Optimizing a modular autonomous vehicle hub-and-spoke public transportation system Routing, scheduling, and repositioning

Wang, Zhimian; An, Kun; de Almeida Correia, Gonçalo Homem

DOI
10.1016/j.tre.2025.104251

Publication date
2025

Document Version
Final published version

Published in
Transportation Research Part E: Logistics and Transportation Review

Transportation Research Part E 201 (2025) 104251

Contents lists available at ScienceDirect

Transportation Research Part E

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

Optimizing a modular autonomous vehicle hub-and-spoke public transportation system: routing, scheduling, and repositioning

Zhimian Wang [a], Kun An [a,*], Gonçalo Homem de Almeida Correia [b]

[a] The Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China
[b] Department of Transport & Planning, Delft University of Technology, The Netherlands

ARTICLE INFO

Keywords:
Modular autonomous vehicles
Hub-and-spoke transportation system
Vehicle routing
Vehicle scheduling and repositioning

ABSTRACT

Modular autonomous vehicle (MAV), as a novel mode of public transportation, is anticipated to reshape the next generation of transportation systems, with the potential to adapt to diverse travel demand patterns in urban areas. In this study, we consider a MAV hub-and-spoke public transportation system (MAV-HSPTS). Each MAV can operate independently for first-mile/last-mile transport of passengers. Multiple MAVs can assemble into a modular bus, operating synchronously on mainline corridors with predetermined routes, stations, and timetables. We formulate a MAV routing, scheduling and repositioning model in a rolling horizon framework to capture the operation of the system. A heuristic algorithm that assigns passenger requests to MAVs is developed to reduce the computation time. The model and solution algorithm are evaluated on a bus transit corridor in Shanghai, China. Results demonstrate that the MAV service can reduce passenger travel time by over 20 % compared to conventional bus service, and over 90 % of the passengers could benefit from the convenience of in-bus transfers. MAV reposition proves to be an effective method to reduce the operational costs, especially in scenarios with imbalanced demand distribution.

1. Introduction

Mass transit systems are widely recognized as effective solutions to alleviate traffic congestion in large cities. However, traditional mass transit options, such as buses or bus rapid transit systems, can be prone to challenges such as long waiting times (Chen et al., 2016; Paudel et al., 2021), inconvenient transfers (Tyrinopoulos and Antoniou, 2008), and limited accessibility due to the first-mile and last-mile problem. These limitations are largely a result of their schedule-based operations and lack of door-to-door service, which can make it difficult to meet the dynamic and high-quality requirements of the travelers (Guo et al., 2017; Chen et al., 2018; Ma et al., 2021; Pei et al., 2019a).

In recent years, the emergence of modular autonomous vehicle (MAV) technology promises to enable dynamic capacity adjustments in public transport supply through the direct physical docking and undocking of carriages with vehicles of varying sizes during the journey. This innovative transportation mode has been developed by NEXT Future Transportation (as shown in Fig. 1) and tested in Dubai (Spera, 2016) and Ohmio LIFT (Ohmio, 2018), with the goal to offer high-quality travel services for passengers (Wu et al., 2021; Tian et al., 2022) and contribute to tackling various challenges associated with urban public transportation systems (Chen et al., 2021; Pei et al., 2021). Several prior studies have explored integrating MAVs with mass transit systems, positioning them as a new type of bus

* Corresponding author.
E-mail addresses: 2310819@tongji.edu.cn (Z. Wang), kunan@tongji.edu.cn (K. An), g.correia@tudelft.nl (G.H. de Almeida Correia).

https://doi.org/10.1016/j.tre.2025.104251
Received 25 January 2025; Received in revised form 22 April 2025; Accepted 8 June 2025
Available online 17 June 2025

capable of flexibly adjusting vehicle capacity either at terminal stations or en-route (Chen et al., 2019; Chen et al., 2020; Shi et al., 2020; Tian et al., 2023), a complement to conventional transit by providing first-mile and last-mile services (Nourbakhsh and Ouyang, 2012; Kim & Schonfeld, 2013), or enabling them to operate seamlessly across both first/last-mile and mainline routes with en-route transfers (Wu et al., 2021; Caros and Chow, 2021; Tang et al., 2023), thereby demonstrating promise for establishing adaptive public transportation systems.

In this study, we consider a novel public transportation system that integrates the flexible features of MAVs with the traditional structure of a bus transit system, aiming to provide passengers with hub-and-spoke services (as illustrated in Fig. 2**)**. The MAV hub-and-spoke public transportation system (MAV-HSPTS) proposed in this paper is inspired by the works of Auad-Perez and Van Hentenryck (2022), who explored the design of On-Demand Multimodal Transit Systems, combining fixed bus and rail routes between transit hubs with on-demand shuttles for first/last mile connectivity. The MAV-HSPTS described in this paper comprises a mass transit corridor with a series of MAV stations, each surrounded by multiple passenger stops. A MAV station serves both for MAV temporary parking and as a transfer hub for MAV docking and undocking operations, whereas a passenger stop is only designated for MAVs to pick up and drop off passengers. An individual MAV within this system function similarly to autonomous taxis or shared autonomous vehicles. These MAVs are initially parked at the stations. An MAV responds to passenger requests by departing from the station, stopping at designated pick-up points to pool multiple requests, and returning to the station. This process is referred to as the origin-to-station trip. For the mainline corridor operation, it can be considered as a virtual towing vehicle travelling along the mainline corridor, with predefined routes and schedules, similar to the framework of a conventional fixed-route bus transit system. The virtual towing vehicle together with its towed assembled MAVs, are collectively referred to as modular buses. At stations, MAVs completing origin-to-station trips can dock with a towing vehicle, allowing synchronized operation for the station-to-station trip. The docking operation is restricted to specific time slots, aligning with the towing vehicle's arrival at MAV stations according to the designated timetable. This coordination makes use of an intermittent dedicated bus lane available to the modular bus, which enhances the operational efficiency and energy consumption of modular bus while minimizing the negative impact on other conventional vehicles. MAVs can also disassemble from the bus at stations, either becoming idle at the station or starting a new station-to-destination trip. A MAV is also allowed to pool passengers from multiple requests during station-to-destination trips. Passengers carpooling for a station-to-destination trip can be relocated to another MAV, which is referred to as in-bus transfer. This MAV later disassembles from the bus and transports these passengers to their respective destination stops.

We address a dynamic routing, scheduling, and repositioning problem of MAVs in the hub-and-spoke public transportation system. The problem presents two key challenges. First, each MAV operates independently during origin-to-station (first-mile) and station-to-destination (last-mile) trips, requiring detailed planning of pick-up/drop-off routes, while multiple MAVs are planned collectively during docking and undocking operations to form a modular bus. Second, there are two distinct matching processes for each passenger request, respectively for origin-to-station trip and station-to-destination trip, which significantly increases the complexity of modeling the relationship between requests and MAVs.

To tackle these challenges, our approach leverages a rolling horizon optimization strategy that makes decisions at discrete time intervals, incorporating both current demand data and predicted future demand within a finite horizon. Based on the updated vehicle and demand status, MAV-to-request matching and routing decisions are continuously optimized through an integer programming model, with the objective of fulfilling passenger demand while minimizing operational costs. Leveraging predicted future demand, an empty repositioning strategy is incorporated to mitigate supply–demand imbalances across different regions, and address the inherent lack of foresight in dynamic scheduling. To ensure efficiency in handling dynamic scheduling operations, we first linearize the model and then develop a heuristic algorithm that reformulates the problem as a simple assignment model. Numerical results demonstrate that the proposed solution method can solve the problem within 10 s, ensuring real-world applications.

The paper is structured as follows. Section 2 reviews the literature on MAVs. Section 3 presents a MAV routing, scheduling and repositioning model within the rolling horizon framework. In Section 4, a linearization approach and a request-to-MAV assignment algorithm are combined to solve the model efficiently. Thereafter we test the performance of the method in the case study of a bus

Fig. 1. Modular autonomous vehicle. (https://www.next-future-mobility.com/).

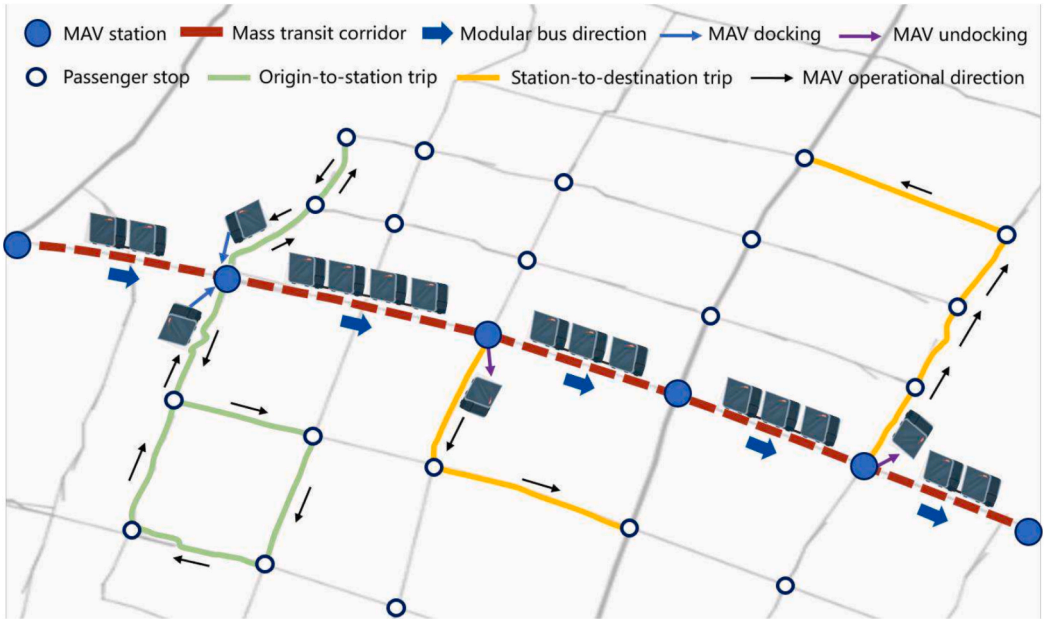

Fig. 2. MAV hub-and-spoke public transportation system.

transit corridor of Pudong District, Shanghai in Section 5, and conclusions follow in Section 6.

2. Literature review

2.1. MAV based fixed-route transit systems

In a fixed route transit system, MAVs can operate on a fixed bus line with predetermined bus stops. Through assembly and disassembly operations at stops, MAVs operate similar to a regular bus with flexible vehicle capacity.

The joint optimization of dispatch headways and vehicle capacity has been a key strategy for balancing vehicle operation costs and passenger travel time in these MAV-based public transit systems. Chen et al. (2019, 2020) proposed both discrete and continuous models to jointly design the dispatch headway and vehicle capacity in a MAV shuttle system under oversaturated traffic conditions. Shi et al. (2020) introduced an operational framework for shared corridors, aiming to minimize passenger waiting time costs and vehicle operation costs while optimizing dispatch headway and vehicle capacity. Chen et al. (2021, 2022) extended the corridor system for MAVs by introducing dynamic vehicle capacity adjustments through docking and undocking operation at intermediate stations. Shi et al. (2021) developed an operational design model for a variable-capacity MAV transit corridor, which groups passengers with the same destination arriving at a station in the same time interval, and allocates boarding passengers proportionally to the group size to mitigate the Oversaturated Queuing Phenomenon. Pei et al. (2021) proposed a modular transit network system where MAVs can be reassembled at stations based on downstream travel demand.

In recent years, a few studies have further addressed the MAV scheduling problem within the framework of fixed route transit, incorporating factors such as station location, fleet size, and vehicle repositioning. Tian et al. (2022) developed a planning method to determine optimal station locations and capacities to facilitate the assembly and disassembly of MAVs on a single transit line. Tian et al. (2023) extended their work by introducing a rolling horizon method to dynamically optimize MAV scheduling and capacity adjustments. Notably, their work explicitly incorporates MAV availability at stations and pioneers a repositioning strategy for the MAV-based transit system. Liu et al. (2023) proposed a joint optimization model for a bidirectional bus line, integrating timetable design, bus formation, and vehicle scheduling. Zhang et al. (2024) proposed a novel MAV service model that combines enroute coupling–decoupling operations with a skip-stop strategy while optimizing the number of MAVs and headways to align supply with demand at the bus-stop level. Tian et al. (2025) proposed an innovative real-time operational framework for modular transit services, leveraging a rolling horizon control method and a learning-based optimization proxy to jointly optimize service schedules and vehicle formations.

However, in the fixed-route transit systems, passengers are assumed to board and alight at designated stations, requiring them to rely on alternative transportation options for first-mile and last-mile travel. This approach underutilizes the flexibility of MAVs, which could otherwise operate independently to serve demand beyond conventional bus routes.

2.2. MAV based hub-and-spoke transit systems

In the hub-and-spoke transportation system, vehicles operate independently to collect and distribute passengers while forming platoons for long-haul trips. It has been identified as one of the most effective strategies for maximizing consumer surplus (Fu and Chow, 2023; Pei et al., 2019b). To the best of our knowledge, only a few studies have explored the MAV systems with this feature. These studies assume that MAVs operate independently to facilitate first- and last-mile connections, while also integrating with mainline bus services through physical platoon assembly.

Zhang et al. (2020) introduced a modular transit system to overcome the first-and-last-mile problem of traditional buses. Their system consists of a modular bus, composed of a virtual towing vehicle and multiple trailer MAVs. While the trailer MAVs can operate independently on minor roads to provide door-to-door services, they must attach to the towing vehicle when traveling on major roads. Caros and Chow (2021) proposed a two-sided day-to-day learning framework to model the performance of an MAV-based mobility service with en-route passenger transfer capabilities. Wu et al. (2021) proposed a modular, adaptive, and autonomous transportation system composed entirely of modular vehicles to offer extensive door-to-door services in urban areas. However, their study primarily focused on MAV trajectory design and passenger transfers at signalized intersections, without explicitly addressing pick-up and drop-off operations. Liu et al. (2021) introduced an operational design for flex-route transit services. In this system, MAVs initially travel collectively along fixed routes as modular buses, then strategically detach at stations to serve nearby requests. Fu and Chow (2023) formulated a dial-a-ride problem for MAVs, incorporating vehicle platooning, request pickup and delivery, and passenger transfers within platoons. Zou et al. (2024) introduced a modular electrified transit system tailored for urban corridor areas, combining the advantages of bus rapid transit and demand-responsive transit.

2.3. Other MAV based innovative transit systems

Beyond balancing supply and demand and offering mobility services, MAVs have also been explored as potential solutions to tackle broader transportation challenges. For example, Hannoun and Menendez (2022) proposed a MAV-based emergency medical transit system, where MAVs couple and decouple to transfer patients from life support vehicles to medical transport units, enabling faster emergency response. Both Hatzenbühler et al. (2023) and Lin and Zhang (2024) explored the integration of passenger and freight transport through MAV platooning. Khan et al. (2023) introduced a bus bunching mitigation strategy using MAV technology, where a modular bus decouples into individual units to perform stop-skipping when facing long headways. Shi et al. (2024) proposed an MAV-based baggage transport system for airports, aiming to reduce greenhouse gas emissions. These studies leverage MAVs' operational flexibility, allowing each MAV to operate independently to complete assigned tasks. Meanwhile, assembling multiple MAVs offers additional advantages, such as platoon-based operations and the direct exchange of passengers or freight within the platoon, enabling seamless task reassignment.

2.4. Research gap summary

The existing research most relevant to our study are summarized in Table 1. Current investigations into MAV deployment within hub-and-spoke transportation systems typically focuses on passenger pickup, drop-off, and en-route transfers, formulating them as static optimization problems. These approaches solve for large request batches over extended periods, resulting in computationally intensive models that struggle to deliver high-quality solutions dynamically. Besides, most models utilize a simplified dummy depot, which can represent any arbitrary parking location in the network after vehicles complete service. A more realistic approach would involve designating MAV storage zones and incorporating proactive repositioning strategies to optimize fleet distribution in anticipation of future demand across different regions. Furthermore, passenger convenience related to en-route transfers is also often overlooked, highlighting the need for a comprehensive assessment of transfer strategies to improve travel comfort and efficiency.

While some recent studies have advanced dynamic MAV scheduling and repositioning through rolling horizon methods, their

Table 1
Summary of related literatures.

Paper	System topology	Model	Solution method	Pickup and drop-off	En-route transfer	Dynamic scheduling	Repositioning
Shi et al. (2021)	Corridor	MILP	Dynamic programming	–	–	–	–
Tian et al. (2023)	Corridor	MINLP	Two-step heuristic	–	–	√	√
Tian et al. (2025)	Corridor	MILP	Learning-based	–	–	√	–
Pei et al. (2021)	Network	MINLP	Solver	–	√	–	–
Liu et al. (2021)	Flex-route	MILP	Two-stage heuristic	–	√	–	–
Wu et al. (2021)	Hub-and-spoke	ILP	A*	–	√	–	–
Zhang et al. (2020)	Hub-and-spoke	ILP	Graph partitioning	√	√	–	–
Fu and Chow (2023)	Hub-and-spoke	MILP	Local search	√	√	–	–
Zou et al. (2024)	Hub-and-spoke	MILP	Bi-level heuristic	√	√	–	–
Our study	Hub-and-spoke	INLP	Dynamic assignment heuristic	√	√	√	√

approach has demonstrated effectiveness primarily in fixed-route transit systems, which assume passengers board and alight at designated stations without considering the pickup/drop-off operations or en-route transfers. They also require that MAVs can only reposition from downstream to upstream stations before subsequent bus arrivals, which may need adaptation to accommodate round-way operations of buses.

In this study, based on a hub-and-spoke public transportation system, we propose a novel dynamic optimization method for MAV routing, scheduling, and repositioning. Our approach leverages a rolling horizon framework to periodically reassess and adjust transit services in response to real-time variations in passenger demand, with decision points distributed across discretized timesteps to simultaneously fulfill current demand while proactively repositioning MAVs for predicted future demand within a finite horizon. Two innovative transfer modes are also incorporated into the proposed system, with a goal to enhance both service quality and operational efficiency. Recognizing the computational challenges inherent to dynamic scheduling problems, we develop a specialized heuristic algorithm capable of delivering near-optimal solutions while maintaining the computational efficiency required for real-world implementation.

3. Problem formulation

3.1. Problem description

The MAV-HSPTS consists of a transit corridor with a sequence of MAV stations $S = \{1, \cdots, s\}$ and a series of passenger stops $N = \{1, \cdots, n\}$ located around the corridor. There are two operational directions for the modular bus, each of which travels along the corridor with fixed timetables and can adjust its capacity by docking or undocking MAVs. For modelling convenience, the two directions of a station are treated as two different stations in the corridor.

Passengers should submit requests to the travel platform with their preferred pick-up stop and drop-off stop, and then the system provides feedback on the stations where they will dock with and undock from modular bus, based on the shortest path strategy. The travel service for a passenger is comprised of three sub-trips: the origin-to-station trip on a MAV, the station-to-station trip on a modular bus, and the station-to-destination trip on a MAV. Firstly, the passenger is picked up by a MAV at the designated pick-up stop and taken to the station where they will dock with modular bus, after completing the origin-to-station trip. The passenger remains onboard at the station waiting its MAV to dock together with the matched modular bus. Upon completing the station-to-station trip, the passenger can get transferred in two ways: either through the MAV which will detach from the modular bus (i.e., in-bus transfer), or by transferring to an idle MAV at the drop off station (i.e., station transfer, get off the modular bus first and board another MAV at the station). Finally, the passenger is dropped-off at the destination stop.

As illustrated in Fig. 3, a passenger P1 boards vehicle A at its origin stop, and then travels to Station 1. There, P1 waits until vehicle A gets docked with a modular bus. On the mainline, at Station 2, vehicle A gets assembled with vehicle B, which has already picked up passengers P2 and P3 from distinct origin stops. At Station 3, passenger P2 is relocated from vehicle B to vehicle A, allowing vehicle B to undock from the modular bus and transport passenger P3 to its destination stop. Finally, at Station 4, vehicle A separates from the mainline with passengers P1 and P2, transporting them to their respective destination stops.

The operation period T is discretized into $|T|$ timesteps. We optimize the decisions regarding MAV routing, scheduling, and repositioning strategies via a rolling horizon framework. At the beginning of each timestep, the MAV operator receives updated information on the locations and availability of modular buses, MAVs, and passenger requests. The passenger requests $r \in R$ can be divided into three categories: requests submitted to the platform, not matched with a MAV yet for the first origin-to-station trip, represented by set R_1; requests already matched with a MAV for the origin-to-station trip, but not matched with a MAV yet for the

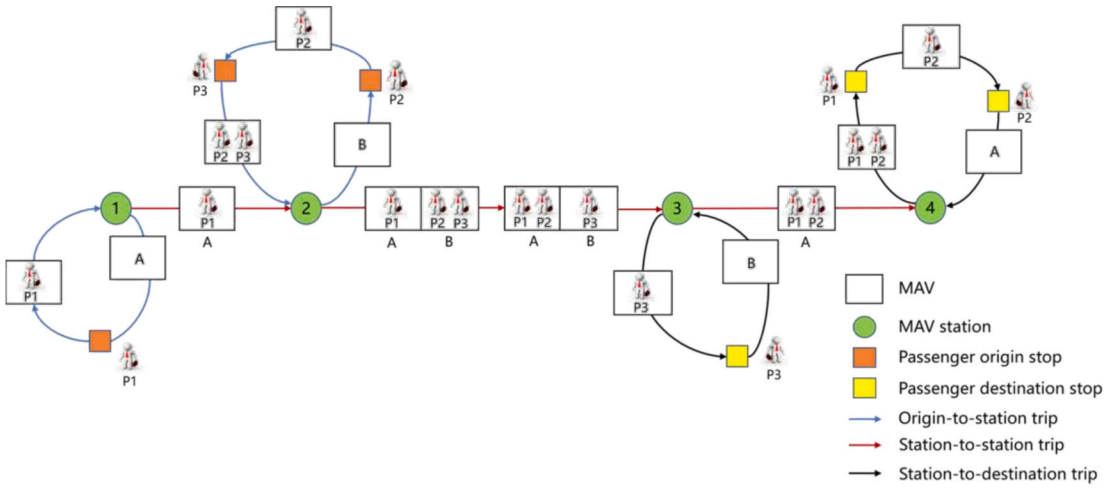

Fig. 3. Passenger service within the MAV-HSPTS.

subsequent station-to-destination trip, represented by set R_2; requests en-route of the station-to-destination trip, which have finished their whole matching process and should be removed from the system, represented by set R_3. The MAVs $u \in U$ can be divided into three sets: free MAVs positioned at the station, represented by set U_1; MAVs transporting passengers from origin to station or assembled in modular bus, represented by set U_2; MAVs on the way of being repositioned to the station or transporting passengers from station to destination, represented by set U_3.

As passenger requests and MAV operations evolve over multiple timesteps, this classification allows real-time updates to their status, ensuring adaptability without requiring a complete rescheduling for all MAVs. Additionally, categorizing requests and MAVs prevents incompatible assignments (e.g., assigning new requests to MAVs already in transit). By narrowing the decision-making scope, the computational complexity of the MAV routing, scheduling, and repositioning problem (MAV-RSRP) is reduced, enabling efficient operation while maintaining quick responsiveness.

Based on the optimization results, the status of MAVs and requests is updated to serve as inputs for the next time-step, as illustrated in Fig. 4. For the origin to station trips, the MAV-request matching and pick-up routes for origin-to-station trip are implemented directly, allowing real-time passenger service. As shown by the solid blue line in the figure, the matched requests transition from R_1 to R_2, while the MAVs serving origin-to-station trip move from U_1 to U_2. For the station-to-destination trips, the MAV-request matching result, and drop-off routes involving requests disembarking from modular bus within the current time-step, are implemented directly. These requests transition from R_2 to R_3, and the MAVs serving these trips either transition from U_1 to U_3 (dashed purple line, representing station transfer) or from U_2 to U_3 (dashed green line, representing in-bus transfer). Requests that remain on the modular bus by the end of the current time-step will stay in R_2, and may be matched again in subsequent time-steps. All the reposition operations are implemented directly, which make the MAVs transition from U_1 to U_3, and then return to U_1 after completing repositioning (solid orange line). Similarly, MAVs transporting passengers from the station to their destination are also classified as U_3, and revert to U_1 after dropping off all passengers (solid black line).

A few assumptions are made in the operations of the MAV-HSPTS.

Assumption 1 The dwell time of modular bus at each station that enables docking and undocking operation is a constant value.
Assumption 2. The headway of modular bus and the maximum allowable number of MAVs assembled per bus are appropriately configured to meet the level of passenger demand. In this assumption, we consider that only a limited number of MAVs can be coupled together in a bus trip due to safety concerns.
Assumption 3 The MAVs are homogeneous in terms of the number of seats and operation speed.
Assumption 4. Passengers' in-bus transfer can be realized en-route before arriving at the station.

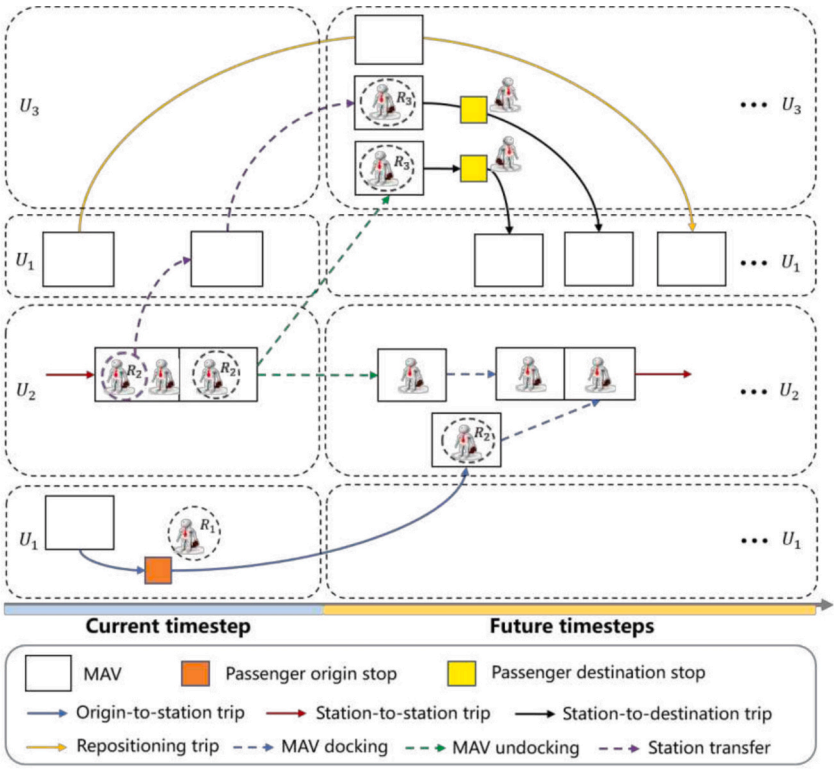

Fig. 4. Scheduling of MAVs.

Table 2
Notations.

Sets	
N	Set of passenger stops on the road network,$N = \{1, \cdots, n\}$
S	Set of MAV stations,$S = \{1, \cdots, s\}$
T	Set of time steps within an operation period,$T = \{1, \cdots, t\}$
T'	Set of time steps for predicted future demand,$T' = \{1, \cdots, t'\}$
B	Set of scheduled trips of modular bus, $B = \{1, \cdots, b\}$
U_1	Set of free MAVs at MAV stations,$U_1 = \{1, \cdots, u_1\}$
U_2	Set of MAVs transporting passengers from origin to station or assembled in modular bus,$U_2 = \{1, \cdots, u_2\}$
U_3	Set of MAVs on the way of repositioning to the station or transporting passengers from station to destination,$U_3 = \{1, \cdots, u_3\}$
R_1	Set of requests submitted to the platform, not matched with a MAV yet for the origin-to-station trip,$R_1 = \{1, \cdots, r_1\}$
R_2	Set of requests already matched with a MAV for the first origin-to-station trip, but not matched with a MAV yet for the subsequent station-to-destination trip,$R_2 = \{1, \cdots, r_2\}$
J_b^i	Set of stations that modular bus $b \in B$ has not yet visited, after having arrived at station i
$\widetilde{J}_b$	Set of stations that modular bus $b \in B$ is yet to visit at the beginning of the current time step
Parameters and Input	
Δt	Duration of a time step
τ_b^s	Index of time step when modular bus trip $b \in B$ arrives at station $s \in S$
ϕ_b^s	Total travel distance from the terminal station when modular bus trip $b \in B$ arrives at station $s \in S$
$\phi^{i,j}$	Distance between stop/station $i \in N \cup S$ and stop/station $j \in N \cup S$
CAP	Number of seats per MAV
p_u^s	Equals 1, if MAV $u \in U_1$ is initially parked at station $s \in S$; 0, otherwise
η_u^s	Equals 1, if MAV $u \in U_2$ is scheduled to dock or already docked with a modular bus at station s; 0, otherwise
$\widetilde{\vartheta}_{u,b}$	Equals 1, if MAV $u \in U_2$ arrives at the station earlier than modular bus trip $b \in B$, or is already docked with the bus trip b; 0, otherwise
$G_u^{s,t'}$	Equals 1, if MAV $u \in U_3$ is scheduled to arrive at station $s \in S$ within time step $t' \in T'$; 0, otherwise
q_r	Number of seats that request $r \in R_1 \cup R_2$ requires
o_r^n	Equals 1, if the origin of request $r \in R_1 \cup R_2$ is located at stop $n \in N$; 0, otherwise
d_r^n	Equals 1, if the destination of request $r \in R_1 \cup R_2$ is located at stop $n \in N$; 0, otherwise
a_r^s	Equals 1, if request $r \in R_1 \cup R_2$ requires docking with a modular bus at station $s \in S$; 0, otherwise
l_r^s	Equals 1, if request $r \in R_1 \cup R_2$ requires undocking from modular bus at station $s \in S$; 0, otherwise
$\rho_{u,r}$	Equals 1, if request $r \in R_2$ is served by MAV $u \in U_2$ during the origin-to-station trip; 0, otherwise
$A^{s,t'}$	Predicted demand within time step $t' \in T'$, which requires docking at station s
L_{max}	Maximum number of MAVs that can be assembled in a modular bus
v_u	Average speed of a MAV
v_b	Average speed of a modular bus
Γ_o^{max}	Upper bound of travel time for the origin-to-station trip
Γ_d^{max}	Upper bound of travel time for the station-to-destination trip
C_t	Unit transport cost of individual MAVs per km
C_r	Unit repositioning cost of individual MAVs per km
C_m	Unit mainline transport cost of assembled MAVs per km
ω_i	Cost of in-bus transfer per passenger
ω_s	Cost of station transfer per passenger
β	Penalty of rejecting a request
γ	Penalty of the insufficient number of MAVs for fulfilling the predicted demand in future time steps
M	A large positive number
Decision Variables	
χ_r	Equals 1, if request $r \in R_1$ is rejected in the current time step; 0, otherwise
y_u^s	Equals 1, if MAV $u \in U_1$ will be docking with a modular bus at station $s \in S$; 0, otherwise
$z_{u,r}$	Equals 1, if MAV $u \in U_1$ picks up request $r \in R_1$ during the origin-to-station trip; 0, otherwise
$x_u^{i,j}$	Equals 1, if MAV $u \in U_1$ consecutively traverses stop/station $i \in N \cup S$ and stop/station $j \in (N \cup S) \backslash \{i\}$ during the origin-to-station trip; 0, otherwise
$\vartheta_{u,b}$	Equals 1, if MAV $u \in U_2$ is scheduled to get docked with modular bus trip $b \in B$; 0, otherwise
$\theta_{r,b}$	Equals 1, if request $r \in R_2$ is scheduled to get aboard modular bus trip $b \in B$; 0, otherwise
$\widetilde{y}_u^s$	Equals 1, if MAV $u \in U_2$ will be undocking from the modular bus at station $s \in S$; 0, otherwise
$\widetilde{z}_{u,r}$	Equals 1, if MAV $u \in U_1 \cup U_2 \cup U_3$ drops off request $r \in R_2$ during the station-to-destination trip; 0, otherwise
$\widetilde{x}_u^{i,j}$	Equals 1, if MAV $u \in U_1 \cup U_2 \cup U_3$ consecutively traverses stop/station $i \in N \cup S$ and stop/station $j \in (N \cup S) \backslash \{i\}$ during the station-to-destination trip; 0, otherwise
$\varpi_u^{s,t'}$	Equals 1, if MAV $u \in U_1 \cup U_2$ is repositioned to arrive at station $s \in S$ within time step $t' \in T'$; 0, otherwise
Auxiliary Variables	
σ_u^n	Equals 1, if MAV $u \in U_1$ traverses stop $n \in N$ during the origin-to-station trip; 0, otherwise
$\widetilde{\sigma}_u^n$	Equals 1, if MAV $u \in U_1 \cup U_2 \cup U_3$ traverses stop $n \in N$ during the station-to-destination trip; 0, otherwise
δ_u	Equals 1, if MAV $u \in U_1 \cup U_2 \cup U_3$ is scheduled to drop off requests during the station-to-destination trip; 0, otherwise
μ_u^i	Auxiliary variables for avoiding sub-tours during origin-to-station trips
$\widetilde{\mu}_u^i$	Auxiliary variables for avoiding sub-tours during station-to-destination trips
$\psi^{s,t'}$	Insufficient number of MAVs at station $s \in S$ at future time step $t' \in T'$

3.2. Mathematical notation

Notations used throughout the paper are summarized in Table 2.

3.3. Model formulation

In this section, we formulate an integer nonlinear programming model to determine the optimal MAV routing, scheduling, and repositioning strategy within one time step.

3.3.1. MAV allocation constraints for the origin-to-station trip

To efficiently transport passengers from origins to the stations where they will dock with modular bus, we schedule free MAVs at stations with constraints formulated as follows.

Constraints (1) impose that every request $r \in R_1$ should be picked up by one available MAV $u \in U_1$, otherwise it will be rejected in the current time step (i.e., $\chi_r = 1$).

$$\sum_{u \in U_1} z_{u,r} = 1 - \chi_r \qquad \forall r \in R_1 \tag{1}$$

Constraints (2) and (3) ensure that each MAV docking with the modular bus should pick up at least one request, while restricted by the number of seats it supplies.

$$\sum_{r \in R_1} z_{u,r} \geq \sum_{s \in S} y_u^s \qquad \forall u \in U_1 \tag{2}$$

$$\sum_{r \in R_1} q_r z_{u,r} \leq CAP \qquad \forall u \in U_1 \tag{3}$$

Constraints (4) impose that a MAV, which is positioned at station $s \in S$ (i.e., $p_u^s = 1$) at the beginning of current time step, can only get docked with the modular bus at station s.

$$y_u^s \leq p_u^s \qquad \forall u \in U_1, \forall s \in S \tag{4}$$

Constraints (5) ensure that each MAV u can pick up request r in the origin-to-station trips only if they both get attached to the modular bus at station s. If there exists $s \in S$ such that $a_r^s = y_u^s = 1$, i.e., request r takes the bus trip at station s and MAV u docks on bus trip at station s, the constraint is equivalent to $z_{u,r} \leq 1$. Otherwise, it is equivalent to $z_{u,r} \leq 0$, prohibiting MAV u from picking up request r.

$$z_{u,r} \leq \sum_{s \in S} a_r^s y_u^s \qquad \forall u \in U_1, \forall r \in R_1 \tag{5}$$

3.3.2. MAV routing constraints for the origin-to-station trip

For the MAVs allocated to origin-to-station trips, we further construct routing constraints in order to pick up all the matched requests.

Constraints (6) ensure that the station should serve as both the start and end of the origin-to-station trip of MAVs in $u \in U_1$.

$$\sum_{i \in N} x_u^{i,s} = \sum_{j \in N} x_u^{s,j} = y_u^s \qquad \forall u \in U_1, \forall s \in S \tag{6}$$

Constraints (7) to (8) impose that MAV $u \in U_1$ should traverse all the origins of the requests it has been scheduled to pick up. In other words, stop $n \in N$ where request $r \in R_1$ is picked up by MAV u (i.e., stop n fulfilling $\sigma_u^n = 1$), should be one of the intermediate stops during the origin-to-station trip of MAV u.

$$\sigma_u^n \leq \sum_{r \in R_1} o_r^n z_{u,r} \leq M \sigma_u^n \qquad \forall u \in U_1, \forall n \in N \tag{7}$$

$$\sum_{i \in N \cup S} x_u^{i,n} = \sum_{i \in N \cup S} x_u^{n,i} = \sigma_u^n \qquad \forall u \in U_1, \forall r \in R_1, \forall n \in N \tag{8}$$

Constraints (9) avoid the occurrence of sub-tours during the origin-to-station trip.

$$\mu_u^i - \mu_u^j + M x_u^{i,j} \leq M - 1 \qquad \forall u \in U_1, \forall i \in N, \forall j \in N \backslash \{i\} \tag{9}$$

The term M represents a large positive constant in both **Constraints (7) and (9)**.

Constraints (10) specifies that the travel time of any MAV during the origin-to-station trip cannot exceed the upper bound Γ_o^{max}, ensuring the duration for passengers from their origins to the station for docking within an acceptable range.

$$\sum_{i \in P \cup S} \sum_{j \in P \cup S} \frac{\phi^{i,j}}{v_u} x_u^{i,j} \leq \Gamma_o^{max} \qquad \forall u \in U_1 \tag{10}$$

3.3.3. MAV docking & undocking constraints for the station-to-station trips

We use the parameter $\widetilde{\vartheta}_{u,b}$ to define the feasible range of modular bus trips that each MAV $u \in U_2$ could dock with. If MAV u is

docked with the modular bus trip b at the beginning of current time step, then $\widetilde{\vartheta}_{u,b} = 1$ for b and $\widetilde{\vartheta}_{u,b'} = 0$ for any $b' \in B \backslash \{b\}$. For MAV u that has not yet docked with any modular bus, $\widetilde{\vartheta}_{u,b} = 1$ if it arrives at the station earlier than modular bus trip b; otherwise, $\widetilde{\vartheta}_{u,b} = 0$. **Constraints (11)** ensure that each MAV $u \in U_2$ can only dock with modular bus trips within the given feasible range. **Constraints (12)** specify that each MAV $u \in U_2$ should be scheduled to get docked with one and only one bus trip.

$$\vartheta_{u,b} \le \widetilde{\vartheta}_{u,b} \forall u \in U_2, \forall b \in B \tag{11}$$

$$\sum\nolimits_{b \in B} \vartheta_{u,b} = 1 \qquad \forall u \in U_2 \tag{12}$$

Constraints (13) impose that if request $r \in R_2$ has been served by MAV $u \in U_2$ during its origin-to-station trip, the request should board the modular bus trip b with which MAV u is scheduled to dock.

$$\theta_{r,b} = \sum\nolimits_{u \in U_2} \rho_{u,r} \vartheta_{u,b} \qquad \forall r \in R_2, \forall b \in B \tag{13}$$

Constraints (14) specify that the MAV $u \in U_2$ can only undock from the modular bus at one of the stations that the modular bus has not yet visited. **Constraints (15)** specify that if the MAV u dock with modular bus at a station, it must undock from modular bus at a subsequent station. **Constraints (16)** ensure that the station where the MAV u undock from modular bus is unique.

$$\sum\nolimits_{b \in B} \left(\vartheta_{u,b} \sum\nolimits_{s \in \widetilde{J}_b} \widetilde{y}_u^s \right) = 1 \qquad \forall u \in U_2 \tag{14}$$

$$\widetilde{y}_u^s \le \sum\nolimits_{b \in B} \left(\vartheta_{u,b} \sum\nolimits_{s' \in S - J_b^s} \eta_u^{s'} \right) \qquad \forall u \in U_2, \forall s \in S \tag{15}$$

$$\sum\nolimits_{s \in S} \widetilde{y}_u^s = 1 \qquad \forall u \in U_2 \tag{16}$$

Constraints (17) specify that for each trip of the modular bus b, the overall capacity of the assembled MAVs should be lower than the remaining passengers on the modular bus. This ensures that all the passengers can be carried by the assembled MAVs during the whole station-to-station trip.

$$\sum\nolimits_{r \in R_2} \left(q_r \bullet \theta_{r,b} \sum\nolimits_{s \in S - J_b^i} \left(\alpha_r^s - l_r^s \right) \right) \le \sum\nolimits_{u \in U_2} \left(CAP \bullet \vartheta_{u,b} \sum\nolimits_{s \in S - J_b^i} \left(\eta_u^s - \widetilde{y}_u^s \right) \right) \qquad \forall b \in B, \forall i \in \widetilde{J}_b \tag{17}$$

Lastly, in view of physical limitations and safety concerns, **Constraints (18)** are formulated to ensure that for each trip of the modular bus, the number of assembled MAVs should be no larger than L_{max}.

$$\sum\nolimits_{u \in U_2} \left(\vartheta_{u,b} \sum\nolimits_{s \in S - J_b^i} \left(\eta_u^s - \widetilde{y}_u^s \right) \right) \le L_{max} \qquad \forall b \in B, \forall i \in \widetilde{J}_b \tag{18}$$

3.3.4. Mav-passenger allocation constraints for the station-to-destination trip

There are two options available to facilitate the MAV-passenger relocation for the station-to-destination trip of request $r \in R_2$. On the one hand, passengers can select a MAV $u \in U_2$ through in-bus relocation. On the other hand, they can alight from the modular bus and transfer by boarding another MAV that has been at the station – either MAVs in U_1 located at the station, or MAVs in U_1 and U_3 that can be repositioned to the station in advance.

Constraints (19) specify that request $r \in R_2$, must be matched to one of three types of MAVs.

$$\sum\nolimits_{u \in U_1 \cup U_2 \cup U_3} \widetilde{z}_{u,r} = 1 \qquad \forall r \in R_2 \tag{19}$$

Constraints (20) specify that the MAV $u \in U_1$, which can be repositioned to station s before time step τ_b^s (i.e., when the modular bus trip b arrives at station s), can transport the passengers who are located in modular bus $b \in B$ and require undocking at station s. **Constraints (21)** impose that the MAV $u \in U_2$ undocking from the modular bus at station $s \in S$ can transport the passengers on the same bus trip to leave from station $s \in S$ to its destination stop. **Constraints (22)** specify that the MAV $u \in U_3$, which is scheduled to arrive at station $s \in S$ before time step τ_b^s, can transport the passengers who are initially located in modular bus $b \in B$ and require undocking at station s.

$$\widetilde{z}_{u,r} \le \sum\nolimits_{s \in S} \sum\nolimits_{b \in B} \left(l_r^s \theta_{r,b} \sum\nolimits_{t'=1}^{\tau_b^s} \varpi_u^{s,t'} \right) \qquad \forall r \in R_2, \forall u \in U_1 \tag{20}$$

$$\widetilde{z}_{u,r} \le \left(\sum\nolimits_{b\in B}\theta_{r,b}\vartheta_{u,b}\right)\left(\sum\nolimits_{s\in S}l_r^s\widetilde{y}_u^s\right) \qquad \forall r\in R_2, \forall u\in U_2 \tag{21}$$

$$\widetilde{z}_{u,r} \le \sum\nolimits_{s\in S}\sum\nolimits_{b\in B}\left(l_r^s\theta_{r,b}\sum\nolimits_{t'=1}^{\tau_b^s}G_u^{s,t'}\right) \qquad \forall r\in R_2, \forall u\in U_3 \tag{22}$$

Constraints (23) and (24) specify that each MAV can be matched with multiple requests for subsequent station-to-destination trip, while the number of passengers from these requests cannot exceed the number of seats it supplies.

$$\sum\nolimits_{r\in R_2}\widetilde{z}_{u,r} \ge \delta_u \qquad \forall u\in U_1\cup U_2\cup U_3 \tag{23}$$

$$\sum\nolimits_{r\in R_2}q_r\widetilde{z}_{u,r} \le CAP\bullet\delta_u \qquad \forall u\in U_1\cup U_2\cup U_3 \tag{24}$$

3.3.5. MAV routing constraints for the station-to-destination trip

For the MAVs allocated to station-to-destination trips, we further construct routing constraints for dropping off all the matched requests.

Constraints (25) to (28) ensure that a MAV should depart from one of the stations, if it has been scheduled to serve a station-to-destination trip (i.e., $\delta_u = 1$). For MAVs in U_2, this station should be where they undock from the modular bus. For MAVs in U_1 and U_3, this station can be either their current location or a future repositioning point.

$$\sum\nolimits_{s\in S}\sum\nolimits_{j\in N}\widetilde{x}_u^{s,j} = \delta_u \qquad \forall u\in U_1\cup U_2\cup U_3 \tag{25}$$

$$\sum\nolimits_{j\in N}\widetilde{x}_u^{s,j} \le \sum\nolimits_{t'\in T'}\varpi_u^{s,t'} \qquad \forall u\in U_1, \forall s\in S \tag{26}$$

$$\sum\nolimits_{j\in N}\widetilde{x}_u^{s,j} \le \widetilde{y}_u^s \qquad \forall u\in U_2, \forall s\in S \tag{27}$$

$$\sum\nolimits_{j\in N}\widetilde{x}_u^{s,j} \le \sum\nolimits_{t'\in T'}G_u^{s,t'} \qquad \forall u\in U_3, \forall s\in S \tag{28}$$

Similar to **Constraints (7) to (8)**, **Constraints (29) to (30)** impose that MAV $u\in U_1\cup U_2\cup U_3$ should traverse all the destinations of the requests during the station-to-destination trip. The stop $n\in N$ where request $r\in R_2$ is dropped off by MAV u (i.e., $d_r^n\widetilde{z}_{u,r} = 1$), should be one of the intermediate stops during station-to-destination trip.

$$\widetilde{\sigma}_u^n \le \sum\nolimits_{r\in R_2}d_r^n\widetilde{z}_{u,r} \le M\widetilde{\sigma}_u^n \qquad \forall u\in U_1\cup U_2\cup U_3, \forall n\in N \tag{29}$$

$$\sum\nolimits_{i\in N\cup S}\widetilde{x}_u^{i,n} = \sum\nolimits_{i\in N\cup S}\widetilde{x}_u^{n,i} = \widetilde{\sigma}_u^n \qquad \forall u\in U_1\cup U_2\cup U_3, \forall n\in N \tag{30}$$

Constraints (31) specify that at the end of the station-to-destination trip, each MAV should return to the station which it has departed from.

$$\sum\nolimits_{i\in N}\widetilde{x}_u^{i,s} = \sum\nolimits_{j\in N}\widetilde{x}_u^{s,j} \qquad \forall u\in U_1\cup U_2\cup U_3, \forall s\in S \tag{31}$$

Constraints (32) avoid the occurrence of sub-loop during station-to-destination trip.

$$\widetilde{\mu}_u^i - \widetilde{\mu}_u^j + M\widetilde{x}_u^{i,j} \le M-1 \qquad \forall u\in U_1\cup U_2\cup U_3, \forall i\in N, \forall j\in N\backslash\{i\} \tag{32}$$

Constraints (33) specify that the travel time of any MAV during station-to-destination trip cannot exceed the upper bound Γ_d^{max}, which ensures that the duration for passengers from the station for undocking to their destinations remains within an acceptable range.

$$\sum\nolimits_{i\in N\cup S}\sum\nolimits_{j\in N\cup S}\frac{\phi^{i,j}}{v_u}\widetilde{x}_u^{i,j} \le \Gamma_d^{max} \qquad \forall u\in U_1\cup U_2\cup U_3 \tag{33}$$

3.3.6. MAV repositioning constraints

In addition to scheduling MAVs for origin-to-station and station-to-destination trips, we also account for repositioning MAVs to stations in need, with constraints formulated as follows.

Constraints (34) specify that a MAV $u\in U_1$ can either dock with a modular bus, stay at its current station, or get repositioned to another station. **Constraints (35)** specify that a MAV $u\in U_2$ is either scheduled to transport passengers for a station-to-destination trip, or repositioned to fulfill potential future demand. **Constraints (36)** impose that a MAV $u\in U_2$ could be allocated to serve the future demand at station s, where it gets undocking from the modular bus.

$$\sum\nolimits_{s\in S}y_u^s + \sum\nolimits_{s\in S}\sum\nolimits_{t'\in T'}\varpi_u^{s,t'} = 1 \qquad \forall u\in U_1 \tag{34}$$

$$\sum_{s\in S}\sum_{t^{'}\in T^{'}}\varpi_u^{s,t^{'}}+\delta_u=1 \qquad \forall u\in U_2 \tag{35}$$

$$\sum_{s\in S}\sum_{t^{'}\in T^{'}}\varpi_u^{s,t^{'}}\le\widetilde{y}_u^s \qquad \forall u\in U_2,\forall s\in S \tag{36}$$

Further, we formulate **Constraints (37) to (38)** to determine whether the MAVs u could be repositioned to any specific station s within each time step $t^{'}\in T^{'}$ through variable $\varpi_u^{s,t}$

$$\sum_{i\in S}p_u^i\frac{\phi^{i,s}}{v_u}\le t^{'}\Delta t+M\left(1-\varpi_u^{s,t^{'}}\right) \qquad \forall u\in U_1,\forall s\in S,\forall t^{'}\in T^{'} \tag{37}$$

$$\sum_{b\in B}\vartheta_{u,b}\tau_b^s\le t^{'}\Delta t+M\left(1-\varpi_u^{s,t^{'}}\right) \qquad \forall u\in U_2,\forall s\in S,\forall t^{'}\in T^{'} \tag{38}$$

With the development of deep learning, statistical methods using historical demand data and other real-time information can make short-term travel origins and destinations predictions of requests. In the existing studies, the accuracy rate of OD prediction for passenger requests has reached 70 % (Qiu et al., 2019; Ke et al., 2021). As demand prediction is not the focus of this study, we do not discuss the details here, but only use the parameter $A^{s,t}$ to describe the deterministic future demand within time-step $t\in T^{'}$ at station s. Then, **Constraints (39)** are formulated to specify that the MAVs at any station $s\in S$ should adequately satisfy the potential demand for any time step $t^{'}\in T^{'}$. Otherwise, the insufficient number of MAVs $\psi^{s,t^{'}}$ will give rise to significantly large penalty costs in the objective function.

$$\begin{aligned}&\sum_{u\in U_1}CAP\left(\sum_{t=1}^{t^{'}}\varpi_u^{s,t}-\delta_u\right)+\sum_{u\in U_2}\left(CAP\sum_{t=1}^{t^{'}}\varpi_u^{s,t}\right)\\&+\sum_{u\in U_3}CAP\left(\sum_{t=1}^{t^{'}}G_u^{s,t}-\delta_u\right)-\sum_{t=1}^{t^{'}}A^{s,t}+\psi^{s,t^{'}}\ge 0 \qquad \forall s\in S,\forall t^{'}\in T^{'}\end{aligned} \tag{39}$$

3.3.7. Objective function

The objective of the problem is to minimize the total costs. We consider the MAVs' operational costs (W_o), passengers' transfer costs (W_t), and penalty costs (W_p).

The MAVs' operational costs are calculated by **Function (40)** which consists of the travel costs during origin-to-station trips, station-to-destination trips, station-to-station trips, and empty repositioning, respectively:

$$\begin{aligned}W_o=&C_t\left(\sum_{u\in U_1}\sum_{i\in N\cup S}\sum_{j\in N\cup S}\phi^{i,j}x_u^{i,j}+\sum_{u\in U_1\cup U_2\cup U_3}\sum_{i\in N\cup S}\sum_{j\in N\cup S}\phi^{i,j}\widetilde{x}_u^{i,j}\right)\\&+C_m\sum_{u\in U_2}\sum_{b\in B}\sum_{s\in S}\vartheta_{u,b}\phi_b^s\left(\widetilde{y}_u^s-\eta_u^s\right)+C_r\sum_{u\in U_1}\sum_{i\in S}\sum_{s\in S}\left(p_u^i\phi^{i,s}\sum_{t^{'}\in T^{'}}\varpi_u^{s,t^{'}}\right)\end{aligned} \tag{40}$$

The passenger transfer costs are calculated via **Function (41):**

$$W_t=\omega_i\sum_{r\in R_2}\sum_{u\in U_2}q_r\widetilde{z}_{u,r}+\omega_s\sum_{r\in R_2}\sum_{u\in U_1\cup U_3}q_r\widetilde{z}_{u,r} \tag{41}$$

The above function includes: (I) the number of passengers which accomplish in-bus transfers by relocating to another assembled MAVs, with a unit cost of ω_i; (II) the number of passengers which accomplish station transfers by relocating to MAVs at stations, with a unit cost of ω_s.

The penalty costs are calculated by **Function (42):**

$$W_p=\beta\sum_{r\in R_1}\chi_r+\gamma\sum_{s\in S}\sum_{t^{'}\in T^{'}}\psi^{s,t^{'}} \tag{42}$$

where β is set as a penalty for request rejection, and γ is a penalty for the insufficient number of MAVs to fulfill the potential demand in future time steps.

The model can be summarized as follows:

$$\min W=W_o+W_t+W_p \tag{43}$$

s.t Constraints (1) – (42).

4. Solution algorithm

4.1. Model linearization

The proposed model is a complex nonlinear programming model, with nonlinear terms in **Constraints (14), (17), and (18)**, which relate to MAV docking/undocking decisions, and in **Constraints (20) to (22)**, which govern MAV-passenger allocation for station-to-

destination trips. The two decision variables, $\vartheta_{u,b}$ and $\theta_{r,b}$, play a significant role in these constraints. The variable $\vartheta_{u,b}$ determines the matching between MAV $u \in U_2$ and bus trip b, while the variable $\theta_{r,b}$ determines the matching between request $r \in R_2$ and bus trip b in the station-to-station trips. The requests and MAVs should dock with the first arriving assembled bus trip. However, due to the **Constraints (18)** on the maximum number of MAVs that can assemble with a bus, some may not be able to dock with the earliest bus. Decisions regarding MAV docking strategies further impact subsequent undocking and matching decisions for station-to-destination trips, resulting in nonlinear constraints. In order to linearize the model, we introduce a bus-trip allocation method to relax the decisions related to MAV docking strategies.

In the bus-trip allocation method, MAVs and their matched requests for origin-to-station trips are initially assigned to the earliest

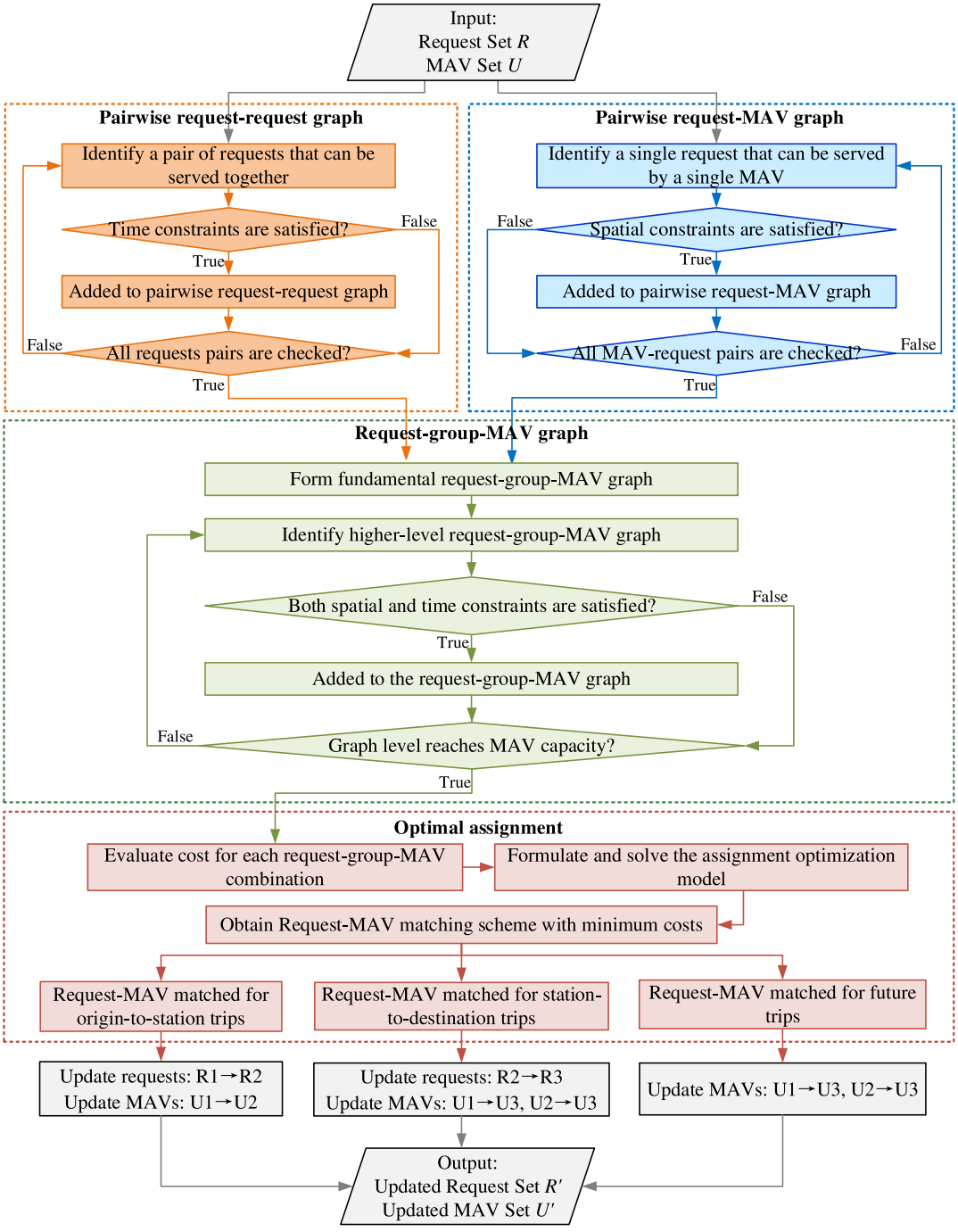

Fig. 5. Overall flowchart of the proposed request-to-MAV assignment algorithm.

modular bus trip based on their arrival time at the station. However, due to the limit on the number of MAVs that can assemble with a single bus trip, not every MAV can dock with its initially assigned trip. To address this challenge, we divide the set U_2 into two sub-sets: $\widehat{U}_2$, representing MAVs transporting passengers from origin to station, and $\check{U}_2$, representing those already assembled in the modular bus. At the beginning of each time step, with the modular bus trips listed in chronological order, we check the assigned MAVs in U_2. If the number exceeds the upper limit L_{max}, we remove MAVs from $\widehat{U}_2$ and reassign them to the next bus trip, until this count is exactly reduced to L_{max}. Accordingly, the requests matched with those MAVs for origin-to-station trip, are also reassigned to the later modular bus trip.

The method is executed before the optimization process of each time step. By using this method, the decision variables $\vartheta_{u,b}$ and $\theta_{r,b}$ can be fixed as constants, transforming the problem into a linear integer programming model.

4.2. Request-to-MAV assignment algorithm

A crucial challenge is multiple MAV-passenger matching in origin-to-station and station-to-destination trips. To improve solution efficiency, we develop a heuristic algorithm for assigning passenger requests to a fleet of MAVs. The overview of the proposed Request-to-MAV assignment algorithm is illustrated in Fig. 5.

The algorithm follows a four-step approach identified in the dash-lined boxes: First, two foundational graphs are constructed to capture spatiotemporal correlations − one for pairwise requests and the other for pairwise MAVs and requests. Then, based on these correlations, feasible request groups are formed, expanding the relationship into a request-group-MAV graph while ensuring compatibility among grouped requests. The Request-Group-MAV graph is expanded continuously until the request group size reaches the MAV's seating capacity. Finally, after evaluating the costs associated with each combination in the request-group-MAV graph, the problem is reformulated as a simple assignment model, where MAVs are optimally matched with request groups to minimize operation costs. The detailed steps are as follows:

4.2.1. Step 1: Pairwise request-request graph

In the first step, we identify which pairs of requests are potentially served by a single MAV and can be combined into pairwise request groups.

We consider any two requests $r_1, r_2 \in R_1$, which require docking with the modular bus at station $s \in S$. Let the path start from station s, traverse the origins of requests r_1, r_2, and eventually return to station s. If the path could fulfill the travel time constraint specified in **Constraints (10)**, the two requests are potentially combined during the origin-to-station trip, and $v(r_1, r_2)$ is added to the first-mile request-request-graph.

For any two requests $r_1, r_2 \in R_2$ initially located in a common modular bus, if both requests require undocking from the modular bus at station $s \in S$, a potential path is generated starting from station s and traversing the destinations of requests r_1, r_2. If the path can satisfy the travel time constraint specified in **Constraints (33)**, the requests are potentially combined during the station-to-destination trip, and $v(r_1, r_2)$ is added to the last-mile request-request-graph.

4.2.2. Step 2: Pairwise request-MAV graph

In the second step, we determine which MAVs can serve which requests individually.

Firstly, we consider every MAV $u \in U_1$ located at station $s \in S$. If a request $r \in R_1$ requires docking with a modular bus at station s, the combination $v(u, r)$ is added to the first-mile request-MAV graph. If a request $r \in R_2$ requires undocking with the modular bus at station s, the combination $v(u, r)$ is added to the last-mile request-MAV graph. Besides, if a MAV $u \in U_1$ can be repositioned to the station where request $r \in R_2$ requires undocking ahead of time, the combination $v(u, r)$ is also added to the last-mile request-MAV graph.

Then, every MAV $u \in U_2$ that gets docked with a modular bus at station $s \in S$ is considered. If a request $r \in R_2$ is located on the modular bus comprising of MAV u, while requiring undocking at a station visited after station s, the combination $v(u, r)$ is added to the last-mile request-MAV graph.

Lastly, we consider every MAV $u \in U_3$ scheduled to arrive at station $s \in S$. If a request $r \in R_2$ is expected to undock at station s later than the arrival time of the MAV, the combination $v(u, r)$ is added to the last-mile request-MAV graph.

4.2.3. Step 3: Request-group-MAV graph

The third step involves constructing groups of requests that can be served by a MAV. Based on the pairwise request-request graph outlined in Step 1 and the pairwise request-MAV graph introduced in Step 2, we initiate the first-mile/last-mile request-group-MAV graph, starting with groups containing only two requests. Subsequently, we construct higher-level graphs that integrate MAVs with groups comprising more than two requests, applying the following approach:

- **First-Mile Request-Group-MAV Graph:** For request group G and MAV u, if the MAV can serve all the requests of the group in a specific order during the origin-to-station trip to station s under the travel time constraint, they are potentially combined and added to $v(G, u, s)$, the first-mile request-group-MAV graph.
- **Last-Mile Request-Group-MAV Graph:** For request group G and MAV u, if the MAV can serve all the requests of the group in a specific order during the station-to-destination trip from station s, while adhering to the travel time constraint, they are potentially combined and added to $v(G, u, s)$, the last-mile request-group-MAV graph.

- **Future Request-Group-MAV Graph:** Since future demand has been grouped by time steps and stations, we can establish the future request-group-MAV graph directly. For each future demand group $\widetilde{G}$ at time step $t' \in T'$ and station $s \in S$, if a MAV $u \in U_1 \cup U_3$ can be repositioned to station s earlier than time step t', the combination $v\left(\widetilde{G}, u, s, t'\right)$ is added to the future request-group-MAV graph. Likewise, if the modular bus comprising of MAV $u \in U_2$ is scheduled to arrive at station s earlier than time step t', the combination $v\left(\widetilde{G}, u, s, t'\right)$ is added to the future request-group-MAV graph.

As there are abundant requests in need of allocation of MAVs for the station-to-destination trip, it takes a relatively long time for the construction of last-mile request-group-MAV graph. To improve computational efficiency, the matching scheme for requests already assigned to a MAV $u \in U_2$ through in-bus relocation in the previous time step can be reused. MAVs from U_2 start with an initial group consisting of the previously matched requests and can expand this group by incorporating additional unassigned requests.

4.2.4. Step 4: Optimal assignment

In this step, we first evaluate the costs associated with each combination in the three types of request-group-MAV graphs. Next, we reformulate a group-MAV assignment optimization model based on these costs. Finally, the optimal assignment solution is obtained using the Gurobi solver.

(1) Cost of combinations within first-mile request-group-MAV graph.

The cost of each request-group-MAV combination presented in the first-mile request-group-MAV graph is given as follows:

$$c_{uG} = C_t \phi_f(G, u, s) \qquad \forall u \in U_1, G \in \zeta_u^1 \tag{44}$$

where $\phi_f(G, u, s)$ denotes the first-mile travel distance associated with station s of MAV u in request group G, ζ_u^1 denotes the set of trips G for which an edge $v(G, u, s)$ exists in the first-mile request-group-MAV graph.

(2) Cost of combinations within last-mile request-group-MAV graph.

The cost of each request-group-MAV combination presented in the last-mile request-group-MAV graph is given as follows:

$$c_{uG} = C_t \phi_l(G, u, s) + W_o(G, u, s) + W_t(G, u, s) \qquad \forall u \in U_1 \cup U_2 \cup U_3, G \in \zeta_u^2 \tag{45}$$

where $\phi_l(G, u, s)$ denotes the last-mile travel distance associated with station s of MAV u in group G, ζ_u^2 denotes the set of trips G for which an edge $v(G, u, s)$ exists in the last-mile request-group-MAV graph.

$W_o(G, u, s)$ denotes the operational costs MAV u to move to station s, where travelers in group G can be served:

$$W_o(G, u, s) = \begin{cases} C_r \sum_{i \in S} p_u^i \phi^{i,s} & \forall u \in U_1, G \in \zeta_u^2 \\ C_m \sum_{b \in B} \left(\vartheta_{u,b} \phi_b^s - \sum_{s' \in S} \eta_u^{s'} \phi_b^{s'} \right) & \forall u \in U_2, G \in \zeta_u^2 \\ 0 & \forall u \in U_3, G \in \zeta_u^2 \end{cases} \tag{46}$$

$W_t(G, u, s)$ denotes the passengers' transfer costs that enables MAV u to serve group G at station s:

$$W_t(G, u, s) = \begin{cases} \omega_s \sum_{r \in G} q_r & \forall u \in U_1 \cup U_3, G \in \zeta_u^2 \\ \omega_i \sum_{r \in G} q_r & \forall u \in U_2, G \in \zeta_u^2 \end{cases} \tag{47}$$

(3) Cost of combinations within future request-group-MAV graph

The individual cost of each group-MAV combination presented in the future request-group-MAV graph is given as follows:

$$c_{u\widetilde{G}} = W_o\left(\widetilde{G}, u, s, t'\right) \qquad \forall u \in U_1 \cup U_2 \cup U_3, \widetilde{G} \in \zeta_u^3 \tag{48}$$

where ζ_u^3 denotes the set of groups $\widetilde{G}$ for which an edge $v\left(\widetilde{G}, u, s, t'\right)$ exists in the future request-group-MAV graph. $W_o\left(\widetilde{G}, u, s, t'\right)$ denotes the operational costs for transporting MAV u to station s, where group $\widetilde{G}$ can be served.

$$W_o\left(\widetilde{G}, u, s, t'\right) = \begin{cases} C_r \sum_{i \in S} p_u^i \phi^{i,s} & \forall u \in U_1, \widetilde{G} \in \zeta_u^3 \\ C_m \sum_{b \in B} \left(\vartheta_{u,b} \phi_b^s - \sum_{s' \in S} \eta_u^{s'} \phi_b^{s'} \right) & \forall u \in U_2, \widetilde{G} \in \zeta_u^3 \\ 0 & \forall u \in U_3, \widetilde{G} \in \zeta_u^3 \end{cases} \tag{49}$$

(4) Reformulation of programming model

Based on the costs evaluated upon various combinations, a group-MAV assignment optimization model is formulated with Objective **Function (56)** and **Constraints (44) to (55)**:

$$\min \sum_{u\in U_1}\left(\sum_{G\in\zeta_u^1} c_{uG}\epsilon_{uG}+\sum_{G\in\zeta_u^2} c_{uG}\epsilon_{uG}+\sum_{\tilde{G}\in\zeta_u^3} c_{u\tilde{G}}\epsilon_{u\tilde{G}}\right)+\sum_{u\in U_2\cup U_3}\left(\sum_{G\in\zeta_u^2} c_{uG}\epsilon_{uG}+\sum_{\tilde{G}\in\zeta_u^3} c_{u\tilde{G}}\epsilon_{u\tilde{G}}\right)\# \\ +\beta\sum_{r\in R_1}\chi_r+\gamma\sum_{s\in S}\sum_{t'\in T'}\psi^{s,t'} \tag{56}$$

s.t.

Constraints (44)-(49)

$$\sum_{G\in\zeta_u^1}\epsilon_{uG}+\sum_{G\in\zeta_u^2}\epsilon_{uG}+\sum_{\tilde{G}\in\zeta_u^3}\epsilon_{u\tilde{G}}=1 \qquad \forall u\in U_1 \tag{50}$$

$$\sum_{G\in\zeta_u^2}\epsilon_{uG}+\sum_{\tilde{G}\in\zeta_u^3}\epsilon_{u\tilde{G}}=1 \qquad \forall u\in U_2\cup U_3 \tag{51}$$

$$\sum_{G\in\zeta_r^1}\sum_{u\in\zeta_G^1}\epsilon_{uG}=1-\chi_r \qquad \forall r\in R_1 \tag{52}$$

$$\sum_{G\in\zeta_r^2}\sum_{u\in\zeta_G^2}\epsilon_{uG}=1 \qquad \forall r\in R_2 \tag{53}$$

$$\sum_{\tilde{G}\in\zeta_{s,t'}^3}\sum_{u\in\zeta_{\tilde{G}}^3} CAP\bullet\epsilon_{u\tilde{G}}-\sum_{t=1}^{t'}A^{s,t}+\psi^{s,t'}\geq 0 \qquad \forall s\in S,\forall t'\in T' \tag{54}$$

$$\sum_{s\in S-J_b^i}\sum_{r\in R_2}\left(q_r\bullet\theta_{r,b}\left(a_r^s-l_r^s\right)\right)\leq\sum_{s\in S-J_b^i}\sum_{u\in U_2}\left(CAP\bullet\vartheta_{u,b}\left(\eta_u^s-\sum_{G\in\zeta_{u,s}^2\cup\zeta_{u,s}^3}\epsilon_{uG}\right)\right) \\ \forall b\in B,\forall i\in\widetilde{J}_b \tag{55}$$

In the reformulated programming model, ϵ_{uG} is a binary variable that denotes whether MAV u is assigned to group G, ζ_r^1 denotes the set of groups that contains request $r\in R_1$, ζ_r^2 denotes the set of groups that contains request $r\in R_2$, ζ_G^1 denotes the set of MAVs u for which an edge $v(G,u)$ exists in the first-mile request-group-MAV graph, ζ_G^2 denotes the set of MAVs u for which an edge $v(G,u)$ exists in the last-mile request-group-MAV graph, ζ_G^3 denotes the set of MAVs u for which an edge $v\left(\widetilde{G},u\right)$ exists in the future request-group-MAV graph, $\zeta_{s,t'}^3$ denotes the set of future demand group at time step t' and station s, $\zeta_{u,s}^2$ denotes the set of groups at station s for which an edge $v(G,u)$ exists in the last-mile request-group-MAV-graph, $\zeta_{u,s}^3$ denote the set of groups at station s for which an edge $v\left(\widetilde{G},u\right)$ exists in the future request-group-MAV graph.

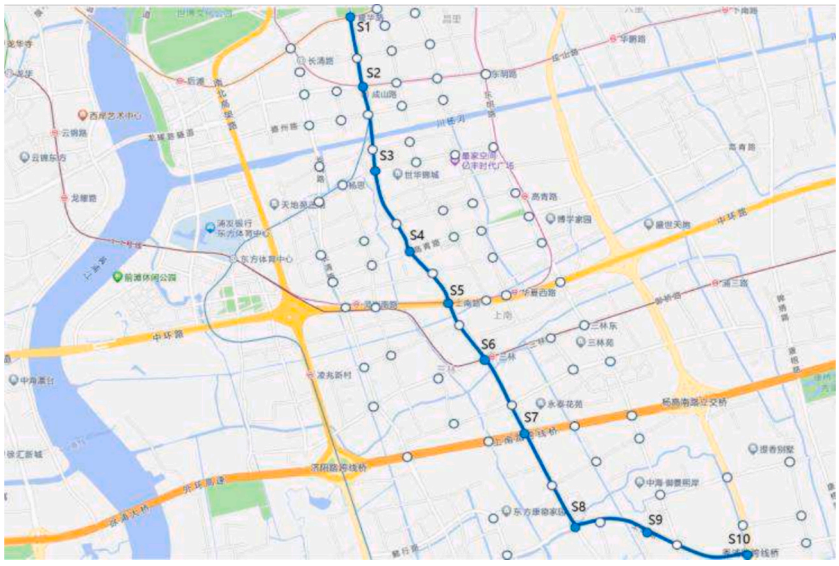

Fig. 6. Operational area of MAV-HSPTS in the case study.

Constraints (50) impose that a MAV $u \in U_1$ can be assigned to at most one group G, for which an edge $v(G,u)$ exists in the first-mile request-group-MAV graph, last-mile request-group-MAV graph, or future request-group-MAV graph. **Constraints (51)** impose that each MAV $u \in U_2$ should be assigned to a group G, for which an edge $v(G,u)$ exists in the last-mile request-group-MAV graph, or future request-group-MAV graph. **Constraints (52)** specify that each request $r \in R_1$ should be assigned to a MAV, otherwise it will be rejected. **Constraints (53)** ensure that the request $r \in R_2$ on the modular bus should be assigned to a MAV for the subsequent station-to-destination trip. **Constraints (54)** impose that the MAVs assigned to future request-group-MAV graph could fulfill the predicted future demand at each station. **Constraints (55)** ensure that all the passengers could be carried by the assembled MAVs during station-to-station trips.

The group-MAV assignment optimization model is a simplified formulation of the original problem that can be solved directly using the Gurobi solver.

5. Application to a real-world case study

In this section, we test the performance of the proposed MAV-HSPTS on a bus transit corridor in a real-world case.

5.1. Case set-up

The MAV-HSPTS is implemented along Shangnan Road, a main bus transit corridor located in the Pudong District of Shanghai, China. The corridor spans 8.6 km. As shown in Fig. 6, we designate 10 MAV stations (i.e., S1, S2, …, S10) along the corridor. Surrounding these stations are 73 passenger stops as passenger origins or destinations.

The modular bus is scheduled along two directions, respectively from Yaohua Road Subway Station to Xiupu Road Overpass Bridge and vice-versa, with an operation period of 1 h. The dispatch headway of the modular bus in both directions is set to be 5 min. The maximum number of MAVs assembled with a modular bus is set to be 8 (i.e., $L_{max} = 8$). The modular bus travels along the transit corridor at an average speed of $v_b = 30km/h$, with a fixed dwell time of 0.5 min at each station. The timetable of the modular bus is detailed in Table 3.

The MAVs are initially distributed uniformly across the 10 bus stations. Each MAV has 6 seats and travels at an average speed of $v_u = 25km/h$. The fixed costs associated with each MAV amount to ¥15 per hour. The choice of the duration of the time step to model the problem is subject to several considerations:

- A relatively long time step allows for more requests to be transferred from origins to stations through carpooling.
- A relatively short time step can decrease the response time to requests, while placing higher demands on the computing power of the operation system.
- The length of time step should be adequate for ensuring the timely decision-making on the allocation of MAVs for the station-to-destination transfer of each request.

Therefore, it is decided to partition the operation period into 20 time steps, each with a duration of 3 min. (i.e.,$\Delta t = 3min$). There are 20 requests generated randomly within each time step, with random origin stop and destination stop. The number of seats required by each request ranges from 1 to 3. The future demand at each station is predicted within the next 4 time steps (i.e., $|T| = 4\Delta t$). Both the upper bound of the travel time of each MAV during the origin-to-station trip and station-to-destination trip are set to be 5 time steps (i.e., $\Gamma_p^{max} = 5\Delta t, \Gamma_d^{max} = 5\Delta t$). The other parameters are given as $C_m =$ ¥0.6/km, $C_t =$ ¥1.2/km, $C_r =$ ¥1.5/km, $\omega_i =$ ¥0.1, $\omega_s =$ ¥0.3.

5.2. Optimization results

5.2.1. Optimization performance under varied number of MAVs

In this section, we conduct numerical experiments to evaluate the overall performance of the proposed MAV-HSPTS under different fleet sizes. To assess the effectiveness of the proposed heuristic algorithm, we perform a comparative analysis against the exact solution obtained using the Gurobi solver. The results, highlighting the trade-off between solution quality and computational efficiency, are presented in Fig. 7.

The proposed heuristic algorithm consistently achieves near-optimal objective values across all three fleet sizes, maintaining solution quality within 12 % of Gurobi's optimal values. As the fleet size increases, the gap between the heuristic and optimal solutions remains stable, demonstrating the scalability of the heuristic approach. The heuristic significantly outperforms the Gurobi solver in computation time. With the number of MAVs increasing from 140 to 180, Gurobi's average computation time rises from 62.82 s to

Table 3
Timetable of modular bus.

		Station									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Bus arrival time (min)	Downward direction	0	2.14	4.62	7.12	9.18	11.28	13.70	16.60	18.78	21.68
	Upward direction	21.68	19.54	17.06	14.56	12.50	10.40	7.98	5.08	2.90	0

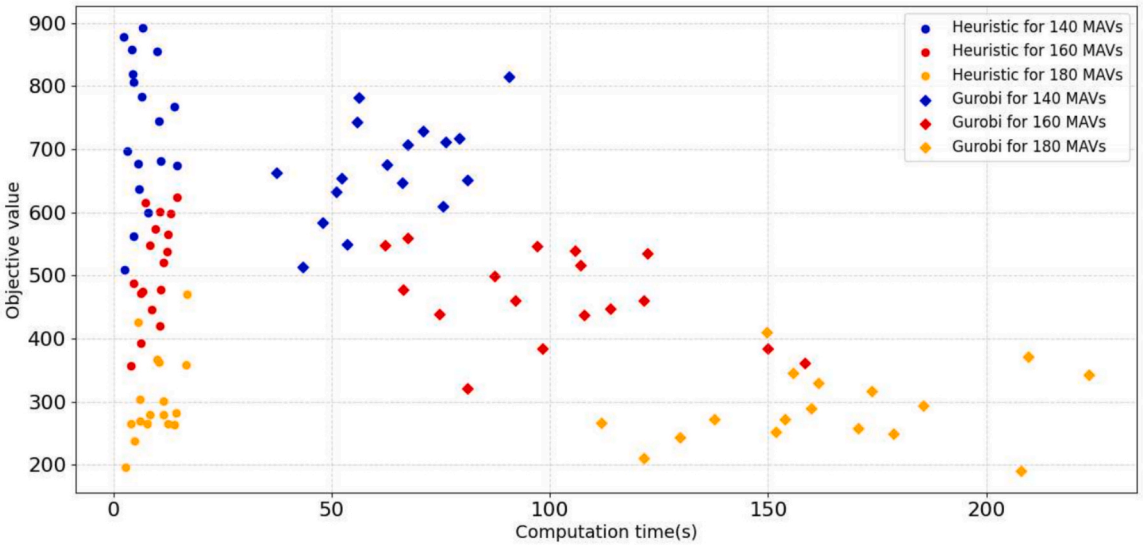

Fig. 7. Comparative analysis on solution performance.

163.69 s, whereas the heuristic's average computation time only increases from 6.98 s to 9.61 s. This efficiency enables the heuristic to solve the problem within 10 s, showcasing its capability to handle dynamic scheduling operations in a rolling horizon framework.

The optimization results on request rejection and total costs are summarized in Table 4. When there are only 140 MAVs in operation, most of the MAVs are allocated to passenger transport, with few allocated for empty repositioning. Consequently, there is an obvious imbalance between demand and supply and making the rejection rate of requests up to 13.97 %. As the total number of MAVs increases from 140 to 240, a greater number of MAVs are relocated from low-demand stations to high-demand stations. This shift results in an increase in repositioning costs from ¥124 to ¥210, reducing the rejection rate of requests to 0.68 %. Additionally, there is a significant increase in transport costs, attributable to the improved matching of requests with the expanded pool of available MAVs. Consequently, the total costs increase from ¥4491 to ¥6447. As the total number of MAVs increases from 240 to 280, there is now a surplus of MAVs capable of meeting the current demand level, leading to a reduction in repositioning costs from ¥210 to ¥154, while reducing the rejection rate of requests from 0.68 % to almost 0.

The performance of passenger service is illustrated in Fig. 8. The average travel times for origin-to-station and station-to-destination trips remain relatively stable, ranging from 2.90 to 3.37 min. Initially, they increase slightly with fleet expansion due to higher carpooling activity but eventually decreases as the number of MAVs continues to grow. For station-to-destination trips, both in-bus and station transfers increase as the total number of MAVs increases from 140 to 220, which can be attributed to the improvement in the rejection rate. However, as the number of MAVs increases from 220 to 280, in-bus transfers continue to rise, while station transfers decrease slightly. Under all the cases of our experiment, over 90 % of the passengers could benefit from the convenience of in-bus transfers.

In summary, adjusting the total number of MAVs in the operational area allows for a balanced consideration of transport costs, repositioning costs, and passenger transfer, leading to more efficient demand fulfillment. In this experiment, 240 MAVs strike an optimal balance, reducing the rejection rate to below 1 % with the acceptable total costs.

5.2.2. Comparison with conventional bus service

In this section, we conduct experiments to compare the performance of the MAV transit service with conventional bus service under varying levels of travel demand. The conventional bus service operates along the transit corridor at an average speed of $30km/h$, with a fixed dwell time of 0.75 min at each station. Each bus dispatch has 24 seats and incurs a transport cost of ¥6/km. Passenger requests are

Table 4
Request rejection rate and operational costs.

Number of MAVs	Rejection rate (%)	Number of trips satisfied	Transport costs (¥)	Repositioning costs (¥)	Fixed costs (¥)	Total costs (¥)
140	13.97	1032	2267	124	2100	4491
160	8.84	1094	2431	135	2400	4966
180	4.88	1141	2531	149	2700	5380
200	3.94	1153	2584	189	3000	5773
220	1.59	1181	2588	200	3300	6088
240	0.68	1192	2637	210	3600	6447
260	0.25	1197	2648	199	3900	6747
280	0	1200	2670	154	4200	7024

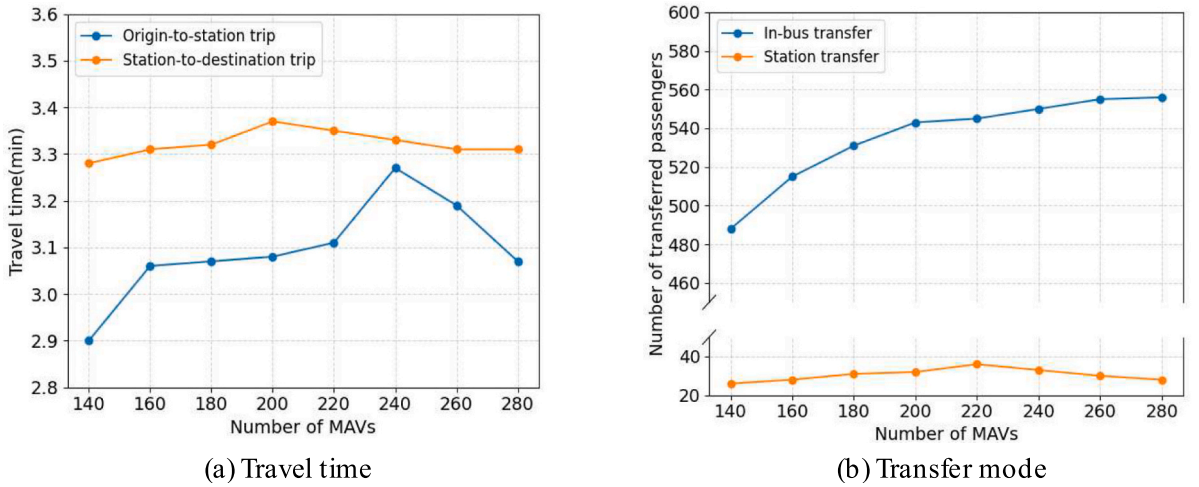

Fig. 8. Performance of passenger service.

still generated at the bus stops designated for the proposed MAV-HSPTS. However, in the conventional bus system, passengers must choose an alternative mode of transportation to complete the first/last-mile trips between their origin/destination stops and the corresponding bus stations. We assume that passengers travel their first/last-mile trips by bike at an average speed of 15km/h.

The arrival demand per hour ranges from 80 to 560, with the number of seats required by each request ranging from 1 to 3. Passenger benefits and system operational costs are evaluated for the two types of bus services, as shown in Table 5. For passenger benefits, we consider the duration of three sub-trips: the origin-to-station trip, the station-to-station trip, and the station-to-destination trip. The operational costs of the conventional bus system are limited to transport costs along the mainline corridor, while the operational costs of the MAV-HSPTS are divided into the transport costs for the three sub-trips and the additional repositioning costs.

The results indicate that compared to the conventional bus service, the modular bus service can reduce total passenger travel time by 16.94 % to 24.90 %. Additionally, compared to the conventional bus service with fixed operational costs, the operational costs of the MAV-HSPTS can be adjusted based on varying travel demand. As travel demand decreases to 80 requests per hour, operational costs are reduced to ¥637 due to the deployment of fewer MAVs, significantly lower than the costs of conventional bus services. This experiment highlights the advantages of MAV-HSPTS in providing efficient first/last-mile transfers and a flexible operational strategy that adapts to varying demand. These benefits are particularly pronounced in low-demand scenarios. However, to maintain reliable performance under higher travel demand, the system requires significantly increased operational costs to ensure sufficient MAVs for origin-to-station and station-to-destination trips, either through repositioning or fleet size expansion—almost 3 times the operational costs of the conventional bus system for the scenario of 560 requests per hour.

Therefore, the MAV-HSPTS might be particularly well-suited to function as a complementary service to traditional bus systems in low demand areas. In high-demand scenarios, MAVs should be primarily scheduled along the mainline corridor, striking a better balance between operational costs and passenger benefits.

Table 5
Comparison between conventional and modular bus system.

Arrival demand per hour	Bus system	Operational costs (¥)	Passenger travel time (min)				
			Origin-to-station trip	Station-to-station trip	Station-to-destination trip	Total	
80	Conventional	1127	4.52	7.79	4.56	16.87	↓24.01
	MAV-HSPTS	637	3.04	6.71	3.07	12.82	%
160	Conventional	1127	4.52	8.03	4.56	17.11	↓24.90
	MAV-HSPTS	1195	3.06	6.70	3.09	12.85	%
240	Conventional	1127	4.52	8.24	4.56	17.32	↓24.71
	MAV-HSPTS	1746	3.12	6.77	3.15	13.04	%
320	Conventional	1127	4.52	8.18	4.56	17.26	↓23.41
	MAV-HSPTS	2317	3.19	6.80	3.23	13.22	%
400	Conventional	1127	4.52	8.28	4.56	17.36	↓21.77
	MAV-HSPTS	2847	3.27	6.98	3.33	13.58	%
480	Conventional	1127	4.52	8.30	4.56	17.38	↓20.02
	MAV-HSPTS	3085	3.35	7.10	3.45	13.90	%
560	Conventional	1127	4.52	8.33	4.56	17.41	↓16.94
	MAV-HSPTS	3277	3.54	7.28	3.64	14.46	%

5.2.3. Sensitivity analysis with MAV capacity

We use the case of 240 MAVs, each with 6 seats, as the baseline case. The total number of seats was set to 1440. We varied the MAV capacity from 4 to 8, and adjusted the number of MAVs accordingly (i.e., 4 × 360, 5 × 288, 6 × 240, 7 × 206, and 8 × 180). In comparison to the 6-capacity scenario, we assume that an additional seat will incur an increase in fixed costs by ¥2 per hour for each MAV. The optimization results are illustrated in Fig. 9.

With an increase in the capacity of MAVs from 4 to 8, the total costs decrease from ¥6673 to ¥6121, while the rejection rate has a significant rise from 0 to 4.25 %. As more requests are accommodated to share a single MAV, the average travel time of origin-to-station trips increases from 3.19 min to 3.42 min, and the average travel time of station-to-destination trips increases from 3.21 min to 3.57 min.

The number of in-bus transfers and station transfers are also sensitive to the MAV's capacity. With the capacity increasing from 4 to 8, in-bus transfers decrease from 553 to 528, while station transfers increase from 22 to 40. As the request rejection rate remains relatively stable, it is evident that the reduction in in-bus transfers is mostly converted to an increase in station transfers. One reason is that a relatively large MAV capacity reduces the flexibility of constructing in-bus transfer groups. Obviously, grouping passengers in sets of 4 is much easier than in sets of 8. With the increase in MAV capacity, reflecting on the need of saving sufficient modular bus capacity for serving station-to-station trip, a larger proportion of passengers cannot be allocated to in-bus transfer groups and must adopt station transfers instead.

Therefore, deploying a relatively large number of MAVs with smaller capacities enhances the flexibility and convenience of the proposed MAV-HSPTS, improving real-time responsiveness to passenger requests and facilitating in-bus transfers. However, this approach also leads to higher costs due to the need to maintain sufficient MAVs in service.

5.2.4. Sensitivity analysis with the future demand prediction

Prediction accuracy and horizon length affect the effectiveness of MAV repositioning strategies. To systematically evaluate their impact, we analyze four prediction accuracy levels (40 %, 60 %, 80 %, and 100 %) across horizon lengths ranging from 2 to 6 time-

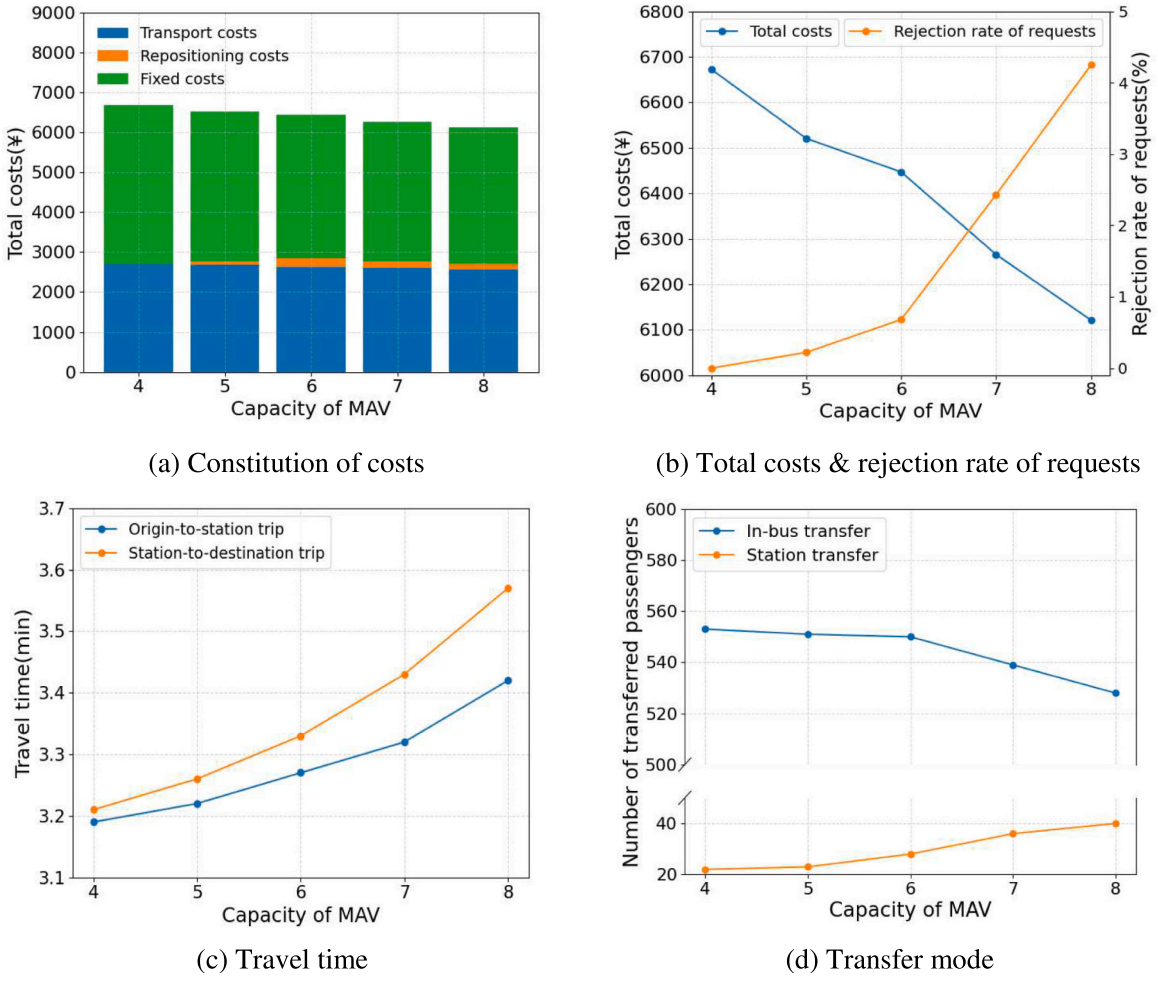

Fig. 9. Sensitivity test on capacity of MAV.

steps. We assess performance using two critical metrics: the ratio of repositioning costs to the total operation costs and the request rejection rate.

For instance, under the 80 % accuracy scenario, we assume 80 % of predicted requests within the horizon are accurate (i.e., will materialize as forecasted), while the remaining 20 % are uncertain. To model this uncertainty, we generate 10 potential demand scenarios for the uncertain requests. Each scenario undergoes independent optimization, yielding distinct repositioning strategies. These scenario-specific results are then averaged to represent the system performance metrics. The experimental results obtained through this methodology are presented in Fig. 10.

Low prediction accuracy (e.g., 40 %) leads to a substantially higher repositioning costs, as frequent and suboptimal adjustments are needed to compensate for forecasting errors. Rejection rates are also increased, as the system cannot adjust the repositioning strategies immediately to handle the unanticipated demand. In contrast, higher accuracy (80–100 %) enables a more efficient strategy, reducing both repositioning costs and request rejections by aligning MAV allocation more closely with actual demand.

Regarding the prediction horizon, while a longer horizon improves decision-making—lowering rejection rates by allowing more time to adapt to demand changes—it also increases repositioning costs due to proactive adjustments over an extended period. Notably, as the horizon extends, the performance gap between different accuracy levels narrows, suggesting that longer foresight can partially mitigate the drawbacks of lower prediction accuracy.

5.2.5. Sensitivity analysis with imbalanced demand distribution and repositioning operation

In this section, we focus on studying the repositioning operation across three cases with varied demand distributions:

Case 1: All the requests are randomly generated at the stops uniformly (50 %/50 %).

Case 2: 60 % of the requests require taking the modular bus in the upward direction, and 40 % of them require taking the modular bus in the downward direction.

Case 3: 70 % of the requests require taking the modular bus in the upward direction, and 30 % of them require taking the modular bus in the downward direction.

The number of MAVs assembled with the modular bus along the mainline corridor is depicted in Fig. 11. In Case 1, with uniform demand, the number of assembled MAVs for the modular bus traveling downstream initially increases across the four upstream stations, peaks at 7 MAVs between Stations 4 and 8, and then decreases across the three downstream stations. In Cases 2 and 3, the number of MAVs assembled in the downward direction gradually decreases, with a peak of only 5 MAVs assembled between Stations 5 and 6 in Case 3. Meanwhile, the upward direction sees a rapid increase in MAVs assembled with the modular bus, reaching the upper limit of 8 MAVs between Stations 6 and 8 in Case 3.

The number of repositioned MAVs at each station within 6-minute intervals is illustrated in Fig. 12, while Fig. 13 shows the time-dependent distribution of MAVs, representing the average number of MAVs at each station during those intervals. In Case 1, the repositioning of MAVs is relatively uniform across various stations. In Case 2 and Case 3, MAVs are predominantly repositioned from upstream to downstream stations, with this pattern being more pronounced in Case 3. Although the unbalanced demand distribution leads to a higher concentration of MAVs at upstream stations, the downstream stations are continuously replenished, ensuring a sufficient reserve to meet potential future demand. Therefore, repositioning operations are crucial for ensuring prompt service and maintaining the robustness of the MAV-HSPTS.

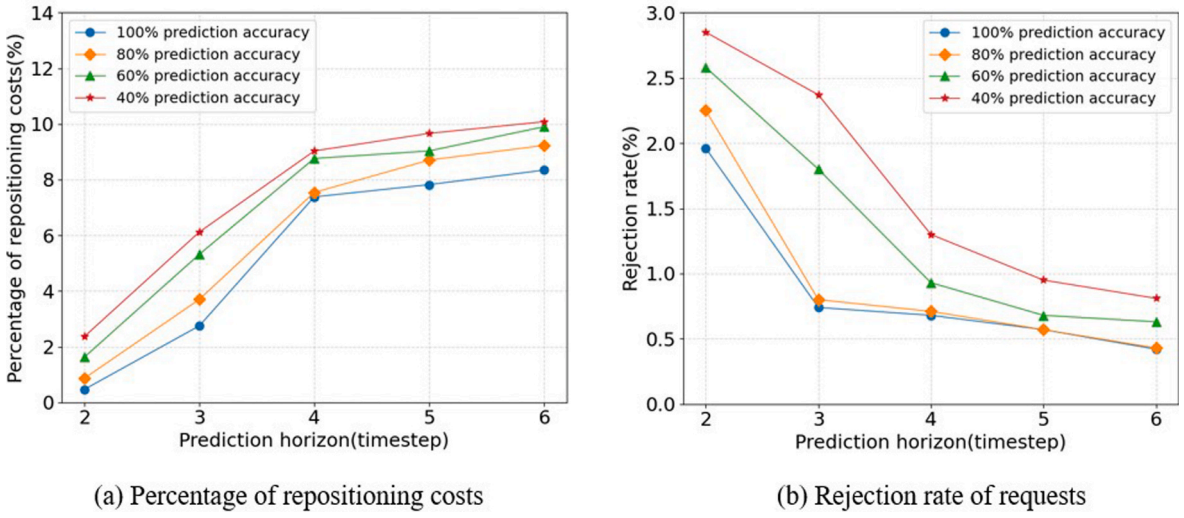

Fig. 10. Sensitivity test on future demand prediction.

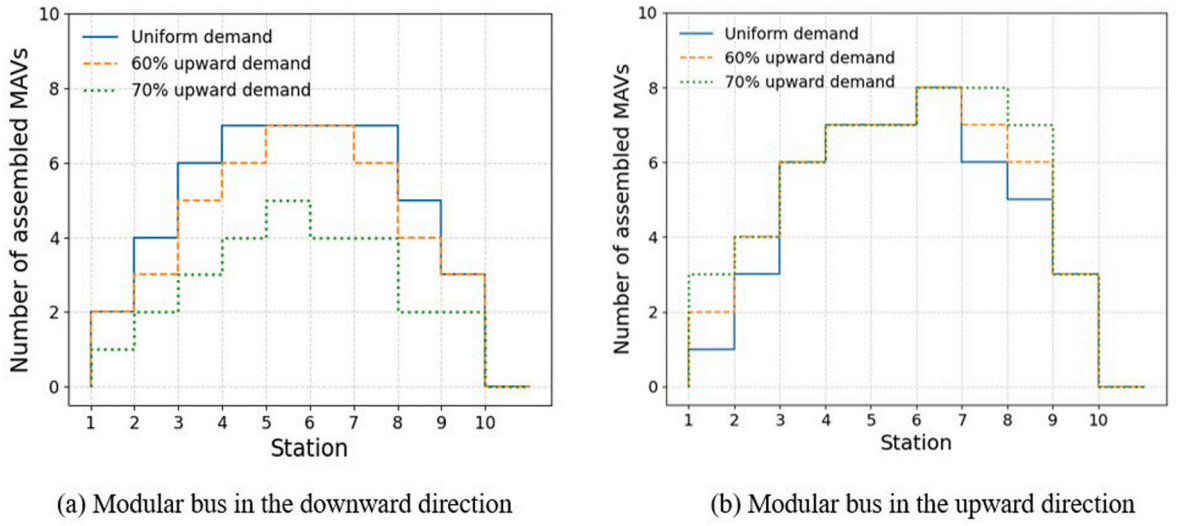

Fig. 11. Number of MAVs assembled in a modular bus.

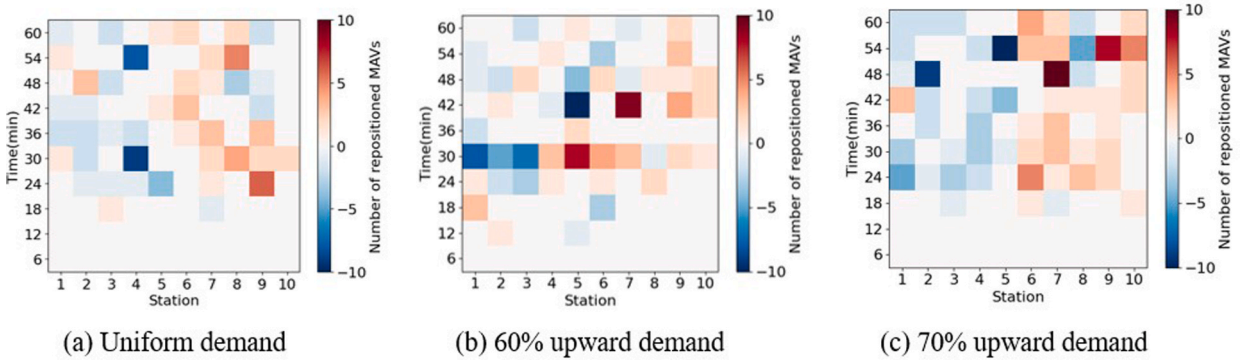

Fig. 12. Repositioning pattern of MAVs under varied demand distributions.

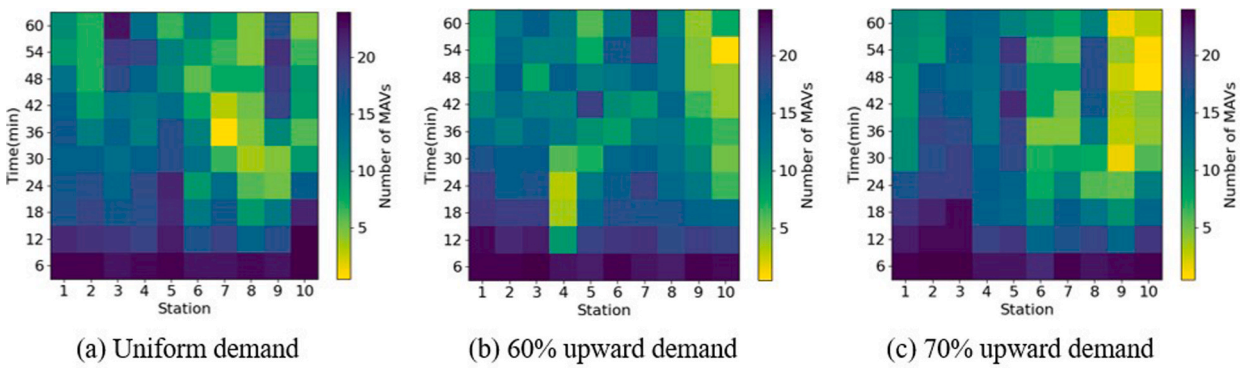

Fig. 13. Station distribution of MAVs under varied demand distributions.

6. Conclusions

With the diversified travel demand of urban residents over these years, modular autonomous vehicle is expected to serve as a complementary to the traditional mass transit system, and thus providing high-level public transportation services in the next generation of transit system. In this study, we conceptualize a door-to-door public transportation system based on modular autonomous vehicles (MAVs). Each MAV can be operated independently, catering to the first-mile/last-mile transport needs of passengers, while

multiple MAVs can be assembled into a modular bus, operating synchronously along the corridor of the road network with fixed routes, stations, and timetables. Based on a rolling horizon framework, we partitioned the operation period of the MAV-HSPTS into discrete time steps, and proposed an integer programming model that determined MAV-request matching schemes, MAV repositioning strategies, and pick-up/drop-off routes within each time step. A linearization approach and a heuristic algorithm that assigns passenger requests to MAVs has been developed to solve the model efficiently.

A case study was presented to demonstrate the performance of the proposed MAV-HSPTS. The results indicated that using the proposed optimization method, the MAV service can reduce total passenger travel time by over 20 % compared to conventional bus service, and over 90 % of the passengers could benefit from the convenience of in-bus transfers for last-mile trips. A relatively large number of MAVs with small capacities could enhance the flexibility of the system, showing superiority in both demand responsiveness and percentage of in-bus transfers. We also highlight the significant role of repositioning operations in balancing the supply of MAVs with passenger travel demands, especially in scenarios with unbalanced demand distribution among various MAV stations.

Several future research directions can build upon this work. First, the charging arrangement for electric-powered MAVs was not specifically optimized in this study. With a specified MAV fleet size, it was assumed that the battery capacity of each MAV could be tailored to the driving range required for daily operations, while charging during idle periods can be utilized to sustain operations and accommodate fluctuations in demand. Future research should focus on integrating charging strategies into the proposed MAV operation framework to further enhance efficiency and reliability. Second, the MAV-HSPTS and the proposed optimization method could be applied to larger-scale corridors with multiple bus lines. Integrating headway adjustment of modular bus with MAV scheduling may also be a crucial factor in minimizing passenger dwell time at stations.

CRediT authorship contribution statement

Zhimian Wang: Writing – original draft, Methodology, Investigation, Formal analysis. **Kun An:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Gonçalo Homem de Almeida Correia:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgements

This research is supported by the National Natural Science Foundation of China (52472349, 72361137005).

References

Auad-Perez, R., Van Hentenryck, P., 2022. Ridesharing and fleet sizing for on-demand multimodal transit systems. Transp. Res. Part C Emerging Technol. 138, 103594.
Caros, N.S., Chow, J.Y.J., 2021. Day-to-day market evaluation of modular autonomous vehicle fleet operations with en-route transfers. Transp. B: Transp. Dyn. 9 (1), 109–133.
Chen, Y., An, K., 2021. Integrated optimization of bus bridging routes and timetables for rail disruptions. Eur. J. Oper. Res. 295 (2), 484–498.
Chen, F., Fang, J., Wu, Q., 2016. Study of standing Passenger Density in Subway Cars based on Passengers' Spatial Comfort: Case Study of Beijing Subway Line 4. Transp. Res. Rec. J. Transp. Res. Board 2540 (2540), 84–91.
Chen, Z., Li, X., Zhou, X., 2019. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Discrete modeling method. Transp. Res. B Methodol. 122, 1–19.
Chen, Z., Li, X., 2021. Designing corridor systems with modular autonomous vehicles enabling station-wise docking: Discrete modeling method. Transp. Res. Part E: Logist. Transp. Rev. 152, 102388.
Chen, Z., Li, X., Zhou, X., 2020. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transp. Res. B Methodol. 132, 76–100.
Chen, Z., Li, X., Qu, X., 2022. A continuous model for designing corridor systems with modular autonomous vehicles enabling station-wise docking. Transp. Sci. 56 (1), 1–30.
Chen, Y., Wang, H., 2018. Pricing for a last-mile transportation system. Transp. Res. B Methodol. 107, 57–69.
Fu, Z., Chow, J.Y.J., 2023. Dial-a-ride problem with modular platooning and en-route transfers. Transp. Res. Part C Emerging Technol. 152, 104191.
Guo, Q.W., Chow, J.Y.J., Schonfeld, P., 2017. Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options. Transp. Res. Procedia 23, 380–399.
Hannoun, G.J., Menendez, M., 2022. Modular vehicle technology for emergency medical services. Transp. Res. Part C Emerging Technol. 140, 103694.
Hatzenbühler, J., Jenelius, E., Gidófalvi, G., et al., 2023. Modular vehicle routing for combined passenger and freight transport. Transp. Res. A Policy Pract. 173, 103688.
Ke, J.T., Qin, X.R., Yang, H., Zheng, Z.F., Zhu, Z., Ye, J.P., 2021. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. Transportation Research Part C-Emerging Technologies 122.
Khan, Z.S., He, W., Menéndez, M., 2023. Application of modular vehicle technology to mitigate bus bunching. Transp. Res. Part C Emerging Technol. 146, 103953.
Kim, M.E., Schonfeld, P., 2013. Integrating bus services with mixed fleets. Transp. Res. B Methodol. 55, 227–244.
Lin, J., Zhang, F., 2024. Modular vehicle-based transit system for passenger and freight co-modal transportation. Transp. Res. Part C Emerging Technol. 160, 104545.
Liu, Z., de Almeida, C.G.H., Ma, Z., et al., 2023. Integrated optimization of timetable, bus formation, and vehicle scheduling in autonomous modular public transport systems. Transp. Res. Part C Emerging Technol. 155, 104306.
Liu, X., Qu, X., Ma, X., 2021. Improving flex-route transit services with modular autonomous vehicles. Transportation Research Part e: Logistics and Transportation Review 149, 102331.

Ma, X., Miao, R., Wu, X., et al., 2021. Examining influential factors on the energy consumption of electric and diesel buses: a data-driven analysis of large-scale public transit network in Beijing. Energy 216, 119196.
Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transp. Res. B Methodol. 46 (1), 204–216.
Ohmio [WWW Document], n.d. URL https://ohmio.com/.
Paudel, J., 2021. Bus ridership and service reliability: the case of public transportation in Western Massachusetts - ScienceDirect. Transp. Policy 100, 98–107.
Pei, M., Lin, P., Ou, J., 2019a. Real-time optimal scheduling model for transit system with flexible bus line length. Transp. Res. Rec. 2673 (4), 800–810.
Pei, M., Lin, P., Liu, R., et al., 2019b. Flexible transit routing model considering passengers' willingness to pay. IET Intel. Transport Syst. 13 (5), 841–850.
Pei, M., Lin, P., Du, J., et al., 2021. Vehicle dispatching in modular transit networks: a mixed-integer nonlinear programming model. TransportationResearch Part e: Logistics and Transportation Review 147, 102240.
Qiu, Z., Liu, L., Li, G., Wang, Q., Xiao, N., Lin, L., Ieee,, 2019. Taxi Origin-Destination demand prediction with contextualized spatial-temporal network. In: 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 760–765.
Shi, X., Chen, Z., Pei, M., et al., 2020. Variable-capacity operations with modular transits for shared-use corridors. Transp. Res. Rec. 2674 (9), 230–244.
Shi, X., Chen, Z., Li, X., et al., 2024. Modular vehicles can reduce greenhouse gas emissions for departure flight baggage transportation. J. Air Transp. Manag. 119, 102633.
Shi, X., Li, X., 2021. Operations design of modular vehicles on an oversaturated corridor with first-in, first-out passenger queueing. Transp. Sci. 55 (5), 1187–1205.
Spera, E., 2016. CAREEM to bring driverless transportation solutions to the MENA region through partnership with NEXT future transportation. linkedin.
Tang, C., Liu, J., Ceder, A., et al., 2023. Optimisation of a new hybrid transit service with modular autonomous vehicles. Transportmetrica a: Transport Science 1–23.
Tian, Q., Lin, Y.H., Wang, D.Z.W., et al., 2022. Planning for modular-vehicle transit service system: Model formulation and solution methods. Transp. Res. Part C Emerging Technol. 138, 103627.
Tian, Q., Lin, Y.H., Wang, D.Z.W., 2023. Joint scheduling and formation design for modular-vehicle transit service with time-dependent demand. Transp. Res. Part C Emerging Technol. 147, 103986.
Tian, Q., Lin, Y.H., Wang, D.Z.W., et al., 2025. Toward real-time operations of modular-vehicle transit services: from rolling horizon control to learning-based approach. Transp. Res. Part C Emerging Technol. 170, 104938.
Tyrinopoulos, Yannis, Antoniou, Constantinos, 2008. Public transit user satisfaction: Variability and policy implications. Transp. Policy.
Wu, J., Kulcsár, B., Qu, X., 2021. A modular, adaptive, and autonomous transit system (MAATS): an in-motion transfer strategy and performance evaluation in urban grid transit networks. Transp. Res. A Policy Pract. 151, 81–98.
Zhang, J., Ge, Y.E., Tang, C., et al., 2024. Optimising modular-autonomous-vehicle transit service employing coupling–decoupling operations plus skip-stop strategy. Transportation Research Part e: Logistics and Transportation Review 184, 103450.
Zhang, Z., Tafreshian, A., Masoud, N., 2020. Modular transit: using autonomy and modularity to improve performance in public transportation. Transportation Research Part e: Logistics and Transportation Review 141, 102033.
Zou, K., Zhang, K., Li, M., 2024. Operational design for modular electrified transit in corridor areas. Transportation Research Part e: Logistics and Transportation Review 187, 103567.