

Operational Decision Support for Evacuating Nursing Home Residents: A Hamburg Benchmark and Metaheuristic Comparison Under a 5-Minute Time Budget

Nick Alexander
nick.alexander@hcu-hamburg.de
HafenCity University Hamburg
Hamburg, Germany

Milva Tannenbaum
milva.tannenbaum@bis.hamburg.de
Ministry of the Interior and Sports
Hamburg, Germany

Jörg Rainer Noennig
joerg.noennig@hcu-hamburg.de
HafenCity University Hamburg
Hamburg, Germany
TU Dresden
Dresden, Germany

Abstract

Evacuating mobility-impaired nursing home residents during urban emergencies requires routing and scheduling heterogeneous vehicles under minute-level deadlines. We introduce the Nursing Home Evacuation Vehicle Routing Problem (NH-Evac-VRP) with open-ended multi-trip shuttling, split pickups, load-dependent service times, and shelter capacities; solutions must be produced within a hard 300 s single-thread budget. We release the Hamburg-NH-Evac benchmark with two expert-specified scenarios: (i) an unexploded-ordnance exclusion-zone evacuation in Altona-Altstadt (479 evacuees, one shelter), (ii) a storm-surge evacuation in Wilhelmsburg (510 evacuees, three shelters), plus (iii) a synthetic stress test (1,348 evacuees, five shelters). Instances use real facility data and asymmetric road-network travel times. Across five scenario–fleet configurations, we compare a constructive dispatcher with a Genetic Algorithm, Memetic Algorithm, and Adaptive Large Neighborhood Search, minimizing demand-weighted average waiting time and makespan with a shelter-overfill penalty. No method dominates: ALNS attains the best makespan in two configurations; in the heterogeneous *Flood–Augmented* case this advantage is driven by capacity-maximizing “sweeper” tours. MA is best in three configurations by favoring rapid shuttle cycles, outperforming ALNS by 27 min in the synthetic mass-transit case. We also provide a web-based decision support system for scenario configuration, optimization, and schedule visualization.

CCS Concepts

• **Computing methodologies** → **Evolutionary algorithms**; *Heuristic function construction*; • **Theory of computation** → **Routing and network design problems**.

Keywords

disaster logistics, evacuation planning, heterogeneous fleet VRP, split pickups, memetic algorithms, ALNS, decision support systems, real-world applications

ACM Reference Format:

Nick Alexander, Milva Tannenbaum, and Jörg Rainer Noennig. 2026. Operational Decision Support for Evacuating Nursing Home Residents: A Hamburg Benchmark and Metaheuristic Comparison Under a 5-Minute Time Budget. In *Genetic and Evolutionary Computation Conference (GECCO '26)*, July 13–17, 2026, San Jose, Costa Rica. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3795095.3805113>

1 Introduction

Urban disasters such as floods, storm surges, industrial incidents, and unexploded ordnance necessitate the rapid evacuation of vulnerable populations. Nursing home residents, who are predominantly mobility-impaired and require assisted boarding, present one of the most challenging evacuation scenarios. The logistical complexity of moving such populations under time pressure demands sophisticated planning tools capable of generating actionable transportation plans within minutes. We therefore impose a strict **300-second (5-minute) planning budget** as a hard constraint on all optimization methods. This threshold aligns with the iterative German command-and-control cycle [2], under which route plans must be revised as updated situational information arrives, and is more generous than the 180-second limit used in comparable bus evacuation optimization [19]. The budget was corroborated as operationally realistic by domain expertise at Hamburg’s Ministry of the Interior and Sports.

Hamburg, Germany, motivates this work due to its dual exposure to evacuation risks. First, unexploded World War II (WWII) ordnance (UXO) is frequently discovered during construction; the Hamburg Fire Department reports 10–20 large aerial bombs defused annually [12]. Recent examples include a 1,000-pound English bomb in Steinwerder requiring a 500-meter exclusion radius [15], a 250-kilogram bomb in Wilhelmsburg with a 300-meter radius [14], and a 500-pound American bomb in the densely populated Sternschanze district [13]. These discoveries require rapid evacuation to emergency shelters, typically schools, that provide temporary accommodation [17]. Second, Hamburg faces flood risks from the Elbe estuary. Approximately 45% of Hamburg’s area (342 km²) is flood-prone, with 326,000 residents and over EUR 10 billion in assets at risk [21]. Despite defenses improved since the catastrophic 1962 flood that claimed 315 lives, the hydrological context is complex. While observed increases in tidal water levels are primarily attributed to river engineering and flood defense measures rather than meteorological changes, climate projections indicate increasingly frequent extreme flood events [23]. These low-lying districts



This work is licensed under a Creative Commons Attribution 4.0 International License. *GECCO '26, San Jose, Costa Rica*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2487-9/2026/07
<https://doi.org/10.1145/3795095.3805113>

contain multiple care facilities requiring robust evacuation planning.

Vehicle routing for disaster response is well-studied, including supply distribution [31], mass-casualty transport [7], and general population evacuation [4, 20]. However, standard benchmarks like Solomon’s instances [35] and CVRPLIB [36], and even the recent 12th DIMACS Implementation Challenge [37], primarily model commercial distribution. While these initiatives successfully benchmark variants like split deliveries in isolation, they fail to capture the distinct operational logic of institutional evacuation, a disconnect extensively documented in humanitarian logistics surveys [3, 11]. Key distinguishing features include: (1) **split pickups**, mandated when facility populations exceed vehicle capacity; (2) **load-dependent service times** due to assisted boarding; (3) **finite shelter capacities**; and (4) **open-ended shuttle topology** (no return-to-depot between trips). While these features individually appear in the Split Delivery VRP [9], Heterogeneous Fleet VRP [24], and Multi-Trip VRP [6], no prior formulation combines all four; the closest evacuation models [4, 34] cover at most three, and their interaction (split decisions coupled with open-ended trip chaining and shelter assignment) creates coordination challenges absent from individual variants. Yet, despite these operational realities, major community efforts in evolutionary computation continue to prioritize the idealized formulations found in standard benchmarks [28], highlighting a critical gap in testing grounds for evacuation-specific constraints.

These constraints fundamentally alter the optimization landscape. The problem is dominated by *synchronization bottlenecks*, where high service times must be balanced against the fast cycle times of high-capacity vehicles. This necessitates coordinating a heterogeneous fleet (mixing specialized medical transport with requisitioned buses) in an open-ended shuttle topology. High demand mandates split pickups [1], requiring multiple vehicles to cooperate in clearing facilities while simultaneously balancing loads against strict shelter capacities. We formally define this as the Nursing Home Evacuation Vehicle Routing Problem (NH-Evac-VRP).

To address this, we evaluate three metaheuristic paradigms. Genetic Algorithms (GA) have shown effectiveness on rich Vehicle Routing Problems (VRPs) through specialized crossover operators [5], while Memetic Algorithms (MA) hybridize evolutionary search with local refinement to intensify exploitation [25, 38]. Adaptive Large Neighborhood Search (ALNS) [33] offers an alternative based on ruin-and-recreate operations guided by adaptive weights [32]. Recent surveys confirm ALNS’s flexibility [39], yet the relative performance of population-based versus single-trajectory methods under strict 300-second deadlines remains non-obvious. Our experiments reveal no single method dominates (Table 2): ALNS attains the best makespan on *UXO-Specialized* and *Flood-Augmented*; in *Flood-Augmented*, this is driven by capacity-maximizing “sweeper” routing with buses. MA achieves the best results on *UXO-Augmented*, *Flood-Specialized*, and *Synthetic-Mass-Transit* by favoring rapid shuttle cycling, underscoring the value of scenario-aware algorithm selection.

We address the *Nursing Home Evacuation Vehicle Routing Problem* (NH-Evac-VRP) and present the following contributions:

- (1) **Hamburg-NH-Evac Benchmark**: Two expert-designed real-world scenarios (UXO Discovery and Flood) and one synthetic stress-test. Scenarios utilize real travel times and

facility data [16], validated by a municipal disaster management professional.

- (2) **Algorithm Comparison**: A systematic evaluation of a Constructive Dispatcher, GA, MA, and ALNS under an identical 300 s planning budget. We optimize the scalar objective $J = \overline{W} + T_{\max} + \beta \cdot P_{\text{over}}$ (Eq. 3) and report $\overline{W}$ and $T_{\max}$ separately, including statistical tests and convergence behavior.
- (3) **Decision Support System (DSS)**: An interactive tool integrating scenario configuration, optimization, and visualization. This addresses the need for interpretable outputs and rapid what-if analysis in emergency workflows [18, 26].

The remainder of this paper is organized as follows. Section 2 defines the NH-Evac-VRP. Section 3 describes the Hamburg scenarios. Section 4 presents the DSS. Section 5 details the optimization methods. Section 6 describes the experimental design, followed by results in Section 7, discussion in Section 8, and conclusions in Section 9.

2 Problem Definition: NH-Evac-VRP

We formalize the NH-Evac-VRP by specifying input entities, service time modeling, solution structure, and optimization objectives.

2.1 Inputs and Entities

The NH-Evac-VRP is defined over a set of pickup locations (nursing homes requiring evacuation), destination nodes (emergency shelters), and a heterogeneous vehicle fleet operating on a road network with precomputed travel times.

Travel Times. Road network travel times are obtained via the OpenRouteService API [22] using the driving-car profile, which accounts for road types, speed limits, and turn restrictions. We precompute a complete origin-destination matrix covering the vehicle staging area, all pickup locations, and all shelters. The resulting matrix is asymmetric, reflecting one-way streets and directional turn costs present in the Hamburg road network. Travel times are treated as static throughout the planning horizon, an assumption appropriate for incidents where evacuation occurs before major traffic disruption or is actively managed by police traffic control, though we acknowledge this as a limitation for scenarios involving concurrent mass evacuation or infrastructure damage. Precomputed matrices are released with the benchmark to ensure reproducibility.

Pickup Nodes (Nursing Homes). Each nursing home i has an associated demand d_i representing the number of residents requiring evacuation, expressed in standard patient equivalents (seats). Facility locations and capacities are drawn from the Hamburg Transparenzportal [16], which provides publicly available data on licensed inpatient care facilities including addresses and registered bed counts. For each scenario, the set of affected facilities is determined by the incident’s geographic extent, for example, all nursing homes within a UXO exclusion zone or within a flood-threatened district.

Shelters. Emergency shelters serve as evacuation destinations, with each shelter s having a finite capacity C_s . In Hamburg, schools are typically designated as temporary emergency shelters [17] due to their large assembly spaces and sanitary facilities. We model an *open-ended shuttle topology*: for each vehicle, the destination

shelter of trip t is the start location of trip $t+1$ (i.e., trips are chained without intermediate returns to a central depot).

Vehicles. The fleet is heterogeneous, comprising a fleet $\mathcal{B}$, where each vehicle $b \in \mathcal{B}$ has a distinct passenger capacity Q_b expressed in standard seat-equivalents. This abstraction allows the model to handle a mix of low-capacity specialized transport and high-capacity mass transit units simultaneously. All vehicles begin at a designated staging area representing pre-positioned emergency resources. Fleet compositions vary by scenario and configuration, as detailed in Section 3.

2.2 Dynamic Service Time Model

Service time at each location depends on the number of evacuees handled, reflecting the assisted boarding requirements of mobility-impaired nursing home residents. For a vehicle picking up or delivering q evacuees, the service time is:

$$\tau(q) = \tau_0 + \alpha q, \quad (1)$$

where $\tau_0 = 3$ min accounts for vehicle positioning and paperwork, and $\alpha = 20$ s/person ≈ 0.33 min/person reflects per-evacuee boarding assistance. This linear model, validated by a senior municipal disaster management professional as a reasonable operational approximation, applies symmetrically to both pickup (boarding) and delivery (alighting) operations.

For example, a vehicle collecting 7 evacuees incurs $3 + 7 \times \frac{20}{60} \approx 5.3$ min of service time at pickup, plus an equivalent 5.3 min upon shelter delivery. At multi-stop trips, service time accumulates independently at each facility visited.

2.3 Decision Variables and Feasibility

Let $\mathcal{B}$, $\mathcal{N}$, and $\mathcal{H}$ denote the sets of vehicles, facilities, and shelters. A solution assigns each vehicle $b \in \mathcal{B}$ a sequence of trips $t = 1, \dots, T_b$; trip t is a tuple $(o_{bt}, h_{bt}, [(n_{btr}, q_{btr})]_{r=1}^{R_{bt}})$, where $o_{bt} \in \{v_0\} \cup \mathcal{H}$ is the origin, $h_{bt} \in \mathcal{H}$ the destination shelter, and $(n_{btr}, q_{btr}) \in \mathcal{N} \times \mathbb{Z}_{\geq 0}$ the facility and collected quantity at stop r . This encodes split pickups, multi-trip shuttling, and heterogeneous capacities.

Three hard constraints must hold:

$$\sum_{b,t,r: n_{btr}=i} q_{btr} = d_i \quad \forall i \in \mathcal{N} \quad (C1)$$

$$\sum_r q_{btr} \leq Q_b \quad \forall b, t \quad (C2)$$

$$o_{b1} = v_0 \quad \forall b, \quad o_{b,t+1} = h_{bt} \quad \forall b, t < T_b \quad (C3)$$

(C1) guarantees full evacuation; (C2) enforces per-trip vehicle capacity; (C3) encodes the open-ended shuttle topology. Route timings are induced by cumulative travel and service times $\tau(q)$, defining t_{btr}^{pickup} and $T_{\max}$. Shelter capacity is a soft constraint penalized via P_{over} (Section 2.4).

2.4 Objectives and Evaluation

We report three metrics capturing distinct operational concerns:

Average Waiting Time. The average waiting time $\bar{W}$ measures mean time from evacuation start until boarding service begins:

$$\bar{W} = \frac{1}{D} \sum_{b,t,r} q_{btr} \cdot t_{btr}^{\text{pickup}}, \quad (2)$$

where $D = \sum_i d_i$ is total demand, q_{btr} is the number of evacuees collected at stop r of trip t of vehicle b , and t_{btr}^{pickup} is the time at which boarding begins at that stop. Under split pickups, evacuees are counted at the start time of the stop on which they are collected.

Makespan. The makespan $T_{\max}$ is the time at which the last evacuee arrives at a shelter, representing total evacuation duration.

Shelter Overfill. The overfill penalty $P_{\text{over}} = \sum_{s \in \mathcal{H}} \max(0, L_s - C_s)$ sums excess load across shelters, where $L_s = \sum_{b,t: h_{bt}=s} \sum_r q_{btr}$ is the load delivered to shelter s and C_s its capacity. Aggregate shelter capacity exceeds total demand in all scenarios, but individual limits remain relevant (e.g., Flood: 510 evacuees across three shelters of capacity 200). Strict feasibility corresponds to $P_{\text{over}} = 0$; small overfill is tolerated but penalized.

Scalar Objective. We optimize the scalar objective:

$$J = \bar{W} + T_{\max} + \beta \cdot P_{\text{over}}, \quad (3)$$

where $\beta = 1$ min/person converts shelter overflow into equivalent system delay. Shelter capacity is treated as a soft constraint reflecting operational reality: in emergencies, a shelter at 105% capacity still accepts evacuees rather than turning them away. The penalty term discourages load imbalance while ensuring the solver always returns a complete evacuation plan. Since aggregate shelter capacity exceeds total demand in all benchmark scenarios, this soft penalty proves sufficient; experimental results (Section 7) confirm that all algorithms converge to strictly feasible solutions ($P_{\text{over}} = 0$).

3 Hamburg Benchmark Suite

This section presents three benchmark scenarios: two expert-designed scenarios based on realistic incident parameters, and one synthetic stress-test evaluating algorithm scalability under resource scarcity.¹

All scenarios share a common vehicle pool drawn from Hamburg's emergency resources. The specialized fleet comprises patient transport vehicles (N-KTW, *Notfall-Krankentransportwagen*) which are equipped similar to emergency ambulances (CEN 1789 Type B) [8]. For the purpose of the NH-Evac-VRP model, we assign these vehicles an operational capacity of 2 seat-equivalents. Multi-purpose vehicles (MZF, *Mehrzweckfahrzeuge*) are modeled with a capacity of 7. The augmented configuration adds requisitioned public transit buses. To ensure safety for mobility-impaired passengers, we define strict *operational capacities* below manufacturer maximums: 40 for standard buses (vs. ~ 106 total [30]) and 80 for articulated buses (vs. 199 total [29]). This conserves space for medical staff and mobility aids while preventing overcrowding. We evaluate five scenario–fleet configurations with per-round capacities of 22 (*UXO–Specialized*), 62 (*UXO–Augmented*), 107 (*Flood–Specialized*), 187 (*Flood–Augmented*), and 240 (*Synthetic–MassTransit*).

3.1 Scenario I: UXO Discovery (Altona-Altstadt)

This scenario models the discovery of unexploded ordnance (UXO), specifically a 500-pound World War II aerial bomb at coordinates

¹Scenario definitions, fleet configurations, precomputed travel-time matrices, algorithm implementations, and reproduction scripts (including random seeds) are available at <https://github.com/NickAlexx/Hamburg-NH-Evac-benchmark>. An archival snapshot is deposited at Zenodo (DOI: <https://doi.org/10.5281/zenodo.19474435>).

(53.5592°N, 9.9592°E) in the Altona-Altstadt district, a recurring event in Hamburg [12]. The exclusion zone encompasses all residents within a 300 m radius of the discovery site.

Three nursing homes fall within the exclusion zone, with demands of 199, 80, and 200 residents (479 total, visualized in Figure 1). A single designated emergency shelter with capacity 500 serves as the destination. This configuration represents a dense urban scenario with short travel distances, removing the load-balancing constraint present in multi-shelter scenarios.

Fleet configurations:

- *Specialized*: 4 N-KTW (capacity 2) + 2 MZF (capacity 7) = 22 seats per round
- *Augmented*: Specialized fleet + 1 Bus (capacity 40) = 62 seats per round

The specialized fleet requires approximately 22 fleet rotations to evacuate all residents, creating a throughput-limited scenario where augmentation with a single bus nearly triples per-round capacity.

3.2 Scenario II: Flood (Wilhelmsburg)

This scenario models a storm surge affecting Wilhelmsburg, a low-lying district on the Elbe river island. The low-lying urban district was devastated by the 1962 North Sea flood that claimed 315 lives in Hamburg [27]. All four nursing homes in the district require evacuation, with a combined demand of 510 evacuees.

Unlike the UXO scenario, evacuees must be distributed across three school shelters:

- Schule Mümmelmannsberg (capacity 200)
- Alexander-von-Humboldt-Gymnasium (capacity 200)
- Schule Maretstraße (capacity 200)

Although aggregate shelter capacity (600) exceeds demand (510), individual shelter limits still apply. Algorithms must balance loads across destinations, as naive greedy assignment to the nearest shelter would overload individual facilities. Shelters must be sited on higher ground outside the flood zone, creating substantial travel distances.

Fleet configurations:

- *Specialized*: 22 N-KTW (capacity 2) + 9 MZF (capacity 7) = 107 seats per round
- *Augmented*: Specialized fleet + 2 Buses (capacity 40) = 187 seats per round

3.3 Scenario III: Synthetic–MassTransit

This synthetic scenario stress-tests algorithm performance under resource scarcity and high vehicle utilization. It encompasses 9 facilities across the Wilhelmsburg and Harburg districts with a combined demand of 1,348 evacuees distributed across 5 shelters (total capacity 1,700).

The fleet consists of only 3 articulated buses (capacity 80 each), providing 240 seats per round. This configuration forces vehicles into extended duty cycles with approximately 6 complete shuttle rotations required, testing algorithms' ability to optimize packing efficiency and minimize deadhead travel over long planning horizons.

Fleet configuration:

- *MassTransit*: 3 articulated buses (capacity 80) = 240 seats per round

4 Decision Support System

To bridge algorithmic output to operational use, we developed a web-based decision support system enabling emergency planners to configure scenarios, execute optimization, and interpret results within a unified interface.

The system provides three predefined scenario presets: UXO discovery, flood, and synthetic stress-test. Each loads appropriate facility sets, shelter locations, and fleet configurations. Planners can modify these defaults by adjusting exclusion zone radii, adding or repositioning shelters via map interaction, and configuring heterogeneous fleets with per-vehicle capacities and custom starting positions. Advanced settings expose service time parameters (base time τ_0 and per-person duration α), enabling scenario-specific calibration. Algorithm selection (Dispatcher, GA, MA, or ALNS) and time budget are specified before execution.

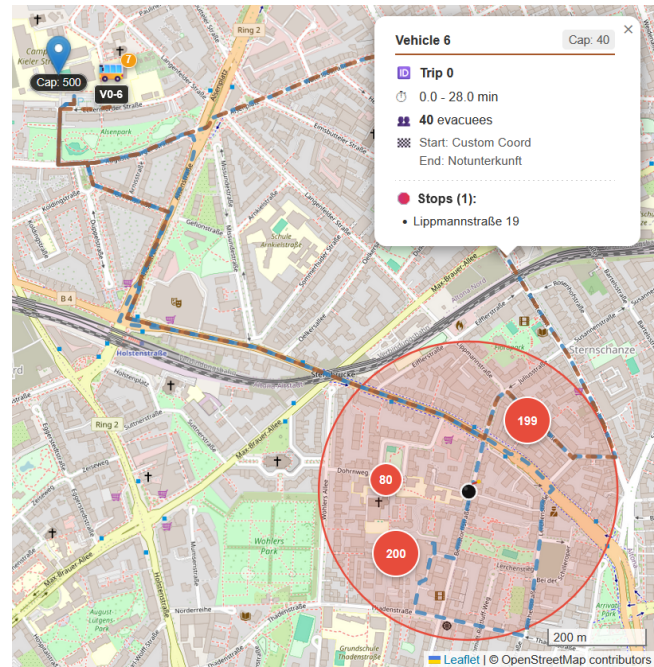


Figure 1: DSS map view of the UXO Discovery scenario showing the exclusion zone, affected nursing homes, shelter, and staging area (V0–6). A selected route is highlighted with trip details. (Map data: © OpenStreetMap contributors).

Visualization integrates an interactive map displaying facility markers, shelter locations, and color-coded vehicle routes (Figure 1); hovering over routes reveals trip details. The interface also provides a Gantt-style timeline showing per-vehicle trip schedules and a route summary modal with text descriptions for field coordination.

The system enforces the specified optimization budget. Shelter overflow, if any, is reported explicitly and enters the objective via P_{over} (Eq. 3). The complete interface, including fleet configuration

and timeline panels, is documented in the supplementary repository. While the interface design was informed by discussions with municipal emergency management personnel, formal usability evaluation remains future work.

5 Optimization Methods

We address the operational constraints of the NH-Evac-VRP using a Constructive Dispatcher baseline and three metaheuristic paradigms: GA, MA [38], and ALNS [33]. This selection covers dominant approaches in VRP literature [10, 39]. All methods operate on a common representation where each vehicle is assigned a sequence of trips specifying origin, destination, and ordered pickup stops. This supports split pickups, multi-trip routing, and heterogeneous fleets, with fitness computed via Equation (3).

5.1 Constructive Dispatcher Baseline

The Constructive Dispatcher models the behavior of an experienced human coordinator with perfect information: full knowledge of demand, travel times, and shelter capacities. It operates in event-driven fashion: the vehicle with earliest availability is assigned to the nearest facility with remaining demand. If the resulting load falls below 60% capacity, a second stop minimizing detour is added. Vehicles deliver to the nearest shelter with sufficient capacity. This greedy process repeats until all demand is satisfied. Despite its simplicity, this baseline produces strong solutions by emulating rational dispatching decisions; it serves as the seed solution for the metaheuristics and provides a meaningful performance floor.

5.2 Genetic Algorithm (GA)

The GA maintains a population of 200 individuals evolved through selection, crossover, and mutation. The initial population combines a Constructive Dispatcher seed, mutated variants, and random solutions. Tournament selection (size 3) chooses parents. Crossover uses a hybrid strategy combining capacity-compatible sub-schedule exchange, best-trip injection, and a temporal time-slice swap; extracted trips are reinserted greedily.

Six mutation operators provide neighborhood exploration: *intra-trip swap* reorders stops within a trip; *relocate stop* moves stops between trips; *add/remove trip* adjusts solution structure; *change depot* modifies shelter assignments; *swap trips* exchanges trips between vehicles; *spatial ruin-recreate* removes and reinserts spatial clusters. A repair function restores feasibility after each operation, ensuring demand satisfaction, capacity compliance, and depot connectivity.

5.3 Memetic Algorithm (MA)

The MA extends the GA with local search applied to selected individuals. A portfolio of 15 operators targets different solution aspects, including: sequence optimization (*intra_trip*), inter-trip moves (*relocate*, *swap_stops*, *move_trip*), depot reassignment (*change_depot*), trip restructuring (*consolidate*, *split_mixed*), makespan reduction (*balance_makespan*, *takeover_gap*), and packing efficiency (*crumb_extract*, *self_consolidate*).

Operators are scheduled adaptively based on improvement-per-second; those failing to improve are temporarily disabled. Operator

weights shift across the optimization timeline: early phases emphasize structural changes, late phases emphasize makespan reduction and consolidation/packing. Local search is applied selectively to maintain generational throughput: elites receive longer budgets (60ms), random offspring shorter budgets (30ms).

5.4 Adaptive Large Neighborhood Search (ALNS)

ALNS iteratively modifies a single solution through destroy-and-repair operations with adaptive operator selection. It initializes from the Constructive Dispatcher solution with several diversification “shakes.”

Six destroy operators remove 8–28% of stop-entries: *random* provides unbiased diversification; *route removal* eliminates entire trips; *worst detour* targets inefficient stops; *Shaw relatedness* removes spatially related clusters; *node cluster* removes all visits to a geographic region; *bottleneck* targets the latest-finishing vehicle. Three repair operators reinsert demand: *greedy* minimizes immediate cost; *regret-2* and *regret-3* prioritize requests that would suffer most from suboptimal placement.

Operator weights update based on outcome quality (new best, improvement, or accepted). Simulated annealing acceptance allows uphill moves with probability decreasing over time (cooling rate 0.999); temperature reheats after prolonged stagnation.

6 Experimental Design

We detail the computational infrastructure, evaluation protocols, and statistical framework used to benchmark the algorithms under the strict 300-second operational time budget.

6.1 Computational Setup

All experiments were conducted on a workstation equipped with an AMD Ryzen 5 2600X CPU (6 cores, 3.6 GHz) and 32 GB of RAM, running Microsoft Windows 11 Pro (Build 26200).

The optimization algorithms were developed as a **custom implementation** in Python 3.11.0 to allow for specialized heterogeneous fleet operators. Key computational libraries included numpy (v1.26.4) for matrix operations and metric calculations. Statistical validation was performed using scipy (v1.15.3). Travel time matrices were queried using the openrouteservice client (v2.3.3). Each algorithm instance was restricted to a single CPU thread.

6.2 Evaluation Protocol

We evaluate five scenario-fleet configurations: *UXO-Specialized*, *UXO-Augmented*, *Flood-Specialized*, *Flood-Augmented*, and *Synthetic-MassTransit*. Each of the four algorithms (Dispatcher baseline, GA, MA, ALNS) was executed under a strict 300-second optimization budget. Travel time matrices were precomputed via OpenRouteService prior to experimentation and excluded from the time budget to ensure fairness and reproducibility; all algorithms received identical cached matrices. Each stochastic algorithm (GA, MA, ALNS) was run 30 times with deterministic seeding to enable statistical comparison.

6.3 Metrics Reported

We report three primary metrics: average waiting time $\bar{W}$, makespan $T_{\max}$, and total cost J as defined in Equation (3). Convergence efficiency is measured as the wall-clock time required to reach 95% of each run's total improvement (i.e., the first time $J(t) \leq J_0 - 0.95(J_0 - J^*)$), where J_0 is the initial cost and J^* is the final best cost at termination.

6.4 Statistical Testing

To assess significance among stochastic methods, we apply paired Wilcoxon signed-rank tests (one-sided, $\alpha = 0.05$) with runs matched by random seed. The deterministic Dispatcher baseline is excluded from statistical comparison. Within each scenario–fleet configuration and metric, we identify the best-performing stochastic method by lowest mean over 30 runs, then test whether each remaining stochastic method is significantly worse. With three stochastic algorithms (GA, MA, ALNS) across five configurations and three metrics, this yields $5 \times 3 \times 2 = 30$ pairwise tests. We apply Holm–Bonferroni correction globally across all 30 tests to control the family-wise error rate. In Table 2, results marked with * indicate statistically significant inferiority to the best result ($p < 0.05$ after correction).

6.5 Parameter Settings

Table 1 summarizes key algorithm parameters. All parameters were held fixed across all scenarios (no scenario-specific retuning).

Table 1: Algorithm parameter settings.

Algorithm	Parameter	Value
GA / MA	Population size	200
	Crossover rate	0.8
	Mutation rate	0.2
	Tournament size	3
ALNS	Destroy fraction	8–28%
	Segment length	50
	Simulated annealing cooling rate	0.999
	Regret- k	$k \in \{2, 3\}$
MA only	LS budget (elite/offspring)	60 / 30 ms

7 Experimental Results

We compare algorithm performance, analyze emergent routing strategies, and examine convergence behavior and model sensitivity.

7.1 Performance Overview

Table 2 summarizes the performance of the Constructive Dispatcher baseline, GA, MA, and ALNS across the five scenarios.

The results demonstrate that no single algorithm dominates across all operational contexts. The ALNS achieves the best Makespan in the *UXO–Specialized* and *Flood–Augmented* scenarios, efficiently managing complex packing constraints. However, the MA dominates the *Synthetic–MassTransit* scenario, outperforming ALNS by 27 minutes in Makespan (480.7 min vs. 507.2 min). The GA proves

competitive in the *UXO–Augmented* and *Flood–Specialized* scenarios, often striking a balance between cost and consistency.

Notably, in the *Synthetic–MassTransit* and *Flood–Specialized* scenarios, ALNS did not consistently improve upon the Dispatcher baseline within the 300 s budget (Table 2). This suggests that in settings where near-direct shuttling is already close to throughput-optimal, packing-oriented ruin-and-recreate moves can introduce structural inefficiencies. Consistent with the rapid plateauing observed (Table 3), these runs often stabilized early, motivating a do-no-harm deployment rule that returns the better of the Dispatcher and metaheuristic outputs (Section 8).

Table 2: Performance comparison (mean $\pm$ std. dev. over 30 runs; times in minutes). All solutions achieved $P_{\text{over}} = 0$. Bold indicates best result. (*) denotes statistically significant inferiority to best stochastic method (paired Wilcoxon signed-rank, one-sided, Holm–Bonferroni corrected, $p < 0.05$). Dispatcher is deterministic and excluded from statistical tests.

Scenario	Method	Avg Wait	Makespan	Total Cost
UXO Specialized	Dispatcher	224.5 $\pm$ 0.0	478.7 $\pm$ 0.0	703.2 $\pm$ 0.0
	GA	224.9 $\pm$ 0.3	477.3 $\pm$ 1.4*	702.2 $\pm$ 1.4*
	MA	225.0 $\pm$ 0.5	473.3 $\pm$ 1.9*	698.3 $\pm$ 1.9*
	ALNS	225.3 $\pm$ 1.0	471.5 $\pm$ 1.5	696.8 $\pm$ 1.9
UXO Augmented	Dispatcher	115.2 $\pm$ 0.0	259.3 $\pm$ 0.0	374.5 $\pm$ 0.0
	GA	114.4 $\pm$ 0.4*	249.1 $\pm$ 1.7	363.5 $\pm$ 1.9
	MA	114.3 $\pm$ 0.4*	248.3 $\pm$ 1.4	362.6 $\pm$ 1.5
	ALNS	113.9 $\pm$ 0.5	251.5 $\pm$ 0.4*	365.4 $\pm$ 0.4*
Flood Specialized	Dispatcher	100.4 $\pm$ 0.0	244.0 $\pm$ 0.0	344.4 $\pm$ 0.0
	GA	100.6 $\pm$ 0.4	229.5 $\pm$ 1.6	330.1 $\pm$ 1.5
	MA	100.6 $\pm$ 0.4	229.3 $\pm$ 1.5	329.9 $\pm$ 1.2
	ALNS	109.0 $\pm$ 1.2*	243.1 $\pm$ 2.7*	352.5 $\pm$ 2.6*
Flood Augmented	Dispatcher	66.2 $\pm$ 0.0	182.4 $\pm$ 0.0	248.6 $\pm$ 0.0
	GA	65.8 $\pm$ 0.7	177.4 $\pm$ 1.9*	243.2 $\pm$ 2.1*
	MA	65.4 $\pm$ 0.2	174.0 $\pm$ 3.1*	239.4 $\pm$ 3.1*
	ALNS	68.5 $\pm$ 0.8*	161.7 $\pm$ 2.7	230.3 $\pm$ 3.0
Synthetic MassTransit	Dispatcher	203.0 $\pm$ 0.0	495.9 $\pm$ 0.0	698.9 $\pm$ 0.0
	GA	201.6 $\pm$ 1.0*	482.8 $\pm$ 2.1*	684.4 $\pm$ 2.7*
	MA	201.0 $\pm$ 0.7	480.7 $\pm$ 1.3	681.8 $\pm$ 1.1
	ALNS	215.1 $\pm$ 5.7*	507.2 $\pm$ 7.6*	722.3 $\pm$ 12.4*

7.2 Forensic Analysis of Routing Strategies

To understand the drivers of these performance differences, we analyzed the routing topology using the average stops per trip as a behavioral proxy. In the *UXO* scenarios, high demand density forced a convergence of strategies: both MA and ALNS adopted direct shuttling (avg. 1.0 stops/trip) to maximize throughput over short distances. However, significant behavioral divergence emerged in the dispersed scenarios. To illustrate these structural differences, Figure 2 shows example schedules illustrating the contrasting strategies.

The “Sweeper” Effect (Flood–Augmented). In this heterogeneous scenario, ALNS outperformed MA by using high-capacity buses (Capacity 40) to consolidate dispersed demand. As visualized in the top panel of Figure 2, ALNS frequently assigned buses to multi-stop

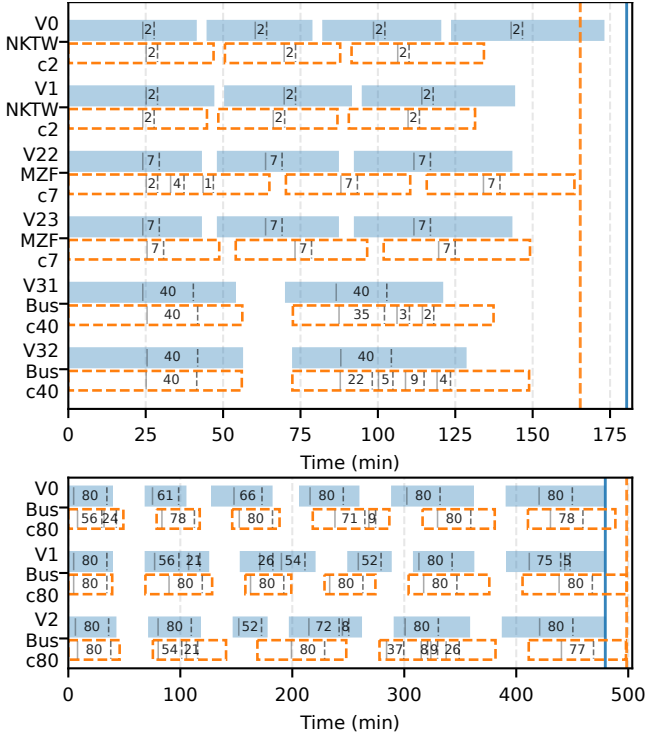


Figure 2: Gantt schedules comparing MA and ALNS on Flood-Augmented (top) and Synthetic-MassTransit (bottom); for readability, only a representative subset of vehicles is shown in the top panel. Solid bars denote MA; dashed-outline bars denote ALNS. Vertical lines mark the makespan $T_{\max}$ for each method. Numbers inside bars denote pickup quantities; the x-position of the labeled segment indicates the time at which the corresponding stop is served.

tours; averaged across all 30 runs, buses performed **1.56 stops/trip** under ALNS compared to **1.22** under MA. For instance, Vehicles 31 and 32 are shown collecting fragmented groups (e.g., 35, 3, and 2 passengers) in single runs. While this increased the cycle time of individual buses, it had a decisive impact on **workload balancing**. By using buses to “sweep” dispersed remainders, ALNS kept the high-capacity fleet active for longer, preventing the small-capacity fleet (Capacity 2–7) from becoming the bottleneck. In contrast, the MA favored single-source pickups for buses. Consequently, high-capacity buses often finished their schedules early, unable to find full loads, while the small fleet was forced to perform disproportionately many trips to clear the remaining fragmented demand, significantly extending the global Makespan.

The “Milk Run” Trap (Synthetic-MassTransit). In the homogeneous large-bus scenario, the ALNS packing strategy proved detrimental (Figure 2, bottom). Logs show ALNS buses frequently performing multi-stop tours (often 3–4 pickups) to fill the final 10–20 seats, driving the average up to **1.37 stops/trip**. The figure highlights a specific instance where an ALNS bus chains pickups of 37, 8, 9, and 26 passengers. Given the significant travel distances

and fixed per-stop service overhead ($\tau_0 = 3$ min), these “milk runs” increased cycle time disproportionately to the capacity gained. Averaged across all 30 runs, ALNS produced **1.37 stops/trip** versus **1.28** for MA, with utilization of 99% versus 93%. The MA instead favored predominantly single-stop shuttles, accepting lower utilization in exchange for rapid vehicle turnover, resulting in a superior Makespan (480.7 min vs. 507.2 min).

Over-Optimization of Small Fleets (Flood-Specialized). In the specialized fleet scenario, ALNS exhibited a tendency to maximize fill rates even for small vehicles (Capacity 2), resulting in **1.05 stops/trip** for ambulances compared to **1.00** for the MA. We observed ambulances performing 2-stop tours to collect single individuals from distant locations. In this dispersed setting, the “detour cost” to collect a single additional passenger significantly degraded the fleet’s effective throughput compared to the MA’s direct-shuttle approach.

7.3 Convergence Behavior

Table 3 reports the mean time required to reach 95% of each run’s total improvement. For ALNS, convergence timing is diagnostic in this benchmark: in scenarios where ALNS achieves the best results (*UXO-Specialized*, *Flood-Augmented*), improvements continue late into the time budget (212.8 s and 275.4 s), whereas in scenarios where it performs poorly (*Synthetic-MassTransit*, *Flood-Specialized*) the search often plateaus early (77.0 s and 147.1 s). MA converges quickly across all scenarios while maintaining consistently competitive performance.

Table 3: Convergence behavior. Mean time (seconds) over 30 runs to reach 95% of each run’s total improvement. For ALNS, later convergence coincides with better final solutions in this benchmark.

Method	UXO		Flood		Synth
	Spec.	Aug.	Spec.	Aug.	Mass.
GA	197.7	81.2	153.4	172.2	109.0
MA	184.8	133.5	116.0	156.6	69.5
ALNS	212.8	107.7	147.1	275.4	77.0

7.4 Ablation: Service Time Modeling

To assess sensitivity to service time modeling, we compared static versus dynamic formulations on the *Synthetic-MassTransit* scenario. The static model applies a fixed penalty based on vehicle capacity ($\tau = \tau_0 + \alpha \cdot Q_b$), while the dynamic model scales with actual load ($\tau(q) = \tau_0 + \alpha \cdot q$).

Under static service times, the Dispatcher baseline cost increased from 698.9 to 793.2 (+13.5%), as partial-load trips incur full-capacity penalties regardless of actual load. For example, a trip collecting 21 then 59 passengers pays 2×30 min under static versus 10 min + 23 min under dynamic. The MA achieved 764.6 ± 5.9 (10 runs), improving over the Dispatcher by 3.6% compared to 2.4% under dynamic times. This widening gap reflects baseline degradation rather than algorithmic improvement: the Dispatcher’s greedy

partial-load trips are disproportionately penalized under static assumptions. We report all primary results under dynamic service times as the operationally realistic and conservative setting.

8 Discussion, Limitations, and Future Work

Across scenarios, performance is driven less by instance size than by **fleet heterogeneity**, **spatial dispersion**, and **fixed per-stop overheads** (τ_0) that trade vehicle utilization against cycle time.

8.1 Proximity to Structural Bounds

Before analyzing behavioral differences, we note that the magnitude of achievable improvement is structurally constrained in several scenarios. The modest performance gaps in UXO–Specialized (0.9% between Dispatcher and ALNS) reflect this reality rather than algorithmic weakness. Routing improvements can only reduce the travel component, which represents just over half of trip duration: for a 7-passenger MZF trip, travel requires 12.0 min versus 10.6 min for boarding and alighting, leaving limited scope for optimization.

As a sanity check, we estimate a throughput-based floor on makespan. Each vehicle type has throughput $r_b = Q_b/t_b$, where Q_b is capacity and t_b is a representative round time (travel plus service). For N-KTW ($Q=2$, $t \approx 19.3$ min): $r \approx 0.10$ pax/min; for MZF ($Q=7$, $t \approx 22.6$ min): $r \approx 0.31$ pax/min. Summing across the fleet yields $4 \times 0.10 + 2 \times 0.31 \approx 1.01$ pax/min, suggesting a duration on the order of $479/1.01 \approx 474$ min—though this ignores first-leg overhead and discrete packing losses. The best observed makespan (471.5 min) falls in this range, consistent with algorithms operating in a regime of diminishing returns.

8.2 The Packing vs. Latency Trade-off

Across instances we observe a packing–latency trade-off driven by fixed stop overheads (τ_0) and heterogeneous bottlenecks. MA tends toward direct shuttles (“block mover”), whereas ALNS more often consolidates residual demand via multi-stop tours (“sweeper”).

In *Synthetic–MassTransit*, additional stops (ALNS: 1.37 vs. MA: 1.28 stops/trip) incur repeated τ_0 and increase tour time more than they improve utilization, so MA wins by higher turnover. In *Flood–Augmented*, bus sweeping (ALNS: 1.56 vs. MA: 1.22) reduces fragmented demand left to the small-capacity fleet and thus lowers the system bottleneck, improving makespan despite longer bus tours.

8.3 Operational Recommendations

Based on the distinct performance profiles observed, we derive three key guidelines for operational deployment:

- (1) **Mass Evacuation (Buses Only):** Consistent with the No Free Lunch perspective that algorithm performance depends on instance structure [40], we find that simple cycle-maximization (MA) outperforms packing-oriented logic in homogeneous contexts. Planners should prioritize direct shuttling here.
- (2) **Mixed Fleet Operations:** ALNS provides superior “sweeping” capability for dispersed demand. Notably, slow ALNS convergence (Table 3) correlates with strong final performance, indicating sustained search rather than computational inefficiency. Conversely, fast ALNS convergence signals early entrapment in suboptimal structures such as the “milk run” pattern; in such cases, the MA solution should

be preferred. Extended time budgets may benefit scenarios exhibiting late ALNS convergence.

- (3) **Baseline Preservation:** As ALNS does not consistently improve upon the Dispatcher baseline in all configurations (e.g., in *Synthetic–MassTransit*), the DSS should enforce a non-degradation rule, returning the better of the constructive plan and the metaheuristic output.

8.4 Limitations and Future Work

Several limitations warrant acknowledgment. Travel times are treated as static throughout the planning horizon. This assumption is appropriate for two operational contexts: acute response where evacuation proceeds before significant traffic disruption develops, and strategic resource planning where planners assess fleet adequacy and pre-position assets based on typical network conditions. However, static times are inadequate for reactive evacuation during ongoing disruption where congestion, road closures, or infrastructure damage dynamically alter travel conditions.

Evacuee heterogeneity is simplified: the model treats all nursing home residents as homogeneous seat equivalents, though in reality mobility levels vary significantly. Wheelchair users require different vehicle configurations than stretcher-bound patients, who also need medical assistance and repositioning before transport (~10 min/patient), overheads not captured in our uniform service-time model. Only nursing homes are considered; hospitals, assisted living facilities, and other care institutions with distinct evacuation characteristics remain out of scope.

Additionally, shelter capacity constraints did not bind tightly in our scenarios; stress-testing with scarcer shelter resources would exercise the load-balancing mechanisms more rigorously.

Future work should address several extensions. Incorporating stochastic travel times and scenario uncertainty would improve robustness. Dynamic replanning is essential for operational use: the optimizer should accept partially-executed states (vehicles mid-route with known positions and current loads) and warm-start reoptimization as situational awareness evolves, enabling continuous adaptation rather than single-shot optimization. Formal usability evaluation of the DSS with emergency management personnel would validate interface design decisions. Finally, extending the benchmark to additional cities and incident types would test applicability across diverse urban contexts.

9 Conclusion

We formalized the NH-Evac-VRP, introduced the Hamburg-NH-Evac benchmark comprising two expert-designed scenarios and a synthetic stress-test, and provided a web-based DSS for scenario configuration, optimization, and schedule visualization. Under a strict 300-second budget, no single algorithm dominated, underscoring the value of scenario-aware algorithm selection.

Acknowledgments

This work was supported by the German Federal Ministry of Research, Technology and Space (BMFTR) under the RESCUE-MATE project (grant no. 13N16840). Generative AI tools (OpenAI ChatGPT, Google Gemini, OpenAI Codex) assisted with drafting and coding; all outputs were reviewed and verified by the authors.

References

- [1] Claudia Archetti and Maria Grazia Speranza. 2012. Vehicle routing problems with split deliveries. *International Transactions in Operational Research* 19 (2012), 3–22. doi:10.1111/j.1475-3995.2011.00811.x
- [2] Ausschuss für Feuerwehrangelegenheiten, Katastrophenschutz und zivile Verteidigung (AFKzV). 1999. *Feuerwehr-Dienstvorschrift 100: Führung und Leitung im Einsatz – Führungssystem. Dienstvorschrift. Ständige Konferenz der Innenminister und -senatoren der Länder, Germany*.
- [3] Vural Bayram. 2016. Optimization models for large scale network evacuation planning and management: A literature review. *Surveys in Operations Research and Management Science* 21, 2 (2016), 63–84. doi:10.1016/j.sorms.2016.11.001
- [4] Douglas R. Bish. 2011. Planning for a bus-based evacuation. *OR Spectrum* 33, 3 (2011), 629–654. doi:10.1007/s00291-011-0256-1
- [5] Olli Bräysy, Wout Dullaert, and Michel Gendreau. 2004. Evolutionary algorithms for the vehicle routing problem with time windows. *Journal of Heuristics* 10, 6 (2004), 587–611. doi:10.1007/s10732-005-5431-6
- [6] Diego Cattaruzza, Nabil Absi, and Dominique Feillet. 2018. Vehicle Routing Problems with Multiple Trips. *Annals of Operations Research* 271, 1 (2018), 127–159. doi:10.1007/s10479-018-2988-7
- [7] Michael D. Dean and Suja K. Nair. 2014. Mass-casualty triage: Distribution of victims to multiple hospitals using the SAVE model. *European Journal of Operational Research* 238, 1 (2014), 363–373. doi:10.1016/j.ejor.2014.03.028
- [8] Deutsches Institut für Normung. 2024. *DIN EN 1789:2024-07: Medical vehicles and their equipment – Road ambulances*. German Standard (Deutsche Norm). Deutsches Institut für Normung, Berlin. 65 pages. doi:10.31030/3508320 Replaces DIN EN 1789:2020-12. Accessed: 2026-01-22.
- [9] Moshe Dror and Pierre Trudeau. 1989. Savings by Split Delivery Routing. *Transportation Science* 23, 2 (1989), 141–145. doi:10.1287/trsc.23.2.141
- [10] Manal El Jaouhari, Ghita Bencheikh, and Ghizlane Bencheikh. 2025. Metaheuristic and Reinforcement Learning Techniques for Solving the Vehicle Routing Problem: A Literature Review. *Journal of Traffic and Transportation Engineering (English Edition)* (2025). <https://jtte.chd.edu.cn/en/article/id/2d453e1a-80c9-4d0f-b3cf-9ac04842281c> In press. Accessed: 2026-01-10.
- [11] A. Esposito Amideo, M.P. Scaparra, and K. Kotiadis. 2019. Optimising shelter location and evacuation routing operations: The critical issues. *European Journal of Operational Research* 279, 2 (2019), 279–295. doi:10.1016/j.ejor.2018.12.009
- [12] Feuerwehr Hamburg. 2024. *Annual Report 2024*. Technical Report. Feuerwehr Hamburg. <https://www.feuerwehr.hamburg/resource/blob/1106020/b0a5cc2880e453da46acd06270cf71ab/d-jahresbericht-2024-englisch-data.pdf> English edition. Accessed: 2026-01-10.
- [13] Feuerwehr Hamburg. 2024. *Entschärfung einer Fliegerbombe in der Sternschanze*. <https://www.presseportal.de/blaulicht/pm/82522/5885224> Press release 2024-10-12. 500 lbs American bomb in Sternschanze. Accessed: 2025-12-13.
- [14] Feuerwehr Hamburg. 2025. *Kampfmittelräumdienst der Feuerwehr Hamburg entschärft erfolgreich den Zünder einer englischen Fliegerbombe*. <https://www.presseportal.de/blaulicht/pm/82522/6053469> Press release 2025-06-11. 250kg English bomb in Wilhelmsburg, 300m exclusion radius. Accessed: 2025-12-13.
- [15] Feuerwehr Hamburg. 2025. *Kampfmittelräumdienst der Feuerwehr Hamburg entschärft erfolgreich den Zünder einer englischen Fliegerbombe in unmittelbarer Nähe*. <https://www.presseportal.de/blaulicht/pm/82522/6058513> Press release 2025-06-18. 1000 Pfund English bomb in Steinwerder, 500m exclusion radius. Accessed: 2025-12-13.
- [16] Freie und Hansestadt Hamburg, Behörde für Arbeit, Gesundheit, Soziales, Familie und Integration. 2025. *Vollstationäre Pflegeeinrichtungen Hamburg. Dataset via MetaVer – Hamburger Metadatenkatalog*. <https://metaver.de/trefferanzeige?cmd=doShowDocument&docuuiid=CE58C4F7-4D45-429E-A556-53AA32FF6C7E> Data licence Germany – attribution – Version 2.0. Metadata updated: 2025-08-06. Accessed: 2026-01-10.
- [17] Freie und Hansestadt Hamburg, Bezirksamt Hamburg-Mitte. 2025. *Notunterkünfte Hamburg [Emergency Shelters]. Dataset via MetaVer – Hamburger Metadatenkatalog*. <https://metaver.de/trefferanzeige?cmd=doShowDocument&docuuiid=C5F3F844-A08D-43EB-9163-01204C9962EC> Provided by Dezernat Steuerung und Service, Katastrophen- und Zivilschutz. Data licence Germany – attribution – Version 2.0. Metadata updated: 2025-08-06. Accessed: 2025-10-07.
- [18] Gabriela Galindo and Rajan Batta. 2013. Review of recent developments in OR/MS research in disaster operations management. *European Journal of Operational Research* 230, 2 (2013), 201–211. doi:10.1016/j.ejor.2013.01.039
- [19] Marc Goerigk and Bob Grün. 2014. A robust bus evacuation model with delayed scenario information. *OR Spectrum* 36, 4 (2014), 923–948. doi:10.1007/s00291-014-0365-8
- [20] Marc Goerigk, Bob Grün, and Philipp Heßler. 2013. Branch and bound algorithms for the bus evacuation problem. *Computers & Operations Research* 40, 12 (2013), 3010–3020. doi:10.1016/j.cor.2013.07.006
- [21] Gabriele Gönner, Birgit Gerkenmeier, and Jan-Moritz Müller. 2012. *Ermittlung des Sturmflutbemessungswasserstandes für den öffentlichen Hochwasserschutz in Hamburg*. Berichte des Landesbetriebes Straßen, Brücken und Gewässer 12/2012. Landesbetrieb Straßen, Brücken und Gewässer (LSBG), Hamburg. <https://lsbg.hamburg.de/resource/blob/784544/51b12e1b6572fc3a3012a321a93fe60d/bericht-nr-12-ermittlung-des-sturmflutbemessungswasserstandes-fuer-den-oeffentlichen-hochwasserschutz-in-hamburg-data.pdf> Accessed: 2025-07-28.
- [22] HeiGIT gGmbH. 2025. *Openrouteservice*. <https://github.com/GIScience/openrouteservice> Version 9.6.0. Accessed: 2026-01-11.
- [23] Birgit Klein, Rita Seiffert, Ulf Gräwe, Holger Klein, Peter Loewe, Jens Möller, Sylvain Müller-Navarra, Jürgen Holfort, and Christian Schlamkow. 2018. *Deutsche Bucht mit Tideelbe und Lübecker Bucht*. In *Hamburger Klimabericht – Wissen über Klima, Klimawandel und Auswirkungen in Hamburg und Norddeutschland*, Hans von Storch, Insa Meinke, and Martin Claußen (Eds.). Springer Spektrum, Berlin, Heidelberg, 55–87. doi:10.1007/978-3-662-55379-4_4
- [24] Çağrı Koç, Tolga Bektaş, Ola Jabali, and Gilbert Laporte. 2016. Thirty Years of Heterogeneous Vehicle Routing. *European Journal of Operational Research* 249, 1 (2016), 1–21. doi:10.1016/j.ejor.2015.07.020
- [25] Noureddine Labadi, Christian Prins, and Mustapha Reghioui. 2008. A memetic algorithm for the vehicle routing problem with time windows. *RAIRO - Operations Research* 42, 3 (2008), 415–431. doi:10.1051/ro/2008021
- [26] Michael K. Lindell and Carla S. Prater. 2007. Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning. *Journal of Urban Planning and Development* 133, 1 (2007), 18–29. doi:10.1061/(ASCE)0733-9488(2007)133:1(18)
- [27] Felix Mauch. 2012. *The Great Flood of 1962 in Hamburg*. *Arcadia* 6 (2012). doi:10.5282/rcc/3733 Rachel Carson Center for Environment and Society. Accessed: 2026-01-10.
- [28] Weiyao Meng, Rong Qu, and Nelishia Pillay. 2024. Machine Learning for Evolutionary Computation - the Vehicle Routing Problems Competition. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, New York, NY, USA, 13–14. doi:10.1145/3638530.3664046
- [29] Mercedes-Benz Buses. 2025. *CapaCity L. Technical data and equipment at a glance*. <https://www.mercedes-benz-bus.com/content/dam/mb/models/capacity/facts-capacity/facts-capacity-int-en.pdf> Accessed: 2025-11-20.
- [30] Mercedes-Benz Buses. 2025. *THE CITARO. Technical data and equipment at a glance*. <https://www.mercedes-benz-bus.com/content/dam/mb/models/citaro/facts-citaro/facts-citaro-int-en.pdf> Accessed: 2025-11-15.
- [31] Lale Özdamar, Erhan Ekinici, and Berrin Küçükyazici. 2004. Emergency logistics planning in natural disasters. *Annals of Operations Research* 129, 1 (2004), 217–245. doi:10.1023/B:ANOR.0000030690.27939.39
- [32] David Pisinger and Stefan Ropke. 2019. Large neighborhood search. In *Handbook of Metaheuristics*, Michel Gendreau and Jean-Yves Potvin (Eds.). Springer, 99–127. doi:10.1007/978-3-319-91086-4_4
- [33] Stefan Ropke and David Pisinger. 2006. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. *Transportation Science* 40, 4 (2006), 455–472. doi:10.1287/trsc.1050.0135
- [34] Shahrooz Shahparvari, Prem Chhetri, Babak Abbasi, and Ahmad Abareshi. 2016. Enhancing Emergency Evacuation Response of Late Evacuees: Revisiting the Case of Australian Black Saturday Bushfire. *Transportation Research Part E: Logistics and Transportation Review* 93 (2016), 148–176. doi:10.1016/j.tre.2016.05.010
- [35] Marius M. Solomon. 1987. Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. *Operations Research* 35, 2 (1987), 254–265. doi:10.1287/opre.35.2.254
- [36] Eduardo Uchoa, Diego Pecin, Artur Pessoa, Marcus Poggi, Anand Subramanian, and Thibaut Vidal. 2017. New benchmark instances for the Capacitated Vehicle Routing Problem. *European Journal of Operational Research* 257, 3 (2017), 845–858. doi:10.1016/j.ejor.2016.08.012
- [37] Eduardo Uchoa, Thibaut Vidal, Diego Pecin, Artur Pessoa, Marcus Poggi, and Anand Subramanian. 2022. The 12th DIMACS Implementation Challenge: Vehicle Routing Problems. DIMACS Center for Discrete Mathematics and Theoretical Computer Science. <http://dimacs.rutgers.edu/programs/challenge/vrp/> Accessed: 2026-01-10.
- [38] Thibaut Vidal, Teodor Gabriel Crainic, Michel Gendreau, Nadia Lahrichi, and Walter Rei. 2012. A hybrid genetic algorithm for multidepot and periodic vehicle routing problems. *Operations Research* 60, 3 (2012), 611–624. doi:10.1287/opre.1120.1048
- [39] Stefan Voigt. 2025. A review and ranking of operators in adaptive large neighborhood search for vehicle routing problems. *European Journal of Operational Research* 322, 2 (2025), 357–375. doi:10.1016/j.ejor.2024.05.033
- [40] David H. Wolpert and William G. Macready. 1997. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1, 1 (1997), 67–82. doi:10.1109/4235.585893