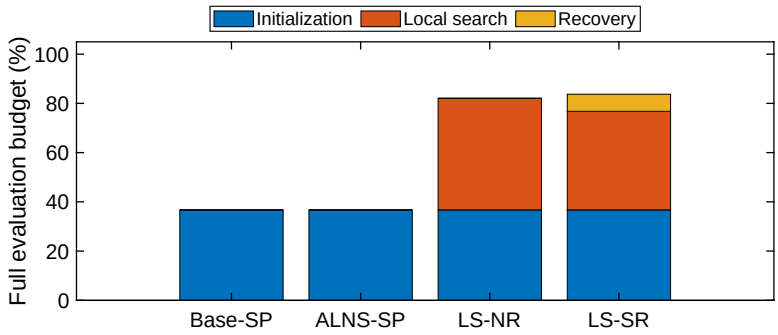


Figure 9. Allocation of the 2500-evaluation budget to construction, local improvement, and recovery. Shares are averaged over twenty runs within each instance and then equally across nine instances. The unstacked remainder covers ordinary search, including ALNS removal and repair. Recovery consumes 6.95% of the LS-SR budget.

4.7. Energy Margin Sensitivity

Energy margin sensitivity was evaluated on S25-1, S100-1, and RC101 using twenty matched seeds and 2500 evaluations per run. The margins were 0%, 5%, the calibrated 9.39%, 15%, and 20%. The available energy $B - B_{\text{res}}$ was set to 72 or 54 Wh, with the 18 Wh reserve retained. Within each energy setting, the sixty plans obtained at the calibrated margin were held fixed for feasibility reassessment, and EDBKA-LS-I was rerun at every margin. Table 12 shows that only two reference plans remained feasible at 15%, and none remained feasible at 20%, in either energy setting. All 600 reoptimized schedules were feasible and served every customer. Compared with zero margin, the 20% margin increased the mean demand exposure by 6.78% and 8.54% at 72 and 54 Wh, respectively. The corresponding mean sortie counts increased from 12.07 to 13.88 and from 15.50 to 19.25. The larger allowance was therefore accommodated by additional sorties, with an associated increase in demand exposure.

Table 12. Effect of the energy margin on fixed-plan feasibility and reoptimized schedules. Fixed feasible gives the number of feasible reference plans out of sixty at each energy setting. Mean values give equal weight to the three instance means, each based on twenty runs.

$B-B_{\text{res}}$ Wh	Margin %	Fixed Feasible	Mean J	Mean Sorties
72	0	60/60	0.09866	12.07
72	5	60/60	0.09984	12.25
72	9.39	60/60	0.09963	12.93
72	15	2/60	0.10139	13.23
72	20	0/60	0.10535	13.88
54	0	60/60	0.10910	15.50
54	5	60/60	0.11183	16.28
54	9.39	60/60	0.11308	17.18
54	15	2/60	0.11653	18.23
54	20	0/60	0.11842	19.25

4.8. Performance Across Operating Conditions

Operating condition sensitivity was evaluated on S25-1, S100-1, and RC101 using twenty matched seeds and 2500 evaluations per method and setting. The demand rates were set to $\lambda_i = \alpha d_i^0/H$, where α is the ratio of demand accumulated over the horizon to the initial demand. One factor was varied at a time across $\alpha \in \{0, 0.5, 1\}$, $K \in \{2, 3, 4\}$, $Q_{\max} \in \{0.25, 0.375, 0.50\}$ kg, and $B - B_{\text{res}} \in \{54, 72, 90\}$ Wh. The baseline used $\alpha = 0.5$, $K = 3$, $Q_{\max} = 0.50$ kg, and $B - B_{\text{res}} = 72$ Wh. The calibrated margin of 9.39%, the 18 Wh reserve, the planning horizon of $H = 7200$ s, and the algorithm parameters were retained. Table 13 gives equal weight to the three instance means. The gain is the percentage reduction in this mean objective relative to ALNS-I. Full-service counts are reported for LS-I and ALNS-I, respectively, out of sixty runs each. W/T/L counts instances with lower, equal, or higher mean J for LS-I.

Table 13 shows that LS-I achieved a lower aggregate mean objective than ALNS-I in all nine configurations. The relative advantage weakened as the demand accumulation rate increased, with objective reductions of 7.12%, 6.92%, and 5.09% at $\alpha = 0$, 0.5, and 1, respectively. Across the tested fleet sizes, the largest reduction was obtained with three UAVs at 6.92%, compared with 5.25% for two UAVs and 6.19% for four UAVs. The relative advantage also varied nonmonotonically with the payload capacity. It was largest at 0.50 kg, with a reduction of 6.92%, followed by 5.81% at 0.25 kg and 4.68% at 0.375 kg. For the available energy, the reduction decreased from 10.22% at 54 Wh to 6.92% at 72 Wh and 4.46% at 90 Wh. Therefore, LS-I exhibits a greater relative advantage under stricter energy constraints.

Table 13. Sensitivity of LS-I performance relative to ALNS-I across operating conditions.

Setting	LS-I J	ALNS-I J	Gain/%	Full Service	W/T/L
Baseline	0.09963	0.10703	6.92	60/59	2/0/1
$\alpha = 0$	0.11577	0.12465	7.12	60/60	2/0/1
$\alpha = 1$	0.09057	0.09543	5.09	60/60	2/0/1
$K = 2$	0.17005	0.17947	5.25	60/59	3/0/0
$K = 4$	0.07096	0.07564	6.19	60/60	2/0/1
$Q_{\max} = 0.25$ kg	0.13339	0.14161	5.81	60/60	3/0/0
$Q_{\max} = 0.375$ kg	0.10969	0.11508	4.68	60/59	2/0/1
$B - B_{\text{res}} = 54$ Wh	0.11308	0.12595	10.22	60/60	3/0/0
$B - B_{\text{res}} = 90$ Wh	0.09556	0.10002	4.46	60/60	2/0/1

At the instance level, lower mean objectives were obtained by LS-I on S100-1 and RC101 in every configuration. On S25-1, LS-I performed better in the three configurations

with two UAVs, a 0.25 kg payload limit, or 54 Wh of available energy. In the remaining six configurations, its mean objective was 0.99% to 3.37% higher than that of ALNS-I, with the largest difference occurring at 90 Wh. These results locate the observed reversals within the smallest instance while confirming the aggregate advantage across all four factors. All 1080 schedules passed independent feasibility checks. Complete customer service was achieved in all 540 EDBKA-LS-I runs and 537 ALNS-I runs. At $H = 7200$ s, EDBKA-LS-I served every customer in all sixty runs with two UAVs and 72 Wh of available energy per sortie and in all sixty runs with three UAVs and 54 Wh. The remaining parameters were held at their baseline values in these two configurations. Thus, complete service was maintained under both the reduced fleet size and the tighter energy setting tested.

The sixty baseline EDBKA-LS-I schedules for S25-1, S100-1, and RC101 were reassessed under common demand rate deviations of −20%, −10%, 0%, 10%, and 20%. Each instance contributed twenty schedules. Parcel masses were recalculated before departure, while the UAV assignments, sortie order, routes, speeds, and completion times were retained. The payload limit, available energy, and energy margin remained at 0.50 kg, 72 Wh, and 9.39%, respectively. For schedules satisfying the payload limit, battery feasibility was checked against both sortie energy and post-service return energy constraints. As shown in Table 14, all sixty schedules remained feasible at the three nonpositive deviations. At +10% and +20%, respectively, 12 and 18 schedules exceeded the payload capacity. Among the remaining 48 and 42 schedules, three and two also violated the energy constraints, reducing joint feasibility to 45/60 and 40/60. Thus, demand rate underestimation primarily affected payload feasibility, while the battery limits excluded additional schedules.

Table 14. Payload and battery feasibility of sixty nominal schedules under demand rate errors.

Rate Deviation	Payload-Feasible Schedules	Jointly Feasible Schedules
−20%	60/60	60/60
−10%	60/60	60/60
0%	60/60	60/60
+10%	48/60	45/60
+20%	42/60	40/60

The performance with variable cruise speeds was evaluated in a controlled energy scenario with $c = 400$ W and all other energy coefficients unchanged. The left-hand side of Equation 14 was 2.3716, so all five speeds in $\mathcal{V} = \{4, 6, 8, 10, 12\}$ m/s were retained. EDBKA-LS-I and ALNS-I searched this set, while an EDBKA-LS-I control used the same operators with the singleton speed set {12} m/s. Each configuration was evaluated on S25-1, S100-1, and RC101 with twenty matched seeds and 2500 complete evaluations per run. The available energy was 72 Wh, and the calibrated 9.39% margin was retained as a fixed planning parameter. Aggregate means assign equal weight to the three instance means. Speed proportions were calculated within each run and then averaged across seeds and instances. Energy denotes the total nominal energy across all sorties in a schedule.

Table 15 shows that variable-speed EDBKA-LS-I reduced the aggregate mean demand exposure by 14.03% relative to variable-speed ALNS-I. The reductions were 13.20%, 10.89%, and 22.65% on S25-1, S100-1, and RC101, respectively. All 180 schedules passed independent feasibility checks. Complete service was achieved in all 60 runs of each EDBKA-LS-I configuration and 57 ALNS-I runs. The mean proportion of EDBKA-LS-I sorties using speeds below 12 m/s was 55.12%. The fixed-speed control achieved a lower aggregate mean objective of 0.11046, compared with 0.13205 for variable-speed EDBKA-LS-I. The latter was 19.54% higher. Thus, the advantage over ALNS-I persisted when speed remained a

decision variable, while fixing the speed at 12 m/s produced better solutions under the same evaluation budget.

Table 15. Scheduling performance when the speed dominance condition fails.

Instance	Method	Mean $J \pm$ SD	Energy/Wh	Full Service	Below 12 m/s %
S25-1	LS-I	0.04114 ± 0.00290	478.90	20/20	41.99
S25-1	ALNS-I	0.04740 ± 0.00475	496.23	20/20	54.97
S25-1	LS-I fixed 12	0.03938 ± 0.00276	480.38	20/20	0.00
S100-1	LS-I	0.26721 ± 0.02491	2052.65	20/20	69.32
S100-1	ALNS-I	0.29988 ± 0.03001	2073.40	17/20	72.47
S100-1	LS-I fixed 12	0.21157 ± 0.00919	1918.22	20/20	0.00
RC101	LS-I	0.08779 ± 0.00538	813.20	20/20	54.06
RC101	ALNS-I	0.11350 ± 0.01367	880.97	20/20	70.06
RC101	LS-I fixed 12	0.08045 ± 0.00414	811.89	20/20	0.00
Aggregate	LS-I	0.13205	1114.92	60/60	55.12
Aggregate	ALNS-I	0.15359	1150.20	57/60	65.83
Aggregate	LS-I fixed 12	0.11046	1070.16	60/60	0.00

4.9. Discussion

Service timing affects mission feasibility because accumulated requests determine the mass to be loaded. The proposed formulation connects delivery urgency represented through deprivation costs [6] with the payload-dependent routing energy [10] through completion-dependent parcel masses. A schedule change is propagated through onboard payloads, energy requirements, and subsequent sorties. Earlier service can therefore reduce both demand exposure and the energy required to carry accumulated supplies. The energy model comparison supports the inclusion of payload effects, with the prediction MAE reduced from 1.827 to 0.993 Wh relative to the refitted payload-free model.

The speed dominance result identifies conditions under which the coupled scheduling problem can be simplified while preserving the best attainable objective value. Dominated-speed preprocessing [15] is formulated here for completion-dependent payloads and successive sorties. Under the conditions stated in Section 3.2, increasing the cruise speed cannot delay service or increase the payload and energy requirements. The reduction of five candidate speeds to 12 m/s illustrates this result for the calibrated model. This analytical simplification is shared by all compared methods.

Compared with the adaptive removal and repair used by ALNS-I, the bounded local search in EDBKA-LS-I directly refines the UAV assignments, sortie order, and customer sequences. The separate component comparisons in Tables 6 and 8 support allocating evaluations to local improvement and screened insertion candidates. With the same insertion screening, EDBKA-LS-I achieved mean demand exposure that was 7.74% below that of ALNS-I under the evaluation budget and 7.93% below it under the time budget. The mean energy consumption was similar in the fixed-evaluation comparison, indicating improved service timing at a comparable energy cost. The relative advantage increased when the available energy was reduced in the operating condition experiments. On S25-1, however, ALNS-I obtained lower mean objectives in six configurations.

EDBKA-LS-I obtained lower mean demand exposure than VNS on the tested instances with 25 or 50 customers. On both instances with 100 customers, VNS produced lower mean values, reversing the ranking observed for the smaller cases.

Applying the calibrated phase power relations to successive delivery legs assumes that they remain valid after each payload reduction. Measurements from complete missions with repeated stops and successive unloadings are needed to assess the accumulated energy error

and calibrate a mission-level margin. Customer-side ground operations should be recorded separately. Demand arrival records would also allow the linear accumulation assumption to be tested and provide input for rolling-horizon scheduling with repeated deliveries.

5. Conclusions

A scheduling method was developed for multi-UAV emergency delivery under demand accumulation and payload-dependent energy. Completion-dependent parcel masses connect service timing with payload and energy feasibility across successive sorties. Including payload effects reduced the energy prediction MAE from 1.827 to 0.993 Wh relative to the refitted payload-free model. For no-wait landing delivery, a sufficient condition was established for removing dominated speeds while preserving the best attainable objective value. This condition reduced the calibrated candidate set from five speeds to 12 m/s. EDBKA-LS-I combines bounded the local search of UAV assignments, the sortie order, and customer sequences with insertion screening. Component comparisons supported the use of both mechanisms to reduce demand exposure. Relative to ALNS-I, the mean cumulative unmet demand exposure was reduced by 7.74% under matched evaluation budgets and 7.93% under matched time budgets. Compared with VNS, EDBKA-LS-I achieved lower mean demand exposure on seven of the nine test instances, whereas VNS obtained lower values on the two 100-customer instances. These results show that the coordinated refinement of delivery schedules can improve service timing while satisfying the modeled payload and energy constraints. Future work will use multi-stop flight measurements to refine the energy margins and observed demand arrivals to assess the linear accumulation assumption and support rolling-horizon scheduling.

Author Contributions: Conceptualization, H.Y., G.C. and J.G.; methodology, H.Y. and Y.L.; software, H.Y.; validation, H.Y. and Y.L.; formal analysis, H.Y.; investigation, H.Y.; resources, G.C. and J.G.; data curation, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, G.C., J.G. and Y.L.; visualization, H.Y.; supervision, G.C. and J.G.; project administration, G.C., J.G. and Y.L.; funding acquisition, G.C. and J.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Guangxi Natural Science Foundation, under grant 2026GXNSFBA00640050; in part by the Guangxi Key Laboratory of Special Engineering Equipment and Control, Guilin University of Aerospace Technology, under Grant SEEC26KF01; in part by the Key Laboratory of Intelligence Integrated Automation in Guangxi Universities, under Grant ZDH26001; and in part by the Guangxi Science and Technology Program, under Grant Gui Ke AA19254021.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The public flight data analyzed in this study are available at https://doi.org/10.1184/R1/12683453 (accessed on 12 September 2026). The related flight energy analysis is reported in [9]. The original data descriptor was published in 2021 and is available at https://doi.org/10.1038/s41597-021-00930-x (accessed on 20 September 2026). Public routing layouts are derived from Solomon's 1987 benchmark collection, available at https://www.sintef.no/projectweb/top/vrptw/100-customers/ (accessed on 13 September 2026). The original benchmark description is available at https://doi.org/10.1287/opre.35.2.254 (accessed on 12 September 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tureci-Isik, H.; Çelik, M.; Sanci, E. The stochastic location-routing problem with parallel truck–drone operations for humanitarian aid delivery. *Eur. J. Oper. Res.* **2026**, *331*, 242–259. [CrossRef]

2. Zhao, L.; Huang, Q.; Wu, C. Distributionally robust scheduling optimization for pharmaceutical delivery using coordinated mother-end drones under post-earthquake road disruptions. *Transp. Res. Part E Logist. Transp. Rev.* **2026**, *205*, 104481. [CrossRef]
3. Attenni, G.; Arrigoni, V.; Bartolini, N.; Maselli, G. Drone-based delivery systems: A survey on route planning. *IEEE Access* **2023**, *11*, 123476–123504. [CrossRef]
4. Duan, J.; Luo, H.; Wang, G. Approaches to the truck-drone routing problem: A systematic review. *Swarm Evol. Comput.* **2025**, *92*, 101825. [CrossRef]
5. Chen, X.Y.; Shi, L.; Li, J.Y. An adaptive sliding-window ant colony optimization approach for multi-UAV delivery routing with safety and lateness objectives. *Algorithms* **2026**, *19*, 570. [CrossRef]
6. Khalaj Rahimi, S.; Rahmani, D. An improved ALNS for hybrid pickup and drones delivery system in disaster by penalizing deprivation time. *Comput. Oper. Res.* **2024**, *170*, 106722. [CrossRef]
7. Mahmoudi, B.; Eshghi, K. The multi-visit split delivery VRP with drones considering en-route launches and rendezvouses: Application to post-disaster relief operations. *Comput. Ind. Eng.* **2025**, *206*, 111232. [CrossRef]
8. Calamoneri, T.; Corò, F.; Mancini, S. Management of a post-disaster emergency scenario through unmanned aerial vehicles: Multi-depot multi-trip vehicle routing with total completion time minimization. *Expert Syst. Appl.* **2024**, *251*, 123766. [CrossRef]
9. Rodrigues, T.A.; Patrikar, J.; Oliveira, N.L.; Matthews, H.S.; Scherer, S.; Samaras, C. Drone flight data reveal energy and greenhouse gas emissions savings for very small package delivery. *Patterns* **2022**, *3*, 100569. [CrossRef] [PubMed]
10. Wu, K.; Lu, S.; Chen, H.; Feng, M.; Lu, Z. An energy-efficient logistic drone routing method considering dynamic drone speed and payload. *Sustainability* **2024**, *16*, 4995. [CrossRef]
11. Guo, J.; Song, R.; He, S. Vehicle and onboard UAV collaborative delivery route planning: Considering energy function with wind and payload. *J. Syst. Eng. Electron.* **2025**, *36*, 194–208. [CrossRef]
12. Shangguan, C.; Zhang, C.; Cao, S. UAV–rider collaborative dispatching under stochastic wind conditions considering nonlinear energy dynamics. *Drones* **2026**, *10*, 174. [CrossRef]
13. Cicek, D.; Simsek, M.; Kantarci, B. Machine learning-driven truck–drone collaborative delivery for time- and energy-efficient last-mile deliveries. *Electronics* **2025**, *14*, 2026. [CrossRef]
14. Ren, G.; Shahzaad, B.; Alkouz, B.; Lakhdari, A.; Bouguettaya, A. Energy-predictive planning for optimizing drone service delivery. *Expert Syst. Appl.* **2026**, *297*, 129251. [CrossRef]
15. Tamke, F.; Buscher, U. The vehicle routing problem with drones and drone speed selection. *Comput. Oper. Res.* **2023**, *152*, 106112. [CrossRef]
16. Büyüksan, A.T.; Demir, K.U.; Utku, D.H.; Özgün, K. Energy consumption optimization of multi-trip UAV routing using surrogate modeling with heuristic and metaheuristic algorithms. *Drones* **2026**, *10*, 430. [CrossRef]
17. Ji, Y.; Liu, L.; Wang, Y.; Wang, X.V.; Wang, L. Time-efficient routing and speed control for truck drone delivery under non-linear energy constraints. *Drones* **2026**, *10*, 466. [CrossRef]
18. Kong, J.; Xie, M.; Wang, H. Integrating autonomous vehicles and drones for last-mile delivery: A routing problem with two types of drones and multiple visits. *Drones* **2025**, *9*, 280. [CrossRef]
19. Jiang, J.; Dai, Y.; Yang, F.; Ma, Z. A multi-visit flexible-docking vehicle routing problem with drones for simultaneous pickup and delivery services. *Eur. J. Oper. Res.* **2024**, *312*, 125–137. [CrossRef]
20. Jiang, Y.; Liu, M.; Jia, X.; Xue, Q. The multi-visit vehicle routing problem with multiple heterogeneous drones. *Transp. Res. Part C Emerg. Technol.* **2025**, *172*, 105026. [CrossRef]
21. Bhuiyan, T.H.; Walker, V.; Roni, M.; Ahmed, I. Aerial drone fleet deployment optimization with endogenous battery replacements for direct delivery of time-sensitive products. *Expert Syst. Appl.* **2024**, *252*, 124172. [CrossRef]
22. Roytvand Ghiasvand, M.; Rahmani, D.; Moshref-Javadi, M. Data-driven robust optimization for a multi-trip truck-drone routing problem. *Expert Syst. Appl.* **2024**, *241*, 122485. [CrossRef]
23. Xing, J.; Guo, T.; Tong, L. Reliable truck-drone routing with dynamic synchronization: A high-dimensional network programming approach. *Transp. Res. Part C Emerg. Technol.* **2024**, *165*, 104698. [CrossRef]
24. Chagas, G.O.; Coelho, L.C.; Laganà, D.; Beraldi, P. A dynamic drone routing problem with uncertain demand and energy consumption. *Transp. Res. Part B Methodol.* **2025**, *202*, 103335. [CrossRef]
25. Sun, W.; Wu, L.; Zhang, F. Robust optimization for truck-and-drone collaboration with travel time uncertainties. *Transp. Res. Part B Methodol.* **2026**, *204*, 103378. [CrossRef]
26. Vidal, T. Hybrid genetic search for the CVRP: Open-source implementation and SWAP* neighborhood. *Comput. Oper. Res.* **2022**, *140*, 105643. [CrossRef]
27. Wouda, N.A.; Lan, L.; Kool, W. PyVRP: A high-performance VRP solver package. *Inf. J. Comput.* **2024**, *36*, 943–955. [CrossRef]
28. Madani, B.; Ndiaye, M.; Salhi, S. Hybrid truck-drone delivery system with multi-visits and multi-launch and retrieval locations: Mathematical model and adaptive variable neighborhood search with neighborhood categorization. *Eur. J. Oper. Res.* **2024**, *316*, 100–125. [CrossRef]

29. Wouda, N.A.; Lan, L. ALNS: A Python implementation of the adaptive large neighbourhood search metaheuristic. *J. Open Source Softw.* **2023**, *8*, 5028. [CrossRef]
30. Voigt, S. A review and ranking of operators in adaptive large neighborhood search for vehicle routing problems. *Eur. J. Oper. Res.* **2025**, *322*, 357–375. [CrossRef]
31. Wu, H.; Bao, X. An adaptive large neighborhood search method for the two-echelon vehicle routing problem with clustered customers. *Algorithms* **2026**, *19*, 387. [CrossRef]
32. Lorenz, C.; Mimmo, N.; Otto, A.; Vigo, D. Very large-scale neighborhood search for drone routing with energy replenishment. *OR Spectr.* **2026**, *48*, 587–626. [CrossRef]
33. Wang, J.; Wang, W.C.; Hu, X.X.; Qiu, L.; Zang, H.F. Black-winged kite algorithm: A nature-inspired meta-heuristic for solving benchmark functions and engineering problems. *Artif. Intell. Rev.* **2024**, *57*, 98. [CrossRef]
34. Liu, W.; He, Y.; Ke, G.Y.; Zhang, L. On-demand meal delivery: Drone scheduling with battery replacement optimization. *Comput. Oper. Res.* **2026**, *185*, 107295. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.