Modular vehicle routing for combined passenger and freight transport

Hatzenbühler, Jonas; Jenelius, Erik; Gidófalvi, Gyözö; Cats, Oded

DOI
10.1016/j.tra.2023.103688

Publication date
2023

Document Version
Final published version

Published in
Transportation Research Part A: Policy and Practice

Transportation Research Part A 173 (2023) 103688

Contents lists available at ScienceDirect

Transportation Research Part A

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

Modular vehicle routing for combined passenger and freight transport

Jonas Hatzenbühler [a,*], Erik Jenelius [a], Gyözö Gidófalvi [b], Oded Cats [c,a]

[a] Division of Transportation Planning, KTH Royal Institute of Technology, Stockholm, Sweden
[b] Division of Geoinformatics, Integrated Transport Research Lab (ITRL), KTH Royal Institute of Technology, Stockholm, Sweden
[c] Department of Transport and Planning, Delft University of Technology, Delft, Netherlands

ARTICLE INFO

Keywords:
Public transportation
Freight transportation
Modular vehicles
Heuristic optimization

ABSTRACT

This study investigates the potential of modular vehicle concepts and consolidation to increase the efficiency of urban freight and passenger transport. Modularity is achieved by connecting multiple vehicles together to form a platoon. Consolidation is realized by integrating passenger and freight demand in the routing problem. Vehicles are specific for each demand type but can be connected freely, allowing the transport of multiple demand types in the same platoon. The routing problem formulation considers travel time costs, travel distance costs, fleet size costs, and unserved requests costs. The operations are modeled in a novel modular multi-purpose pickup and delivery problem (MMP-PDP) which is solved using CPLEX and Adaptive Large Neighborhood Search (ALNS). In an extensive scenario study, the potential of the modular vehicle type is explored for different spatial and temporal demand distributions. A parameter study on vehicle capacity, vehicle range and platoon cost saving is performed to assess their influence on efficiency. The experiments indicate a cost saving of 48% due to modularity and an additional 9% due to consolidation. The reduction mainly stems from reduced operating costs and reduced trip duration, while the same number of requests can be served in all cases. Empty vehicle kilometers are reduced by more than 60% by consolidation and modularity. A large-scale case study in Stockholm highlights the practical applicability of the modular transport system. The proposed model and optimization framework can be used by companies and policy makers to identify required fleet sizes, optimal vehicle routes and cost savings due to different types of operation and vehicle technology.

1. Introduction

In recent years the volume of e-commerce and the number of urban deliveries have steadily increased (EU, 2022). This ongoing trend, which has accelerated during the pandemic crisis, directly affects the number of vehicles needed to meet transport demand, as well as the number of delivery trips made by these vehicles. To be able to serve the increasing demand without requiring more and larger vehicles, an efficient delivery system is required. Furthermore, a reduction of vehicles and distance traveled, due to a new operation system, can have a positive impact on the sustainable development goals set by the European Commission (Sachs et al., 2021).

As a complement to electrifying vehicle fleets (Aggoune-Mtalaa et al., 2015; Mello Bandeira et al., 2019), autonomous systems (Mourad et al., 2020; Münster et al., 2020; Schlenther et al., 2020), and on-demand transport services (Ronald et al.,

* Corresponding author.
E-mail addresses: jonashat@kth.se (J. Hatzenbühler), erik.jenelius@abe.kth.se (E. Jenelius), gyozo@kth.se (G. Gidófalvi), O.Cats@tudelft.nl (O. Cats).

https://doi.org/10.1016/j.tra.2023.103688

Available online 3 May 2023

(a) Vehicle concept NXT by Scania (2020) (b) Modular vehicle concept NEXT by NEXT (2022)

Fig. 1. Illustrations of the modular vehicle concept.

2016; Winter et al., 2018; Wen et al., 2019), collaboration and modularity have recently been studied. The term collaboration is typically used to describe the consolidation of the same type of demand, e.g., freight, between multiple operators. Such collaboration has been studied in the context of vehicle routing problems by Gansterer and Hartl (2018), Wang et al. (2018b), Cleophas et al. (2019), Ferrell et al. (2020), Vargas et al. (2020), Los et al. (2020). Past studies have used well-established game theory concepts to demonstrate potential reductions in vehicle kilometers traveled if multiple freight operators share delivery requests (see Lozano et al., 2013; Guajardo and Rönnqvist, 2016; Wang et al., 2018a). Vehicle routing problems that incorporate collaboration concepts focus mainly on the delivery of goods, whereas most urban trips serve the movements of people. As is highlighted by Los et al. (2020), urban freight and passenger transport system could improve their efficiency if both types of demand are consolidated into a joint vehicle routing framework.

A related concept are modular vehicles. Pei et al. (2021) describe a passenger transport system consisting of multiple medium-sized vehicle pods that can be coupled to form variable-length platoons. In their model, the authors allow passengers to transfer between pods to further increase the flexibility of the system. The authors formulate a mixed-integer nonlinear model which is solved on small test scenarios. The authors conclude that the modular system reduces total costs by 6.63%, operation costs by 25.16% and passenger waiting time costs by 6.16%, compared with fixed-capacity shuttle buses. Lin et al. (2022) describe a bimodal system which enables the integrated last-mile transport of passengers and goods. The vehicles are driven autonomously and can couple and de-couple individually along their routes to dynamically form platoons. The authors discuss challenges and potentials of such systems and conclude that modularity and integrated transport solutions can result in more efficient, and flexible transport systems.

In this work consolidation and modularity are jointly modeled in a novel pickup and delivery problem. The necessary vehicle concepts for this novel operation system has been illustrated by several vehicle manufacturers. In Fig. 1 two example illustrations are shown: Fig. 1(a) shows a vehicle concept announced by Scania (2020) and Fig. 1(b) shows a vehicle concept designed by NEXT (2022). The core innovation of the Scania NXT vehicle concept is its ability to transport different demand types (e.g., freight or passengers) using different vehicle modules. This is enabled through the exchange of modules at dedicated places. However, the total demand that can be transported in one trip is limited to the module capacity. In the NEXT vehicle concepts all modules are able to individually drive and connect to platoons, allowing for a more flexible capacity along one trip. Additionally, each platoon can be formed by multiple types of modules. Hence, the concept allows for consolidation and modularity, where the consolidation stems from the combined consideration of multiple demand types (e.g., freight and passengers) and the modularity stems from the versatile module configuration of each platoon.

This work focuses on the impacts of the modular multi-purpose vehicle technology when simultaneously optimizing the passenger and freight transport systems by determining optimal fleet size, vehicle routes, platoon configuration and pickup and delivery times for each request. The impacts are evaluated in terms of travel distance, travel time, vehicle fleet size, and unserved demand. An extensive set of experiments has been devised, where the scenarios differ in the temporal and spatial demand distribution. Additionally, a sensitivity analysis is performed with regard to the capacity, range, and cost parameters of the vehicles.

The main contributions of this work can be summarized as follows. First, we present a mixed-integer program of the novel modular multi-purpose pickup and delivery problem (MMP-PDP). Second, an Adaptive Large Neighborhood Search (ALNS) algorithm is presented and used to efficiently solve the MMP-PDP. The ALNS is validated using CPLEX and smaller demo scenarios. Third, this paper extends the aforementioned studies on collaborative and modular vehicle routing problems (VRP) by performing an extensive parameter study and scenario analysis.

In the remainder of the paper, a literature review of the relevant and related works is presented in Section 2. Second, the formulation of the optimization problem is provided and the ALNS algorithm described in Section 3. The experimental design and parameter study are detailed in Section 4. In Section 5 the ALNS algorithm is validated and the results are reported and discussed. The paper concludes with a discussion of study limitations and an outlook on future research directions in Section 7.

2. Literature review

Our work can be categorized as a specialized version of the widely studied vehicle routing problem (VRP). The general VRP describes the routing of a single vehicle serving a set of requests, where all trips start and end at a single depot. Since the problem was first described in Dantzig and Ramser (1959) many alternative formulations and extensions have been developed. In this section we present the most relevant recent trends and variations for the proposed model. Following a brief discussion of the Pickup and Delivery Problem (PDP), routing problems with modular vehicles, transfer and transshipment operations, and concepts of platooning are elaborated. The section highlights the key features of each variant, mentions key publications, and their results, while underlining the differences in relation to the model proposed in this work. The section closes with a brief discussion on multi-purpose transport concepts and solution approaches for the respective routing problems.

The PDP as introduced by Desrochers et al. (1988) and Savelsbergh and Sol (1995) is characterized by the possibility to model dedicated pickup and delivery processes for each request. This means that requests are not only delivered from a depot to their destinations but also could be picked up along the route. Adding this feature allows modeling of on-demand services and better captures the operations of passenger delivery systems. Toth and Vigo (2014), Koç et al. (2016) and Dündar et al. (2021) provide comprehensive overviews of the various versions of VRP and PDP beyond the scope of this article. In addition to the definition of pickup and delivery nodes, the PDP typically also includes time window definitions, capacity constraints, range constraints, and multiple depots. All of these features are shared with the proposed model in this work. The main distinctive features of the model proposed in this work is the modularity of combining multiple types of demand and the possibility to form a platoon of vehicles along the route. Both features have been separately studied in previous works, but the combination has not hitherto been investigated.

In a related study by Hatzenbühler et al. (2022), we propose a PDP using single-unit multi-purpose vehicles to serve multiple demand types sequentially. In the model of that paper, the purpose of a vehicle can be changed multiple times during a vehicle route by exchanging modules at dedicated depots. The model is applied to several scenarios in Stockholm, Sweden, where it is shown that the operation of multi-purpose vehicles on consolidated demand can lead to a cost reduction of 13%, while vehicle trip durations can be reduced by 33% on average. In contrast to the model proposed in Hatzenbühler et al. (2022), this study considers the simultaneous transport of multiple demand types. Additionally, the vehicle operations considered in this study allow for vehicle platoons of variable length and with different configurations. The model forms platoon configurations at each depot before departure and the configuration cannot be changed en-route. The different vehicle operations require problem-specific heuristic optimization algorithms to guarantee feasible solutions and efficient convergence.

A related research topic is the concept of vehicle platooning. Bhoopalam et al. (2018) categorize different platooning strategies and create an extensive list of works for each platooning strategy. The authors distinguish between opportunistic, real-time, and scheduled planning, of which the last concept is the closest related to this work. In addition, the authors review work related to fixed and flexible routing strategies. The main motivation to form platoons in the listed works is expected cost savings due to reduced operational costs, reduced emissions, and improved traffic flow. Gong and Du (2018) develop a cooperative control for mixed platoons of connected, autonomous and human-driven vehicles. By controlling the speed of each vehicle in the platoon, traffic oscillations can be minimized, and traffic flow smoothness is improved. To investigate the operational cost benefit of platooning Liang et al. (2014) estimate the benefits of opportunistic platoon formation and find that on long-haul trips in Europe, a spontaneous platooning rate of 1.2% can be achieved, which in turn leads to fuel savings of 0.07%. This result indicates that a certain level of coordinated platooning is necessary. Sokolov et al. (2017) and Nourmohammadzadeh and Hartmann (2016) study scheduled platooning operations for long-haul and metropolitan areas. The works utilize exact and heuristic solvers to estimate the cost savings on small and large scenarios. Nourmohammadzadeh and Hartmann (2016) identify a potential fuel reduction of up to 5% for large scenarios.

Also the public transport sector may benefit from the formation of platoons along certain route segments. In Zhang et al. (2019) the authors investigate the efficiency of semi-autonomous and fully autonomous buses on trunk-and-branch networks. The authors formulate an analytical model that integrates passenger-related costs (e.g., waiting time and in-vehicle travel time) into the total cost formulation. The authors conclude that the platooning along network corridors carries the potential to reduce waiting time and operating costs. The theoretical cost reduction is sensitive to the demand level and the operating speed of the vehicles. For scenarios with higher speeds and higher demand along the corridor, higher cost savings can be achieved.

Masson et al. (2013), Wolfinger (2021), Dakic et al. (2021), Fu and Chow (2022), Gong et al. (2021), Liu et al. (2021), Zhang et al. (2020), Chen and Li (2021) and Chen et al. (2022) propose multiple approaches to integrating transfers and transshipments into modular passenger routing problems. Different approaches have been proposed to analyze this concept, such as continuous approximation, heuristics, and exact algorithms. These studies have in common to simplify the problem by either considering single-line networks or limiting the potential transfer or transshipment locations. The general conclusion of these works is that modular transport systems can improve conventional systems, while gains are mainly present for passenger travel, since shorter travel times, reduced waiting times, and fewer transfers do not greatly affect the cost of freight transport.

In contrast to the works mentioned above, our work focuses on the combination of passenger and freight transport within the same system. In this study, the platoon configuration is determined before departure from the depot and remains unchanged until its arrival at the destination depot. Since complex in-vehicle transfers and transshipments are not modeled, our model allows us to study larger and more complex scenarios.

A problem similar to the modular vehicle problems mentioned above including transfers and transshipment is the truck-and-trailer routing problem (see Derigs et al., 2013). Here, the fleet consists of two types of vehicles, trucks and trailers. Compared to the M-VRP, the truck and trailer problem considers only three possible vehicle configurations: (1) A truck in combination with a trailer, (2) a truck which has temporarily detached its trailer, and (3) a truck without a trailer. In the M-VRP no limitations on the possible vehicle configurations are imposed. On the other hand, in the truck and trailer problem vehicles can change configuration during their route, e.g., de- and re-attach a trailer, while in the proposed model no en-route reconfigurations are possible. In addition, in the proposed model all vehicles can operate within the entire network, whereas in the truck and trailer problem certain vehicle configurations are prohibited to enter specific areas of the network.

Table 1 gives an overview of the most relevant works as discussed in this section. The different papers are categorized based on the problem formulation, vehicle operations, solution approaches, transport concepts modeled (e.g., collaboration, vehicle platoons) and the scenario/case study analyzed.

During the last decade, several simultaneous consolidation approaches as described above (i.e., transport of passengers and freight in the same platoon) have been tested and demonstrated in pilot projects. SteadieSeifi et al. (2014), Cochrane et al. (2017), Ozturk and Patrick (2018), Behiri et al. (2018) and Li et al. (2021) provide several examples of consolidation operations. These transport systems have been mostly limited to specific routes and purposes, i.e., grocery delivery using a dedicated tram line from the warehouse to the central store or waste transportation from dedicated collection points in urban areas to the recycling station outside the city center. Typically, these applications were rail-bound and dependent on political support and financial subsidies; hence none of these pilots has remained operational. A recent summary of sustainable collaboration examples and problems can be found in Aloui et al. (2021). Despite unsuccessful long-term demonstrations, Los et al. (2020) highlight the large potential for integration and collaboration in modern transport systems. Especially, the utilization of vehicles for multiple purposes is expected to yield high efficiency gains in an urban environment.

The mentioned literature, the past pilot projects, and recent upcoming vehicle concepts all point to the high need and potential for efficient models to evaluate the impact of modular vehicle systems for multi-purpose transport systems. Since the practical operation of modular platoons that can perform en-route reconfiguration is technically not yet feasible, we focus on determining the platoon configuration before departure. This reduces the complexity of the problem and allows for the study of larger and more complex scenarios. Additionally, we close the identified research gap in modular vehicle systems for multi-purpose transport systems by developing a mixed integer linear problem which models the consolidated pickup and delivery of passenger and freight requests in an urban environment. In addition to the optimal route, our model also identifies the optimal platoon configuration to maximize vehicle utilization. Additionally, we propose a cost formulation that accounts for both travel time and travel distance-related costs and accounts for the perspective of both passengers and freight.

3. Methodology

The transport system modeled in this article is a variation of the pickup and delivery problem and consists of a number of available vehicle modules. Each module has a certain capacity and range. Modules can be of two types: module type one can only transport passengers, and module type two can only transport freight. Multiple modules can form a platoon, which allows the simultaneous transport of passenger and freight. The two commodity types are transported in separate modules, and transfers between modules within a platoon are not possible. A platoon configuration is created at the beginning of a route and this configuration remains unchanged throughout the entire route. Only one driver per platoon is required, resulting in a reduction in operational costs compared to operating each module individually. The objective of this operation is to utilize the flexible capacity configurations and operational cost savings to create a more efficient transport system. The number of modules in each platoon and their type are optimized to serve the demand in the most efficient way. In the problem considered, not every request needs to be served, but unserved requests are penalized in the objective function.

This section focuses on highlighting problem-specific challenges to first formulate the proposed pickup and delivery problem and then develop an efficient solution algorithm. The section closes with a description of the used performance indicator and the performed optimization parameter tuning approach.

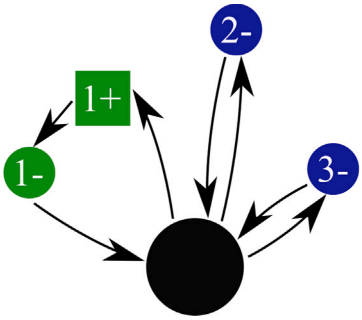
3.1. Proof of concept

The hypothesized benefits of the proposed transport system are visualized in Fig. 2 . The scenario consists of one depot (black marker), one passenger request (green marker), and two freight requests (blue markers). The passenger request has a pickup node (square) and a drop off node (circle), while the freight requests are picked up at the depot and dropped off at the round blue nodes. The numerical values in Fig. 2 represent request identifiers including '+' for pickup nodes and '-' for drop off nodes. For each request a demand of 1 is assumed and the capacity of each module is also 1. For simplicity, no time windows are defined for any of the nodes. In Fig. 2(a) the optimal solution for a conventional transportation system is shown. One vehicle serves the passenger request and returns directly to the depot. Two vehicles are required to serve the freight requests and return to the same depot. The solution in Fig. 2(b) illustrates the benefits of modularity. Here only two trips are performed in the optimal solution. One trip serves the passenger demand, while the other two modules form a platoon and sequentially serve the two freight requests. Fig. 2(c) shows the additional benefits of consolidating the two demand types and forming a platoon, which further reduces the operational costs. Instead of simultaneously operating two vehicles with two drivers, only one driver and one route are needed. In this example, modularity and consolidation can reduce both the total vehicle kilometers and the empty kilometers, which in turn reduces emissions.

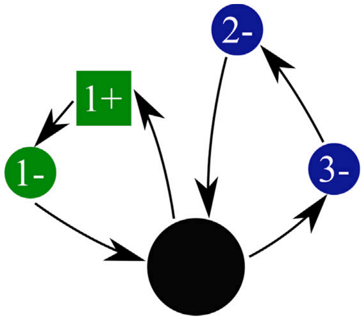
The effect of flexible capacity over time due to vehicle modularity is not displayed in this simple example. If the temporal distribution of requests form peak and off-peak periods, the modular vehicle concept is able to utilize the vehicle capacity more efficiently.

Table 1
Literature overview. Pass.: Passenger, MINLP: Mixed-integer non-linear program, GA: Genetic algorithm, ALNS: Adaptive large neighborhood search, DP: Dynamic Programming, PD: Pickup and Delivery, C: Collaborative logistics, S: Vehicle sharing, P: Vehicle platoons.

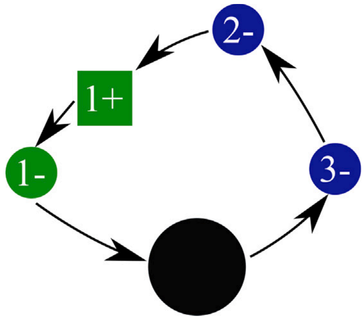
Authors	Operations	Demand	Objective	Solution	Scenario	Type	PD	C	S	P
Wang et al. (2018b)	Multi-depots and shared vehicles	Freight	Transportation, maintenance and depot costs	NSGA-II	Chongqing city, China	MILP	x	x	x	
Pei et al. (2021)	Modular platoon formation	Pass.	Vehicle operations and pass. trip times	Gurobi	Guangzhou, China	MINLP	x			x
Hatzenbühler et al. (2022)	Sequential multi-purpose vehicles	Pass. and freight	Vehicle operations, customer travel time costs, and unserved demand	ALNS and CPLEX	Synthetic created real size problems	MILP	x	x	x	
Sokolov et al. (2017)	Platoon formation	Freight	Fuel consumption	PO-LARIS/GAMS	Synthetic created	MINLP				x
Nourmohammadzadeh and Hartmann (2016)	Platoon formation	Freight	Fuel consumption	LINDO and GA	Synthetic created	ILP				x
Masson et al. (2013)	PD with transfers	Freight	Travel distance	ALNS	Synthetic and real instances	MILP	x			x
Wolfinger (2021)	PD with transshipments	Freight	Travel and transshipment costs	ALNS	Benchmark and case study	MILP	x			
Fu and Chow (2022)	PD with transfers	Freight	Travel distance, and waiting/transfer time	Ruin & repair	Synthetic and real instances	MILP	x			x
Gong et al. (2021)	Modular bus systems with transfers	Pass.	Operational cost, pass. travel cost, and unserved customers	PSO, CPLEX	Chengdu, China	MINLP	x		x	x
Liu et al. (2021)	Flexroute modular transit	Pass.	Pass. waiting time, travel time, and operation costs	Heuristics and DP	Beijing, China	MILP	x		x	x
Zhang et al. (2020)	Modular transit	Pass.	Maximize the number of served trip requests	Commercial solver	Benchmark instances	ILP	x		x	x
Chen and Li (2021)	Modular autonomous vehicles	Pass.	Pass. waiting time and total operational cost	Branch and bound	Benchmarks and Beijing, China	MILP	x		x	x
Ozturk and Patrick (2018)	Freight delivery using urban rail	Freight	Minimization of inventory levels and waiting time	Heuristics and DP	Synthetic generated	MILP	x		x	
Li et al. (2021)	Collaborative urban rail	Pass. and freight	Revenue considering operation cost, loading cost, and inventory cost	Iterative based heuristics algorithm	Two single line studies cases	MILP		x	x	
This work	Modular vehicle platoons	Pass. and Freight	Vehicle operations, customer travel time costs, and unserved demand	ALNS and CPLEX	Synthetic created real size problems	MILP	x	x	x	x



(a) Solution assuming conventional vehicle operations



(b) Solution assuming modular vehicle operations and separated demand types



(c) Solution assuming modular vehicle operations and consolidation

Fig. 2. Proof of concept for different vehicle operations and consolidation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Problem formulation

The mixed-linear program is formulated with the parameters, variables, sets, and decision variables from Table 2.

Using a directed graph $G = (N, A)$ with a set of nodes N and a set of arcs A, the proposed modular multi-purpose pickup and delivery problem (MMP-PDP) can be formulated as follows. The nodes N are the union of all depot nodes ($N^{orig} \cup N^{dest}$) and nodes associated with customer requests (N^r), i.e., $N = N^{orig} \cup N^r \cup N^{dest}$. Here, N^{orig} is the set of nodes that correspond to the depot locations. Each of these nodes is the starting point of a trip. N^{orig} is defined as $N^{orig} = \{0, \dots, n_d\}$ where n_d is the number of depots. Nodes representing trip destinations are defined as $N^{dest} = \{n_d + 2 \cdot n_r + 1, \dots, 2 \cdot n_d + 2 \cdot n_r\}$, where n_r is the number of requests. This is split into passenger and freight requests, i.e., $n_r = n_{pr} + n_{fr}$. Note that the number of origin nodes and destination nodes is the same; hence, each origin node has a corresponding destination node with the same location. This pair of nodes represents a physical depot.

Customer requests (N^r) are the set of nodes $N^r = \{n_d + 1, \dots, n_d + 2 \cdot n_r\}$. The requests are separated into passenger requests and freight requests, $N^r = N^r_p \cup N^r_f$. Each request has a dedicated pickup node (set N^+) and drop off node (set N^-). A request can then be modeled as a pair of nodes $(i, i + n_r)$ with $i \in N^+$ and $i + n_r \in N^-$. The pickup nodes are further separated into passenger pickups and freight pickups, $N^+ = N^+_p \cup N^+_f$, and drop offs are separated into passenger drop offs and freight drop offs, $N^- = N^-_p \cup N^-_f$.

Each arc $(i, j) \in A$ of the fully connected graph contains the travel time information ($t_{i,j}$) between the nodes $i \in N$ and $j \in N$. The travel time is calculated as $t_{i,j} = d_{i,j}/v + t^s_i$, where $d_{i,j}$ is the travel distance between the nodes i and j, t^s_i is the service time at node $i \in N$, and v is the constant travel speed of the network. The service time at a node represents the duration for loading or unloading. It is independent of the level of demand and the type of request.

For each node $i \in N$, demand q_i, service duration t^s_i and time windows (a_i, b_i) are defined. For depot nodes the demand and service duration are set to zero and the time windows are defined so that each depot is active for the entire planning period. Demand at the pickup and drop off nodes represents the total of items or passengers entering or leaving the modules. The demand at each drop off node is the negative value of the corresponding pickup node demand, that is $q_i + q_{i+n_r} = 0$.

The set of available module types (K) is defined as $K = \{0, 1\}$, for passenger modules and freight modules respectively. The set can be freely extended to accommodate additional types of modules. The set of available platoons is defined as $L = \{1, \dots, l_{max}\}$, where l_{max} is the maximum number of platoons in a solution. Hence, we predefine the maximum number of platoons, but a feasible solution is not constrained to form l_{max} platoons. Each platoon carries out a single trip and contains a certain number of modules ($m \in M$) of each type, with $M = \{0, \dots, Z_{max}\}$. The zero is needed to allow for platoons with only one module type. The set of platoon lengths (P) contains all the length values that a platoon can have. In the problem formulation, the length of the platoon is given by the index p and the number of modules of type (k) per platoon (l) is given by the index m. The maximum trip distance per platoon is R, hence the range is assumed to be independent of the platoon length.

Vehicle operating costs are formulated with respect to the platoon length. Two assumptions for the operation of modular vehicles are made. First, the scalar W_p with $p \in P$ accounts for the reductions in distance-related costs. Due to the formation of platoons and the potential reduction of energy consumption a reduction of distance related operating costs is assumed, which asymptotically reaches 5% for long platoons. This goes in line with the results reported in Nourmohammadzadeh and Hartmann (2016). W_p is defined as follows.

$$W_p = \frac{1 + 0.95 \cdot (p - 1)}{p} \qquad \forall p \in P \tag{1}$$

Second, it is assumed that operating costs related to fleet size, including vehicle capital costs, driver costs, and infrastructure costs (see Militão and Tirachini, 2021) are reduced due to the operation of modular vehicles. This cost reduction factor U_p is calculated as

$$U_p = \frac{1 + (1 - \eta) \cdot (p - 1)}{p} \qquad \forall p \in P \tag{2}$$

Table 2
Nomenclature and parameter values for the modular multi-purpose pickup and delivery problem (MMP-PDP).

Notation	Description
i	node index
j	node index
l	platoon index
p	platoon length index
m	module index
k	module type index
N^r	set of all pickup and drop off nodes for all types of request
N	union of N^r and depot nodes
N^+	set of all passenger and freight pickup nodes
N^-	set of all passenger and freight drop off nodes
N^{orig}	set of all origin nodes (depots)
N^{dest}	set of all destination nodes (depots)
P	set of all available platoon lengths
M	set of all available modules
L	set of available platoons
W	set of distance related cost reduction values
U	set of fleet related cost reduction values
K	set of available module types
a_i	lower time-window limit at node $i \in N$
b_i	upper time-window limit at node $i \in N$
$d_{i,j}$	distance between node $i \in N$ and $j \in N$
$t_{i,j}$	travel time between node $i \in N$ and $j \in N$
t_i^s	service time at node $i \in N$
q_i	demand at node $i \in N$
n_r	number of requests
n_d	number of depots
H	large positive number
R	maximum range for a platoon
l_{max}	maximum number of platoons
Z_{max}	maximum number of modules in one platoon
Z_k	maximum number of modules per type $k \in K$
Q_k	module capacity for module type $k \in K$
α_1	travel distance related cost parameter
α_2	fleet size related cost parameter
α_3	trip duration related cost parameter
α_4	unserved requests related cost parameter
β_p	passenger to freight unserved request relative cost
$x_{i,j,l}$	binary variable is 1 if a platoon $l \in L$ is driving between node i and $j \in N$
$y_{l,m,k}$	binary variable is 1 if platoon $l \in L$ consists of modules $m \in M$ of type $k \in K$
$e_{l,p}$	binary variable is 1 if platoon $l \in L$ consists of $p \in P$ modules
$f_{i,j,l,p}$	binary variable is 1 if a platoon $l \in L$ of length $p \in P$ is driving between node i and $j \in N$
s_i	continuous variable describing the arrival time at node $i \in N$
$g_{l,i}$	continuous variable describing the load for platoon $l \in L$ at node $i \in N$

where η is the platoon incentive parameter that models the reduction in operational costs due to forming a platoon. This cost reduction, which asymptotically reaches $\eta \cdot 100\%$ for long platoons, is motivated by two factors. First, driver costs can be significantly reduced if modules are operated in a platoon configuration compared to operated separately. Second, the creation of platoons may result in fewer trips, which in turn simplifies the fleet management processes. The numerical value for the parameter η is subject to the conducted parameter study.

To model the proposed vehicle operations, six decision variables are required. The first decision variables are all binary and model the vehicle routing, platoon characteristics, and request assignment. $x_{i,j,l}$ is a binary variable that is 1 if the platoon $l \in L$ drives between nodes i and $j \in N$ and is 0 otherwise. $y_{l,m,k}$ is a binary variable which is 1 if platoon $l \in L$ consists of $m \in M$ modules of type $k \in K$ and 0 otherwise. $e_{l,p}$ is a binary variable which is 1 if platoon $l \in L$ consists of $p \in P$ modules and 0 otherwise. This variable is needed in order to compute the fleet size related cost term. $f_{i,j,l,p}$ is a binary variable which is 1 if platoon $l \in L$ of length $p \in P$ is driving between node i and $j \in N$ and 0 otherwise. This variable is needed to linearize the distance-related cost term. The remaining two decision variables represent the arrival time s_i at node $i \in N$ and the load $g_{l,i}$ for platoon $l \in L$ at node $i \in N$, respectively.

The objective function consists of four terms (see Eq. (3)). The first term computes the cost for the total distance traveled, the second term computes the fleet size dependent cost, the third term accounts for travel time costs and the fourth cost term computes the penalty costs for unserved requests. The cost terms are scaled with the cost parameters α_1, α_2, α_3, and α_4.

Travel time costs are computed by subtracting the depot starting time from the depot arrival times for each trip. Note that the problem formulation permits waiting at each node, therefore the travel time is not equal to the traveled distance as computed in term 1. This cost term represents the perspective of the user about using the proposed transportation system. The cost for unserved requests is computed by subtracting the served pickup nodes from the total number of passenger and freight requests, respectively. The unserved passenger requests are multiplied by β_p, which represents the cost of passenger requests relative to freight requests. The total value is multiplied by α_4, which represents the average operational costs to serve an average node.

$$\begin{aligned}\min_{x,s,e,f} \quad & \alpha_1 \cdot \sum_{i,j \in N, l \in L, p \in P} d_{i,j} \cdot W_p \cdot f_{i,j,l,p} \quad + \quad \alpha_2 \cdot \sum_{l \in L, p \in P} p \cdot U_p \cdot e_{l,p} \quad + \\ & \alpha_3 \cdot \sum_{i \in N^{orig}} \left(s_{i+2\cdot n_r + n_d} - s_i \right) \quad + \quad \alpha_4 \cdot \left[\beta_p \left(n_{pr} - \sum_{i \in N_p^+, j \in N, l \in L} x_{i,j,l} \right) + \left(n_{fr} - \sum_{i \in N_f^+, j \in N, l \in L} x_{i,j,l} \right) \right]\end{aligned} \tag{3}$$

The problem formulation for the MMP-PDP is completed with the following constraints.

$$\sum_{j \in N, l \in L} x_{i,j,l} \leq 1 \qquad \forall i \in N^+ \cup N^{orig}, \tag{4}$$

$$\sum_{j \in N} x_{i,j,l} = \sum_{j \in N} x_{i+n_r,j,l} \qquad \forall i \in N^+, \forall l \in L, \tag{5}$$

$$\sum_{i \in N^{orig}, j \in N} x_{i,j,l} \leq 1 \qquad \forall l \in L, \tag{6}$$

$$\sum_{j \in N} x_{i,j,l} = \sum_{j \in N} x_{j,i+2\cdot n_r + n_d,l} \qquad \forall i \in N^{orig}, \forall l \in L, \tag{7}$$

$$\sum_{j \in N} x_{j,i,l} = \sum_{j \in N} x_{i,j,l} \qquad \forall i \in N^r, \forall l \in L, \tag{8}$$

$$x_{i,i,l} = 0 \qquad \forall i \in N, \forall l \in L, \tag{9}$$

$$\sum_{j \in N} x_{i,j,l} = 0 \qquad \forall i \in N^{dest}, \forall l \in L, \tag{10}$$

$$s_i + t_{i,j} - H \cdot (1 - x_{i,j,l}) \leq s_j \qquad \forall i, j \in N, \forall l \in L, \tag{11}$$

$$s_i + t_{i,i+n_r} - H \cdot (1 - \sum_{j \in N} x_{i,j,l}) \leq s_{i+n_r} \qquad \forall i \in N^+, \forall l \in L, \tag{12}$$

$$a_i \leq s_i \leq b_i \qquad \forall i \in N, \tag{13}$$

$$s_i \leq H \cdot \sum_{j \in N, l \in L} x_{i,j,l} \qquad \forall i \in N^{orig} \cup N^r, \tag{14}$$

$$\sum_{i,j \in N} d_{i,j} \cdot x_{i,j,l} \leq R \qquad \forall l \in L, \tag{15}$$

$$\sum_{m \in M, k \in K} m \cdot y_{l,m,k} \leq Z_{max} \qquad \forall l \in L, \tag{16}$$

$$\sum_{l \in L, m \in M} m \cdot y_{l,m,k} \leq Z_k \qquad \forall k \in K, \tag{17}$$

$$g_{l,i} \leq Q_k \cdot \sum_{m \in M} m \cdot y_{l,m,k} \qquad \forall i \in N_k^r, \forall l \in L, \forall k \in K, \tag{18}$$

$$g_{l,j} + H \cdot \left(1 - x_{i,j,l}\right) \geq g_{l,i} + q_j \qquad \forall i \in N, \forall j \in N_k^r, \forall l \in L, \forall k \in K, \tag{19}$$

$$g_{l,i} = 0 \qquad \forall i \in N^{orig}, \forall l \in L, \tag{20}$$

$$\sum_{m \in M} y_{l,m,k} \leq 1 \qquad \forall l \in L, \forall k \in K, \tag{21}$$

$$\sum_{m \in M, k \in K} m \cdot y_{l,m,k} = \sum_{p \in P} p \cdot e_{l,p} \qquad \forall l \in L, \tag{22}$$

$$f_{i,j,l,p} \leq x_{i,j,l} \qquad \forall i, j \in N, \forall l \in L, \forall p \in P, \tag{23}$$

$$f_{i,j,l,p} \leq e_{l,p} \qquad \forall i, j \in N, \forall l \in L, \forall p \in P, \tag{24}$$

$$x_{i,j,l} + e_{l,p} - 1 \leq f_{i,j,l,p} \qquad \forall i, j \in N, \forall l \in L, \forall p \in P, \tag{25}$$

$$x_{i,j,l} = \{0, 1\} \qquad \forall i, j \in N, \forall l \in L, \tag{26}$$

$$y_{l,m,k} = \{0, 1\} \qquad \forall l \in L, \forall m \in M, \forall k \in K \tag{27}$$

$$e_{l,p} = \{0,1\} \qquad \forall l \in L, \forall p \in P, \tag{28}$$

$$f_{i,j,l,p} = \{0,1\} \qquad \forall i,j \in N, \forall l \in L, \forall p \in P \tag{29}$$

$$\{s_i \in \mathbb{R} \mid s_i \geq 0\} \qquad \forall i \in N, \tag{30}$$

$$\{g_{l,i} \in \mathbb{R} \mid g_{l,i} \geq 0\} \qquad \forall l \in L, \forall i \in N \tag{31}$$

The combination of constraints (4) and (5) guarantee that each request node is visited at most once and that the nodes of a request pair are served by the same vehicle. Constraints (6) ensure that a trip starts at an origin node, while the constraints (7) force each trip to terminate at the corresponding destination node. The one-sided constraints (4) and (6) allow for unserved requests. Constraint (8) forces a platoon that enters a request node to also leave the node. Constraint (9) prohibits looping within the same node, while constraint (10) manifests that destination nodes cannot be the origin of a route.

The determination of arrival times is modeled in constraints (11)–(14). Constraint (11) models the arrival times for two nodes along a served trip segment (i,j). Note that the travel time between nodes i and j includes the service time s_i for node i. The number H is a large positive number that is defined as $H = \max\left(0, \max\left(b_i + t_{i,j} - a_j \, \forall i,j \in N\right)\right)$. Similarly, constraint (12) requires that the arrival times of the pickup nodes be lower than the arrival times of the corresponding drop off nodes. The definition of time windows is considered in constraint (13). Constraint (14) explicitly enforces that unused depots have an arrival time of 0, which implicitly sets the arrival time at the corresponding destination node to 0 as well. This constraint is required to correctly compute the travel time costs in the third cost term.

The range of each trip is constrained by (15). In constraints (16) and (17) the maximum number of modules per trip and the maximum total number of modules per type are restricted to Z_{max} and Z_k, respectively.

With constraints (18)–(20) the platoon capacity is modeled. This group of constraints connects the two main decision variables x and y related to the routing and the platoon configuration of a trip. Constraint (18) ensures that the load in a platoon at any node is less than the available capacity of that platoon. If a platoon visits a node, we assume that all requests at that node are either picked up or remain unserved. Since a platoon cannot visit the same node twice the demand cannot be split in multiple parts. The available capacity in the platoon is calculated by multiplying the module capacity Q_k by the number of modules of the same type. Constraint (19) ensures that the load of the arrival node j is larger or equal to the load of the departure node i plus the demand at node j if segment (i,j) is served. This effectively models the pickup and drop off processes at each served node. Note that the demand q_i at drop off nodes ($i \in N^-$) is negative. Constraint (20) ensures that every platoon starts its trip empty.

The group of constraints (21)–(25) are necessary to achieve a linear objective function. First, the total length of a platoon is computed. In constraint (21) the number of modules per type and platoon are defined, each platoon can consists of m modules for each type k. The total length of a platoon can be registered using constraint (22). In this equation the total number of modules p in a platoon l is set to the sum of modules m for each type k in the same platoon. Second, the information about the platoon length is added to the platoon route by constraints (23)–(25). Constraint (23) guarantees that the helper variable $f_{i,j,l,p} = 0$, if $x_{i,j,l} = 0$. Constraint (24) guarantees that $f_{i,j,l,p} = 0$, if $e_{l,p} = 0$. Constraint (25) guarantees that $f_{i,j,l,p} = 1$, if $x_{i,j,l} = 1$ and $e_{l,p} = 1$. In combination, these three constraints ensure the desired combination of x and e and allow the linear formulation of the first term in Eq. (3).

The remaining constraints (25)–(31) define the decision variable domains for platoon routing ($x_{i,j,l}$), platoon configuration ($y_{l,m,k}$), platoon length ($e_{l,p}$), route characteristics ($f_{i,j,l,p}$), node arrival times (s_i) and platoon load ($g_{l,i}$), respectively.

3.3. Adaptive large neighborhood search

The proposed linear program is a variation of the widely studied vehicle routing problems. The complexity of these combinatorial optimization problems has been proven to be NP-hard (Lenstra and Kan, 1981), therefore, the proposed model is also NP-hard. In order to solve larger instances of such complex problems, several heuristic algorithms have been developed over the years. In Ropke and Pisinger (2006) the Adaptive Large Neighborhood Search (ALNS) algorithm is described and presented for the first time. The authors show that the ALNS can outperform the best-known solutions of VRP benchmark problems by 50%. The proposed algorithm shows high robustness, repeatability, and convergence speed. Since then the ALNS algorithm has been adapted and applied in many VRPs (see Pisinger and Røpke, 2010, Masson et al., 2013, Ghilas et al., 2016, Li et al., 2016, and Hatzenbühler et al., 2022).

Due to the good performance and versatility of the ALNS optimization algorithm, we use this algorithm to solve the proposed MMP-PDP. The basic algorithm is identical to the one proposed in Ropke and Pisinger (2006). The implemented heuristics are adjusted to consider the modularity and consolidation in the model. In Algorithm 1 the general outline of the ALNS implemented is given.

First, an initial feasible solution x, which is also the current global solution x^*, is created using one of the repair operators (see Line 1 in Algorithm 1). Second, the main iterative loop is executed. In this loop destroy and repair operators are selected based on a random wheel selection process (see Eq. (32)). at each iteration the selected destroy and repair heuristics are applied sequentially to the current solution to create a new candidate solution x'. This is done . The new candidate solution is always feasible. In Eq. (32), $w_{i,j}$ is the weight of operator i, $p_{i,j}$ is the probability to choose operator i at iteration j, and O is the set of operators.

$$p_{i,j} = \frac{w_{i,j}}{\sum_{k \in O} w_{k,j}} \tag{32}$$

In the first iteration, the weights are set to 1. The weight for each operator is updated at the end of each iteration (see Line 22 in Algorithm 1) based on their performance score. In Eq. (33) this updating step is formalized, where $s_{i,j}$ is the score (σ_1, σ_2, σ_3, or σ_4)

Algorithm 1: General outline of ALNS framework (see Ropke and Pisinger (2006))

```
   Data: passenger/freight demand, depot positions, parameter settings
   Result: best solution (x*) for the MMP-PDP
1  create a feasible solution (x), set x* := x;
2  repeat
3  |  random wheel selection (see Eq. (32)) for a destroy & a repair operator using weights;
4  |  create a destroyed solution (x_d) using the chosen destroy operator on x;
5  |  create a candidate solution (x') using the chosen repair operator on x_d;
6  |  if Objective(x') < Objective(x*) then
7  |  |  set x* := x';
8  |  |  set x := x';
9  |  |  set score for chosen destroy & repair operator to σ1;
10 |  else
11 |  |  if x' is accepted (Simulated Annealing) then
12 |  |  |  set x := x';
13 |  |  |  if Objective(x') < Objective(x) then
14 |  |  |  |  set score for chosen destroy & repair operator to σ2;
15 |  |  |  else
16 |  |  |  |  set score for chosen destroy & repair operator to σ3;
17 |  |  |  end
18 |  |  else
19 |  |  |  set score for chosen destroy & repair operator to σ4;
20 |  |  end
21 |  end
22 |  update weights using new operator scores (see Eq. (33))
23 until maximum number of iterations, or objective variation threshold;
24 return x*
```

of the operator i at iteration j. The parameter $\delta \in [0, 1]$ is used to determine the decay of the weights. If δ is large the adjustment rate is low, and vice versa.

$$w_{i,j+1} = w_{i,j} \cdot \delta + (1 - \delta) \cdot s_{i,j} \tag{33}$$

The score is computed based on the objective function value of the new candidate solution x' (see Lines 6–21 in Algorithm 1). If the candidate solution is the new best global solution, the current solution and the best global solutions are updated accordingly and the score is set to σ_1. Otherwise, the candidate solution is evaluated using a simulated annealing (SA) approach. If the candidate solution is accepted and its objective value is better than the current solution, the current solution is replaced and the operator score is set to σ_2. If the candidate solution is accepted and its objective value is worse than the current solution, the current solution is replaced with the candidate solution, and the operator score is the to σ_3. If the candidate solution is not accepted the operator score is set to σ_4.

The main iteration loop is repeated until one of the two termination criteria is met: (1) The algorithm ends when the maximum number of iterations (λ) are computed, or (2) the algorithm ends when the changed in objective value over a certain number of iterations (ω) is below a threshold (ϵ). The second criteria is only active after a minimum of iterations (λ_{min}) are computed. Eq. (34) formalizes the second termination criterion.

$$\frac{\sum_{j=i-2\cdot\omega}^{i-\omega} z_j}{\sum_{j=i-\omega}^{i} z_j} - 1 \leq \epsilon \quad \text{for } \lambda_{min} \leq i \leq \lambda, \tag{34}$$

3.3.1. Heuristic operators

In this section, the implemented operators are briefly discussed. The discussions focus on the adjustments made to the operators to facilitate the modularity of the vehicles. In a related study by Hatzenbühler et al. (2022) the authors propose heuristic operators to solve a multi-purpose PDP with sequential transportation of different demand types. In this work the operators are designed to efficiently optimize the PDP for simultaneous transportation of multiple demand types, resulting in different destroy and repair heuristics. For a more detailed description of the basic operators and the reasoning for their implementation, we refer to the original paper from Ropke and Pisinger (2006).

Destroy operators. Several destroy operators are implemented to diversify the search process and facilitate the intensification process. The implemented operators are *Random removal*, *Module removal*, route removal, *Shaw removal*, and *Worst removal*. The operators differ mainly in two ways. First, the number of requests which are removed from a solution is different for each operator, resulting in small or large neighborhood variations. Second, the selection process for which request should be removed differs. Operators like

Worst removal or *Shaw removal* remove requests based on deterministic decision rules, while operators like *Random removal*, *Module removal* and route removal are based on pseudo-random decisions. All destroy operators remove node pairs from a solution, hence the depot origin node and depot destination node, as well as request pickup and drop off nodes are removed.

- *Random removal*: a random selection of served requests are removed from the solution. The removed requests are added to the list of unserved requests.
- *Module removal*: a random selection of vehicle modules is removed from the solution. The reduced platoon capacity triggers the removal of requests, which are added to the list of unserved requests.
- *Platoon removal*: a randomly selected platoon is removed from the solution. All included requests and depots are added to the list of unserved requests.
- *Shaw removal*: As proposed by Shaw (1997) and implemented in Ropke and Pisinger (2006) this heuristics first computes a relatedness index (see Eq. (36)) between two requests. Using this index the operator then removes highly related requests from the solution.
- *Worst removal*: If the objective value of a solution is reduced by removing a specific request from the solution, this request is considered to have a high contribution to the cost. The removal operator removes a number of requests that contribute most to the objective value.

The consideration of modular platoon configurations requires some special adjustments of the above-mentioned operators. All operators, except for *Module removal* and *Platoon removal*, solely remove requests and/or depots from a solution, hence the platoon configuration is not affected. However, they allow exploring the order of served requests and by that define the route and arrival times of the platoon. The number of removed requests (N) is computed using the random distribution in Eq. (35) at each iteration. The number of removed request ranges between a minimum number (ι) and maximum number $n_r \cdot \xi$, where n_r is the total number of requests in a scenario. If the number of served requests n_{served} is smaller than $n_r \cdot \xi$, N ranges between ι and n_{served}.

$$N \sim U\left(\iota, \dots, min\left(n_{served}, n_r \cdot \xi\right)\right), \tag{35}$$

The *Module removal* and *Platoon removal* operators affect the fleet size and number trips performed, as well as the request order, further diversifying the solution and therefore allowing for a simultaneous optimization of platoon route, fleet size, and platoon configuration. If a module is removed from a solution, a number of randomly selected requests of the corresponding type are removed until the capacity of the removed module is met. Therefore, the number of requests removed can vary between iterations.

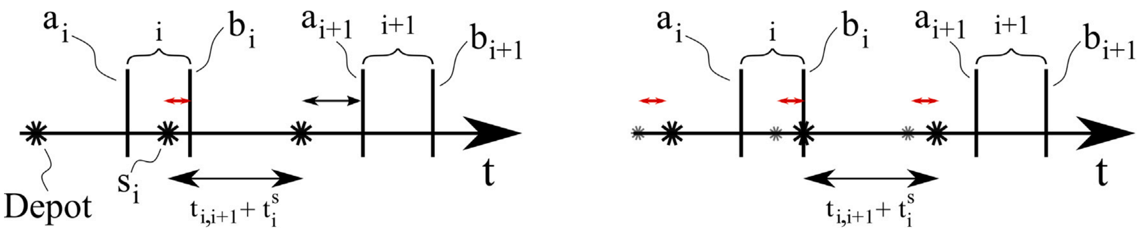
The computation of the relatedness index in the *Shaw removal* operator differs slightly from the implementation as proposed by Shaw (1997). In Eq. (36) the formulation as implemented in this work is shown. The relatedness ($R_{i,j}$) between two requests i and j and their pickup (a_i, a_j) and drop off (b_i, b_j) nodes is computed using the distance, travel time and request load information, terms one to three respectively. In Eq. (36) the distance term is computed using the travel distance $d_{i,j}$ between two nodes i and j. The travel time term uses the difference between node arrival times (s_i) and the load term is computed as the difference between the loads (q_i) of two nodes i and j. Each term is weighted using the parameter ϕ, χ, and ψ respectively.

$$R_{i,j} = \phi \cdot \left(d_{a_i,a_j} + d_{b_i,b_j}\right) + \chi \cdot \left(|s_{a_i} - s_{a_j}| + |s_{b_i} - s_{b_j}|\right) + \psi \cdot \left(|q_i - q_j|\right), \tag{36}$$

Repair operators. The goal of repair operators is to intensify the search process. Each repair operator has two main steps. The first step is to randomize the list of unserved requests and depots. In the second step these unserved requests and depots are inserted into the previously destroyed solution. Each repair operator has a different heuristic how the insertion is computed. If a request or depot cannot be feasibly inserted into a solution, this request or node remains unserved. At the end of each repair operator a feasible solution is created, which is the new candidate solution (x') (compare Algorithm 1).

- *First fit insert*: a request or depot is inserted into the first feasible location of the temporary solution.
- *Inter route insert*: a request or depot is inserted into the best feasible location within *one* platoon. The platoon in which the request is inserted is the same platoon from which it has previously been removed. For the first insertion of a request/depot a random platoon is chosen.
- *Best insert*: a request or depot is inserted into the best feasible location of *all* platoons.

Arrival time computation. Due to the integration of route duration into the objective function the optimal solution for a given scenario depends not only on the sequence of visited nodes but also on the arrival time at each node. The route duration is computed as the time difference between the departure time at the origin depot and the arrival time at the destination depot. In order to guarantee the shortest route duration for a specific route and with that implicitly guarantee an optimal solution the arrival times at each node are determined in a two step approach. First, the arrival time of node s_{i+1} is computed as the sum of arrival time at node s_i, the travel time $t_{i,i+1}$ and the service time t^s_i. Here i is the index of nodes for a given route. If $i = 0$, i.e., the beginning of a route at a depot, the arrival time represents the departure time at the depot and is determined by subtracting $t_{i,i+1}$ and the service time t^s_{i+1} from the lower time limit of node $i+1$ (a_{i+1}). If the resulting time s_i is larger than b_i, s_i is adjusted to $s_i = b_i$ to respect the time window constraints. In a second step, the arrival time s_{i+1} is evaluated. If s_{i+1} is larger than b_{i+1} the solution is not feasible. If s_{i+1} is between a_{i+1} and b_{i+1} the arrival time is accepted, and the next node pair in the route is evaluated. If s_{i+1} is smaller than a_{i+1} this means the vehicle experiences waiting time. In such cases a recursive arrival time adjustment is performed. This calculation steps aims to minimize the duration of the route up to $i+1$ by potentially minimizing the waiting time.



(a) Arrival times before waiting time adjustment

(b) Arrival times after waiting time adjustment

Fig. 3. Minimizing waiting times to reduce trip duration.

Table 3
ALNS parameters.

Notation	Description	Value
σ_1	Operator weight for new best overall solution	7
σ_2	Operator weight for better candidate solution	2
σ_3	Operator weight for accepted solution	9
σ_4	Operator weight for rejected solution	1
δ	Operator decay per iteration	0.8
λ	Iterations	10000
λ_{min}	Min. iterations	5000
ω	Iterations look-back	1000
ϵ	Objective improvement threshold	0.001
T_{start}	Start temperature simulated annealing	90
T_{end}	End temperature simulated annealing	0.0001
ν	Step size simulated annealing	0.9999
ϕ	Relatedness parameter for distance	9
χ	Relatedness parameter for time	4
ψ	Relatedness parameter for load	9
ρ	Randomizing selection parameter for *Shaw removal* requests	6
ρ_{worst}	Randomizing selection parameter for *Worst removal* requests	4
ξ	Max. request removal factor	0.32
ι	Min. request removals	1

The concept of recursive arrival time adjustment is visualized in Fig. 3. For each arrival time the time gap to its upper time window is computed. These time gaps are compared with the waiting time and the arrival times of all nodes moved forward by the shortest time gap or the waiting time. This procedure guarantees feasible solutions while generating the shortest route duration. Both steps are performed for each node of a route until all nodes are evaluated.

3.3.2. Parameter tuning

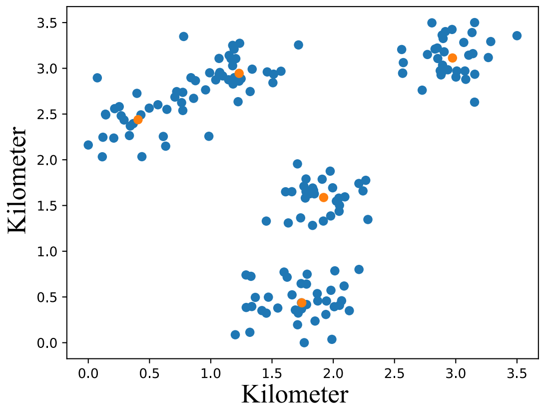
For each ALNS specific parameter (see Table 3) a tuning procedure has been performed. The procedure is aligned with that described in Ropke and Pisinger (2006), where the authors propose an iterative extensive search over a set of possible parameter values. For each parameter a set of 10 values was explored. A varied set of small and medium sized scenarios are solved using one parameter setting at a time. After all values for one parameter have been evaluated the parameter value resulting in the lowest average optimally gap with respect to the best available solution is chosen. This procedure is continued until all parameters are set. All the results in the remainder of this paper are computed with the parameter settings from Table 3.

4. Experimental design

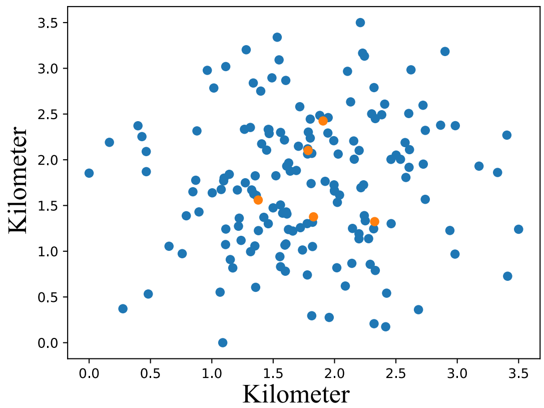
In this study two sets of experiments are performed. The first set evaluates the proposed transportation system with respect to the variation of temporal and spatial demand. The second set is a parameter analysis with respect to capacity, range, and vehicle operation costs. All the described scenarios and parameter studies are solved using the proposed ALNS. Each problem instance is solved as one ensemble run consisting of 10 individual optimization runs. This is to increase the robustness of the presented results.

4.1. Scenario definition

The scenarios included in this study are created synthetically. The created scenarios aim at representing typical transport scenarios in medium-sized to large European city. This entails an area size of $12.25\,\mathrm{km}^2$ and a planning period of 6 am to 10 pm.



(a) clustered scenario



(b) distributed scenario

Fig. 4. Illustration of spatial scenario variations.

Each scenario consists of 80 requests and 5 depots. The scenarios differ in the spatial and temporal distribution of the requests. Each request has associated demand $q \sim U(1, 15)$ which follows a uniform distribution. Furthermore, the service time for each node ranges uniformly between 1 min and 5 min, i.e. $t^s \sim U(1, 5)$.

Spatial distribution. The spatial variation is distinguished in clustered and distributed scenarios. The two scenarios allow us to study the effect of the demand pattern on the efficiency of the modular transport system. In the clustered scenarios the node locations are forming clusters following a Gaussian distribution using the depot locations as centers. The depots themselves are positioned following a Gaussian distribution on the area using the areas center as mean location. In Fig. 4(a) an example clustered scenario is given. The x and y coordinates are given in kilometers. In distributed scenarios (see Fig. 4(b)) both the depot and the node locations are spread over the area using a Gaussian distribution with the center of the area as the mean location. For passenger requests pickup and drop off nodes are spread over the case study area, whereas for freight requests the pickup location is at one of the depot locations and the drop off location is spread over the case study area.

Temporal distribution. The temporal demand variation is distinguished in evenly distributed and peak demand levels. The scenarios allow us to study the effect peak demand has on the proposed transport system. In scenarios with even demand distribution, the requests are spread out evenly over the planning period. This is achieved by assigning evenly spread arrival time definitions for each passenger pickup node and freight drop off node. In Fig. 5(a) an example distribution is shown. Here, the colors indicate the different demand types, the x-axis shows the time, and the y-axis shows the number of requests.

In the peak demand scenarios, the number of requests in each time period is computed so that it follows a demand curve with a peak during the middle of the time period. The total number of requests remains unchanged. For each passenger request the pickup and drop off time windows are designed based on the time distribution assumed. The pickup time windows ($i \in N_p^+$) are defined as $a_i \sim U(t - \Delta, t + \Delta)$ and $b_i = a_i + \Delta$, with $\Delta \sim U(5\,\text{min}, 20\,\text{min})$. The drop off time windows for passenger requests ($j \in N_p^-$) are computed with $a_j \sim U(a_i - t_{i,j}, a_i + 60\,\text{min})$ and $b_j = a_j + \Delta$, with $\Delta \sim U(5\,\text{min}, 20\,\text{min})$. For freight requests the pickup time windows range over the entire planning period, hence $a_i = 360\,\text{min}$ and $b_i = 1320\,\text{min}$ $\forall i \in N_f^+$. The drop off time windows are designed in the same way as the passenger pickup time windows, i.e. $a_i \sim U(t - \Delta, t + \Delta)$ and $b_i = a_i + \Delta$, with $\Delta =\sim U(5\,\text{min}, 20\,\text{min})$ and t being the computed time based on the assumed temporal distribution. With these time window definitions, each request can be served by an individual platoon so that not all requests are infeasible by design. In Fig. 5(b) an example of the peak demand levels is shown.

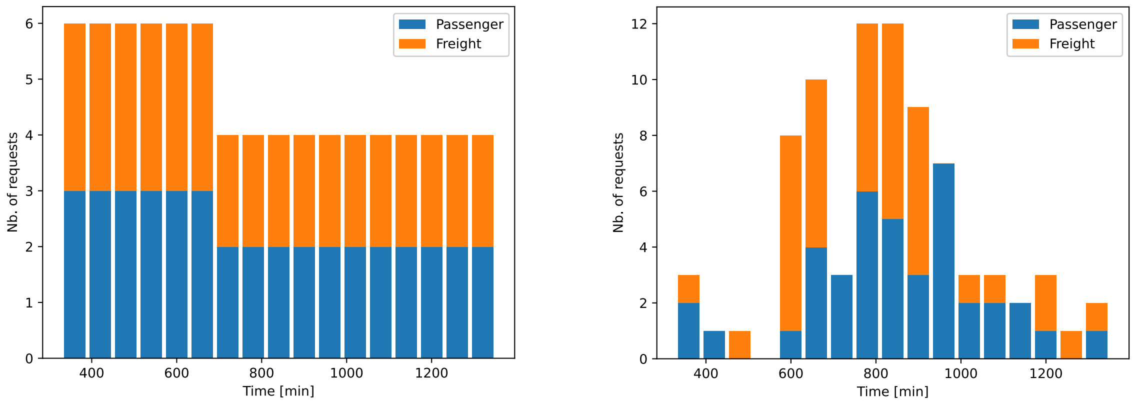
The conducted experiments contain a combination of the spatial and temporal distributions, hence there are four different scenario classes investigated, clustered-even, clustered-peak, distributed-even, and distributed-peak scenarios. In total, there are 5 different instances of each scenario type computed, hence the reported results are based on a total of 20 different scenarios.

Consolidation. In order to study the impacts of consolidation of passenger and freight requests on the efficiency of the transport system, each scenario is solved once without consolidation and once with consolidation. If no consolidation is assumed the passenger and freight requests are solved separately and summed up. For conventional vehicle operation the solutions with and without consolidation are identical, since each vehicle module can only transport one type of demand.

4.2. Parameter settings

The parameters used in the performed experiments are shown in Table 4. If not specified further, these parameters are employed in all scenarios. The values for the number of requests, number of depots, service time and demand per request are set so that the computation time is reasonable and the complexity of the problems is comparable with real-sized problems.

The values for module range (R), module capacity (Q_k) and maximum platoon length (Z_{max}) are in line with reported values from research reports and prototype vehicles (NEXT, 2022). The number of modules per type (Z_k) is defined to provide sufficient



(a) even scenario

(b) peak scenario

Fig. 5. Illustration of temporal scenario variations.

Table 4
Parameter values used in numerical experiments.

Notation	Description	Value
n_r	number of requests	80
n_d	number of depots	5
R	maximum range for a platoon	200 km
l_{max}	maximum number of platoons	20
Z_{max}	maximum number of modules in one platoon	10
Z_k	maximum number of modules per type $k \in K$	10 per type
Q_k	module capacity for module type $k \in K$	15 per type
α_1	travel distance related cost parameter	0.096 EUR/km
α_2	fleet size related cost parameter	19.37 EUR/h
α_3	trip duration related cost parameter	6.9 EUR/h
α_4	unserved requests related cost parameter	19.37 EUR/h
β_p	passenger to freight unserved request relative cost	1
η	operational cost reduction	0.6

supply to absorb the demand, hence it is feasible to serve every request with the provided combination of fleet size and module capacity. The maximum number of platoons l_{max} is set to the sum of available modules per type to allow one trip per module. Hence it is feasible to operate the fleet of modules as conventional, individual vehicles. In real applications, this parameter should be set based on what is feasible for the operator regarding monetary budget, garage space, staff capacity, etc. If the constraint based on these empirical considerations is non-binding in the optimization problem, then l_{max} can be reduced to save computational time.

The cost parameters α_1 and α_2 are based on the values reported in Militão and Tirachini (2021). The authors estimated the distance-based and time-based operation costs of electric vehicles based on available data from the city of Munich, Germany, and proposed a linear estimation model based on vehicle capacity to estimate the two cost parameters. On the basis of the published data, the models are $\alpha_1 = 0.003599 \cdot Q_k + 0.04162$ and $\alpha_2 = 0.1753 \cdot Q_k + 16.8$. The parameter α_2 is scaled with the duration of the planning period.

The parameter α_3 scales travel time costs and is based on the value-of-time for public transport users in Stockholm, Sweden as reported by Börjesson and Eliasson (2014). In Eq. (3) the parameter α_3 corresponds to the value of time for passenger travel. Hence, it increases the overall cost of total travel time per route in the objective function. Since this value is also applied on freight requests, which can be assumed to have a lower value of time, this implies an overestimation of the total travel time costs. However, the different value of time for passenger and freight requests is implicitly considered using the time window definitions. In the scenario definition passenger time windows are set to be tighter compared to freight requests, indicating the higher importance of on-time operations for passengers. The parameter α_4 penalizes the unserved requests. The value for this parameter is chosen to scale with the fleet size related cost parameter. The reasoning behind the value is that it should not lead to a reduced objective value if a single module serves a single request. Hence, the penalty for keeping demand unserved equals that of α_2 multiplied with the duration of the planning period and the average demand for the request. The value of η is estimated to be 0.6 as it is reported in Zhang et al. (2019) who provide cost reduction estimations for bus platoon operations based on data published in project reports.

The difference between conventional vehicle operations and the proposed modular vehicle operations is modeled by adjusting the Z_{max} parameter. For the case of conventional vehicle operations $Z_{max} = 1$ and for modular vehicle operations $Z_{max} = 10$. The other cost parameters remain unchanged to investigate the impact of the changed vehicle operations. When optimizing a scenario using conventional vehicle operations, the best solution for the corresponding scenario using modular vehicle operations with separated demand is used as the initial solution. That way a high quality initial solution is used which improves the optimization process by reducing computation time and achieving high robustness.

4.3. Parameter analysis

A parameter analysis is performed in order to explore the influence different parameter settings have on the objective value and the benefits modular vehicles might have on the transportation system . The three parameters varied in this analysis are the vehicle capacity, vehicle range and the platooning operational cost reduction parameter. For each of these parameters an array of discrete values is created. The vehicle capacity spans from 15 to 45, the vehicle range spans from $50\,\mathrm{km}$ to $250\,\mathrm{km}$ and the cost reduction parameter (η) spans from 0.2 to 1. All other parameters remain unchanged. The parameter study is performed on scenarios with distributed-peak demand characteristics. Each of the 5 created scenarios with this temporal and spatial demand definition is computed with all parameter settings. The reported values are the average values for all scenarios and its ensemble runs. In total $3 \cdot 5 \cdot 10 \cdot (5+5+4) = 2100$ experiments are evaluated for parameter analysis.

4.4. Key performance indicators

In addition to the cost terms (route distance, route duration, fleet size, and unserved demand), the analysis includes the following key performance indicators.

Fill rate. The fill rate is computed by adding the demand of all the pickup nodes on all the routes of a solution. This total demand units transported is divided by the total capacity of all modules of all platoons used in the solution. Therefore, this indicator provides an understanding of the overall utilization of the supplied capacity. The higher the fill rate, the better the module capacity is used for a solution.

Request kilometers. The request kilometers are computed as the total traveled distance over all requests (passenger and freight). The traveled distance is the distance from pickup node to drop off node along the traveled route.

Request minutes. Similar to the request kilometers this indicator sums the traveled time from the pickup node to the drop off node of all passenger and freight requests. In this indicator, also potential waiting times along a requests travel path are included.

Load per platoon distance. This indicator is calculated by dividing the total number of passenger and freight served with the total distance of all platoons in that solution.

Empty vehicle kilometers. The empty kilometers of a solution are computed as the sum of the length of all segments in each route where the platoon is driving without any load on board. This indicator also includes the empty kilometers at the beginning and end of a route before returning to the depot.

5. Results

5.1. ALNS performance analysis

The validation of the implemented heuristic optimization algorithm is performed by solving small and medium-sized problem instances with ALNS and the exact solver CPLEX. The small- and medium-sized instances are created as described in Section 4. In Table 5 the results of the benchmark tests are summarized. For each problem instance the table gives the number of nodes, number of depots, the objective values for CPLEX and best objective value for ALNS, and the computation times for CPLEX and ALNS. The ALNS algorithm is run 10 times for each problem instance. The standard deviation of the objective value and the mean ALNS computation time of all ensemble runs are reported in Table 5. The column "Mean First Found" shows the mean number of iterations across all ensemble runs until the final objective value was reached for the first time, indicating the convergence speed of the ALNS. The reported computation times are recorded from the start of the ALNS algorithm until the first convergence criterion is met. CPLEX is configured to have a maximum computation time of $1\,\mathrm{h}$ per instance and the accepted optimally gap is set to 0.01%.

The ALNS algorithm is able to find the optimal solution determined by CPLEX in all instances. In four instances (ID 4, ID 15, ID 28, ID 30), the objective value determined by the ALNS is slightly lower than the CPLEX value. The difference between the two values lies within the predefined CPLEX optimality gap of 0.01%, hence these instances can be assumed as solved to optimality as well. The ALNS ensemble runs for each instance have a small standard deviation, indicating a robust convergence. In 22 out of the 30 instances, all 10 ensemble runs per instance converge to the same optimal solution.

Even for some smaller instances (ID 8, ID 12), ALNS outperforms CPLEX in computation time. This finding is emphasized by the medium sized instances of 12 and 16 nodes where the ALNS converges faster to all optimal values. The computation time of ALNS is more related to the size of the problem, whereas the computation time of CPLEX is fluctuating within one problem size (see instances 29 and 30). For instances 20 and 29 ALNS outperforms the CPLEX algorithm in both computation time and objective value by a large margin. In these instances CPLEX does not converge to a global optimal solution within 1 h. The last column in Table 5 shows the mean number of iterations across all ensemble runs until the final optimal solution is first found. This metric shows the fast convergence speed of the ALNS algorithm and motivates the chosen convergence criteria. Using this metric and the average computation time per iteration, the ALNS finds the optimal solution for all benchmark instances within $0.5\,\mathrm{sec}$.

Table 5
ALNS validation results.

ID	Nodes	Depots	Obj. CPLEX	Obj. ALNS	Stdev. Obj. ALNS	Time [s] CPLEX	Mean time [s] ALNS	Mean first found
1	8	1	547.465	547.465	0	10.3	17.17	7.3
2	8	1	800.808	800.808	0	15.6	17.8	10.7
3	8	1	676.032	676.032	0	16.6	9.92	8
4	8	1	674.499	674.451	0	6	14.38	5.3
5	8	1	797.586	797.586	0	6	12.48	3.1
6	8	2	641.608	641.608	0.66	236.5	15.77	108
7	8	2	572.472	572.472	0	213.1	12.62	104
8	8	2	765.994	765.994	0	68.6	7.91	15.8
9	8	2	1015.083	1015.083	0.06	273.1	9.45	34.2
10	8	2	768.393	768.393	0.03	273	7.31	1271.6
11	8	2	849.648	849.648	0	15.5	8.45	8.4
12	8	2	845.769	845.769	0	156.5	9.2	8.8
13	8	2	847.112	847.112	0	38.7	8.35	8.3
14	12	1	705.163	705.163	0	149.4	27.07	125.7
15	12	1	925.901	925.881	0	23	30.15	11.6
16	12	1	579.082	579.082	0	186.6	15.24	29.3
17	12	1	580.514	580.514	0	121.8	36.08	113.6
18	12	1	801.68	801.68	0	21	38.67	12.3
19	12	2	654.029	654.029	0.09	1083.8	32.53	801.4
20	12	2	704.316	580.303	0	3600.2	17.18	83.9
21	12	2	1020.164	1020.164	1.79	3601.2	12.19	430.2
22	12	2	967.319	967.319	0.14	668.9	24.29	21.4
23	12	2	649.579	649.579	0.23	3604	11.07	784.4
24	12	2	701.256	701.256	0	3601.5	13.74	149.5
25	12	2	907.161	907.161	0	818.8	12.23	52.6
26	12	2	781.22	781.22	0	229.8	12	36.9
27	12	2	1099.148	1099.148	38.99	3600.2	21.55	408.6
28	12	2	676.335	676.334	0	260.1	28.49	60.1
29	16	1	832.738	708.393	0	3600.4	40.15	54.9
30	16	1	708.483	708.482	0	56.8	55.51	65.7

5.2. Scenario analysis

In this section the results for the different scenarios in terms of spatial and temporal demand distributions are presented. The numerical values are the average values over all computed scenarios and their ensemble runs. The analysis focuses on the effects of vehicle technology and consolidation on system performance.

Fig. 6 shows the objective values with each color representing a cost term. It can be seen that the cost of the fleet size and the cost of the trip time dominate the objective value. The distance-related cost is negligible which can be explained by the parameter settings in relation to the higher unit costs for modules. When analyzing the impacts of vehicle technology and consolidation, it can be seen that, due to a large reduction in fleet size and trip time costs, modular operations result in lower overall costs, while the unserved demand remains unchanged. Additional savings can be achieved by introducing consolidation between passenger and freight demand. These cost savings are independent of the spatial and temporal demand distribution. In addition to the effect of vehicle technology and consolidation, the results show that spatially distributed demand patterns lead to slightly lower total costs.

In numerical values, the use of modular vehicles results in an average total cost reduction of 48%, and consolidation further reduces the objective value by 9%. The fleet size costs, distance costs, and trip time costs can be reduced by 50%, 15%, and 58% with modular vehicles and by further 9%, 4%, and 18% with consolidation, respectively. The costs for unserved demand remain unchanged.

Some scenarios have unserved requests due to the time windows for these requests, which imply that it would require one one-module platoon to serve each request. By design (fleet size costs, penalty for unserved demand and travel distance/time cost), serving a single request with a single module results in a higher objective value compared to not serving it.

The large reduction in fleet costs is mainly due to the reduction in operational costs made possible by the formation of platoons. Fig. 7(a) indicates that on average 1.8 modules form a platoon. The average platoon length increases to 2.1 if passenger and freight demand are consolidated. The large reduction in total fleet size cost is also due to a reduction in the number of required modules for the different types of operation.

Fig. 8(a) shows the average number of passenger and freight modules used for each scenario. In conventional operation, approximately the same number of modules is used for passengers and freight, and the reduction in the number of modules in the other cases mainly stems from the reduction of required passenger modules. This is due to a change in the delivery order for the served freight requests.

Due to the connection of multiple modules, more passengers can be jointly transported, allowing the time window constraints to be met more efficiently. Such exploitation might result in longer travel distances which are compensated by the reduction in the required number of modules. The longer travel distances for the various types of operation can be seen in Fig. 8(b), where

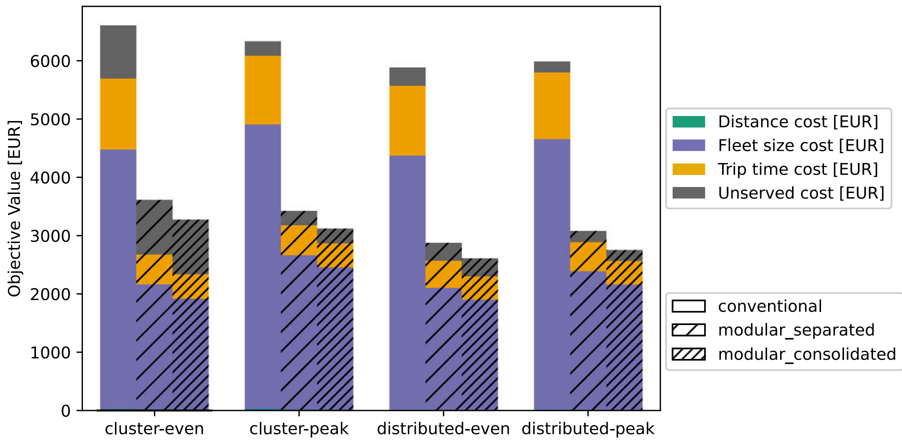
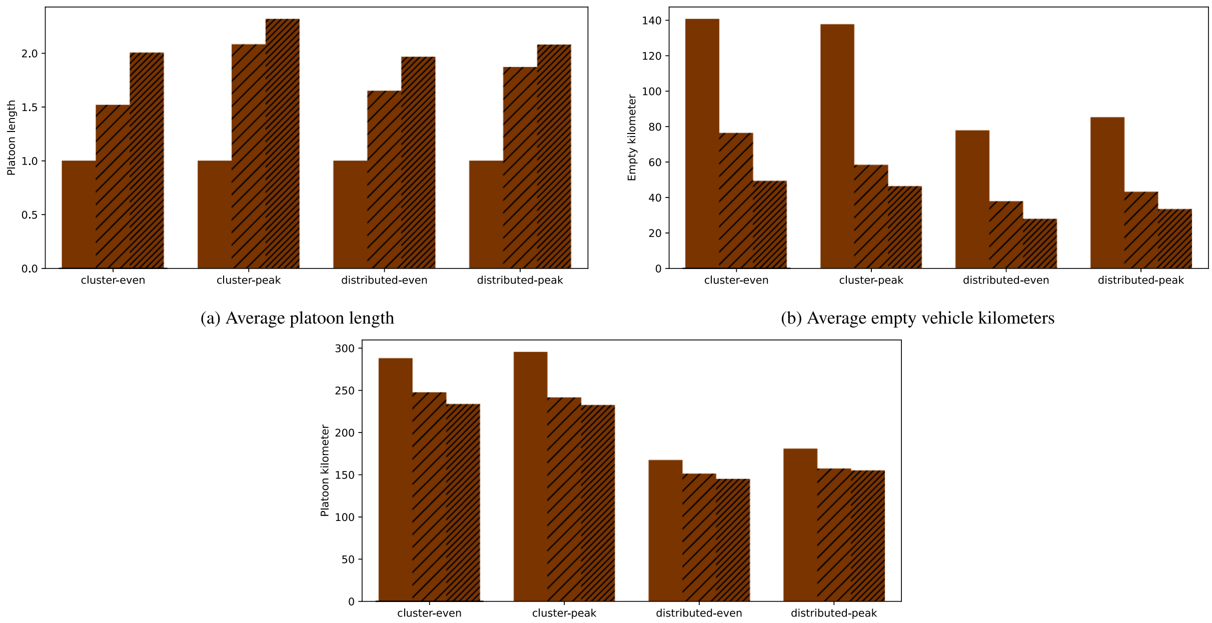


Fig. 6. Objective function values per scenario.



(a) Average platoon length

(b) Average empty vehicle kilometers

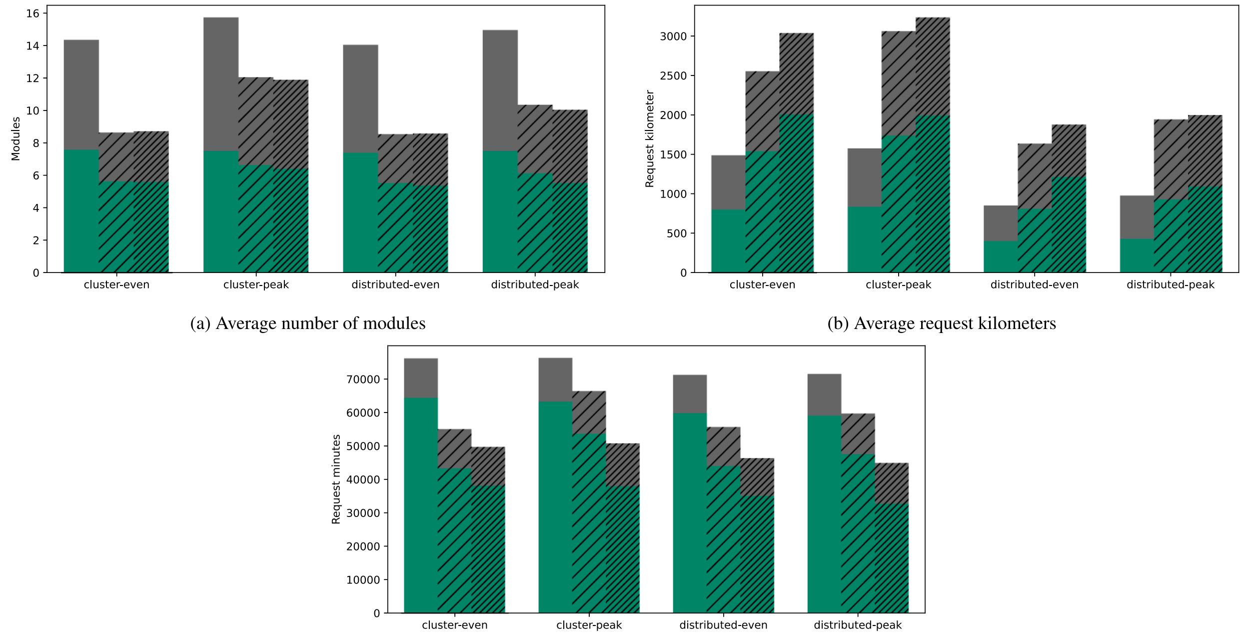
(c) Average platoon kilometers

Fig. 7. Effects of spatial and temporal demand distributions. The vehicle operations are indicated by different patterns, conventional (), modular without consolidation (), and modular with consolidation ().

the kilometers traveled for each request demand unit increases with modularity and consolidation. It can also be seen that fewer traveled kilometers are needed to serve the demand in distributed demand patterns compared to clustered demand patterns. This is expected since the average distance between nodes is lower for distributed scenarios.

When looking at the empty vehicle kilometers for each scenario (see Fig. 7(b)) two clear trends can be seen. First, the introduction of modular vehicle types reduces empty kilometers by more than 50% in all cases, while an additional 27% can be saved by combining passenger and freight demand. This is another explanation for the reduction in the number of modules required. The empty vehicle kilometers are lower in distributed demand scenarios than in the clustered scenarios, i.e., a higher capacity utilization can be reached. The difference between even and peak demand patterns is not significant.

Fig. 7(c) underscores the previous results. Although the request kilometers (see Fig. 8(b)) increase by 90% with modular vehicles and increase by an additional 10% with demand consolidation, the reduced number of required modules leads to a reduction of platoon kilometers required to serve the demand. This directly translates to a reduction of emissions and congestion. The positive effect is most prominent for clustered scenarios and scenarios with peak demand.



(a) Average number of modules

(b) Average request kilometers

(c) Average request minutes

Fig. 8. Effects of spatial and temporal demand distributions. The vehicle operations are indicated by different patterns, conventional (▭), modular without consolidation (▨), and modular with consolidation (▨). The colors indicate the demand type, passenger demand (■) and freight demand (■). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In contrast to the increase in request kilometers, modular operations without consolidation reduces the request minutes by an average of 20%, and consolidation reduces the request minutes by another 19% (Fig. 8(c)). Most of the reduction is in freight request minutes, which are reduced by 23% in both operation modes. Passenger request minutes vary by less than 1%, resulting in an unchanged level of service for users of such transport systems. The discrepancy between freight and passenger request minutes can be explained by the design of the scenarios. Freight requests have loose time windows for the pickup nodes, meaning that longer waiting times between multiple requests and therefore longer in-vehicle times, are feasible. The passenger time windows are defined tightly and therefore do not allow for large pickup variations. The average in-vehicle time is approximately 91 min for freight requests and 17 min for passenger requests.

5.3. Sensitivity analysis

The sensitivity analysis is performed using the distributed and peak demand distribution. By analyzing the graphs in Figs. 9 and 10, it can be seen that the solutions do not depend on the module range settings. We therefore conclude that, first, the range setting of 50 km is sufficient for the study area chosen, and second, when providing a sufficiently large fleet of vehicles the range for urban deliveries is not a limiting factor.

As can be seen in Fig. 9 larger module capacities in lead to a reduction in overall costs. However, conventional vehicles are more sensitive to vehicle size since the benefits of larger capacities are greater, and a total cost reduction of around 50% can be observed when increasing the capacity from 15 to 45. The overall cost for modular vehicle operations can also be reduced by around 15% when increasing the module capacity; however, the cost reduction saturates at a module capacity of 35, indicating a turning point of optimal fleet characteristics. The cost reduction is mainly due to the reduced number of modules needed to serve the demand (see Fig. 10(b)). For capacities above 35 the disadvantages of fewer models, i.e., longer traveled distances and longer travel times, outweigh the cost savings gained by further reducing the number of modules (see Figs. 10(f) and 10(h)). The distances traveled and the travel time increase by 13% and 7%, respectively, while the number of modules is reduced by only 5% from vehicle capacity 35 to 45. The unserved demand stays unchanged throughout all parameter settings.

A similar observation can be made for the empty kilometers and average platoon length, both decreasing with larger modules. The empty kilometers can be drastically reduced for conventional operations but only by a small margin with modular vehicle operations. As expected the optimal platoon length is shorter with larger module capacities, and this trend saturates at a module capacity of 35.

The average fill rate per module is shown in Fig. 10(e). The fill rate for modular operations is higher than for conventional vehicles and is less affected by larger module capacities. Furthermore, with larger module capacities, the fill rate gets lower, indicating less efficient use of available capacity. The jump for conventional operations indicates a shift in operation for larger

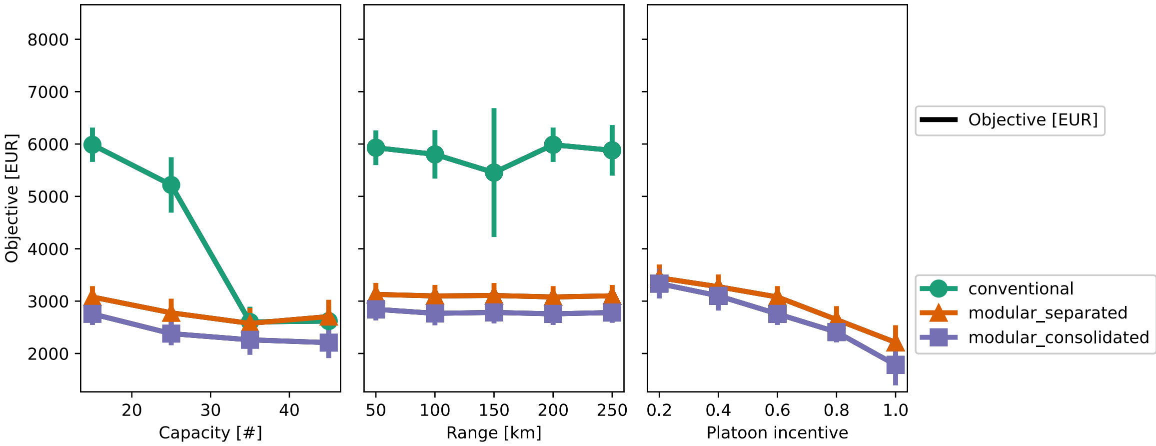


Fig. 9. Effects of module capacity, range and platoon incentive parameters on the objective value.

capacities conventional operations and modular operations with separated demand operate identical. Interestingly, the fill rate for the consolidated operation is lower than if each commodity is transported separately.

As shown in Figs. 10(b) and 10(h) the passenger transport contributes significantly less than the freight transport to the overall cost. The total time needed to transport all passengers and the number of passenger modules are lower for all parameter settings and for all types of modular vehicle operations. As discussed in the previous section, this shows that passengers can be transported together with freight without adding much additional trip time. As the module capacity increases, the change in request minutes shows that the total delivery time for each request is increasing. Additionally, the same number of passenger modules is needed in the conventional and modular operation for higher values of vehicle capacity, which reduces the benefits of modular operations.

Looking at the load transported per platoon kilometer (see Fig. 10(g)), the modular and consolidated operations outperform the conventional mode of operation. The load transported per platoon kilometer increases slightly with an increase in module capacity.

The influence of the platoon incentive in Figs. 9 and 10 mainly shows two effects. First, the higher the incentive to form platoons, the longer the platoons. This results in more used modules, since the number of trips is not significantly reduced. Since the platoon incentive parameter is directly related to the objective value, the total cost for operating such systems decreases. Second, however, for larger platoon incentives, the request kilometers and request minutes increase, leading to a longer average trip distance and average trip duration. A similar negative effect can be seen in relation to the fill rate, which is slightly reduced for higher platoon incentives. This indicates that the reduction of operational costs leads to more supply which in turn is not used efficiently to serve the demand.

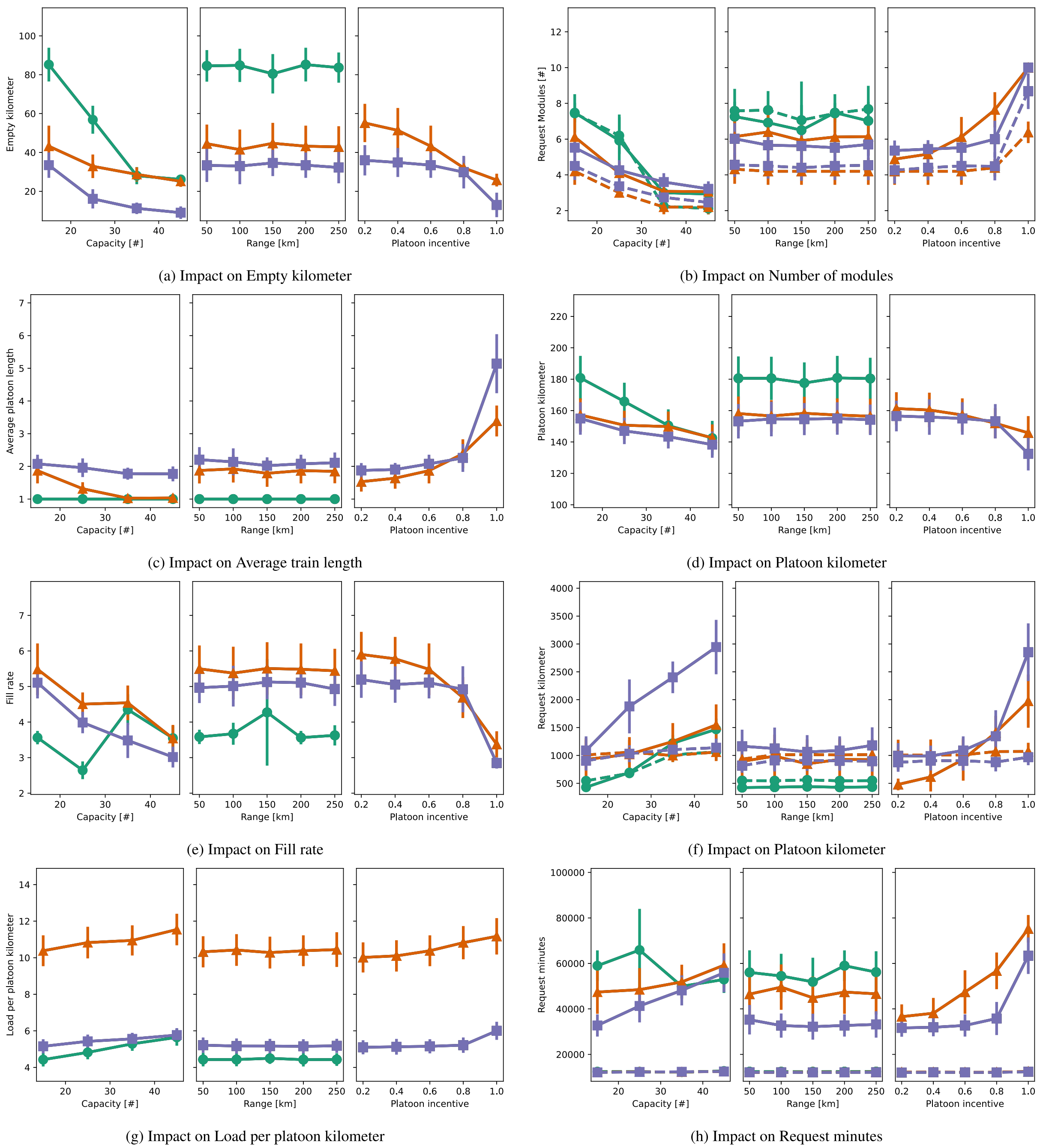
The results discussed in this section are not impacted by different β_p values. Only 1%–3% requests are unserved in the scenarios and therefore changing the value for β_p has no significant effect.

6. Case study

The case study is situated in Stockholm, Sweden. In Fig. 11 the pick-up and drop-off locations including the two depots are visualized. Each customer request has a certain demand level and time windows assigned. The spatial passenger demand patterns and levels represent expected daily passenger movements in Stockholm, while the freight demand is evenly distributed over the case study area. The freight pick-up locations are at the depots. The time windows follow a peak-hour distribution as exemplified in Fig. 5(b). The ALNS and model parameters of this case study are in line with the values shown in Tables 3 and 4. The travel distances and travel times used in the graph representation of the MMP-PDP are based on the underlying road network. For each node pair the shortest path is computed, and its distance recorded. The road network information is extracted from OpenStreetMap contributors (2017).

The results of the case study show similar trends as the analysis in Section 5. The overall objective value is substantially reduced when operating modular vehicles and is reduced further when consolidating passenger and freight demand (see Fig. 12). The main cost reduction is due to the reduced fleet size cost and the reduction in trip times. Multiple modules form platoons (see Fig. 13(c)) and in some scenarios fewer modules are needed to serve the demand. This does not lead to a higher number of unserved requests compared to the conventional operations, indicating an unchanged level of service.

In Fig. 13(a) the effect on empty kilometers is shown. The large reduction in empty kilometers for modular operations highlights the higher utilization of available capacity. In modular operations fewer empty trips towards the depot need to be made since the platoons can serve more requests in one trip. This also results in a changed delivery order. The changes in platoon kilometer as shown in Fig. 13(b) follow a similar trend as discussed in Section 5 and shown in Fig. 7(c). The overall platoon kilometers are drastically reduced with modular vehicles, and additional savings are achieved with consolidation. This trend reflects the changed



(a) Impact on Empty kilometer

(b) Impact on Number of modules

(c) Impact on Average train length

(d) Impact on Platoon kilometer

(e) Impact on Fill rate

(f) Impact on Platoon kilometer

(g) Impact on Load per platoon kilometer

(h) Impact on Request minutes

Fig. 10. Results for the parameter study. The different operations are marked with different colors and symbols, conventional operations (●), modular without consolidation (▲), and modular with consolidation (■). The demand type is illustrated using different line patterns, freight demand (▬▬) and passenger demand (▬ ▬). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

delivery orders and the more efficient serving of the requests, and it indicates the reduced number of modules needed to serve the demand.

Fig. 13(c) shows that the average platoon length is approximately 2.7 for modular and separated operations and approximately 4.1 for modular and consolidated operations. When comparing this to the scenario analysis in Section 5, two observations can be made. First, the trend of longer platoons for modular and consolidated operations is persistent in all experiments. Second, the higher demand levels and different demand patterns in the Stockholm case study lead to longer platoons as well as a larger fleet size to serve the demand.

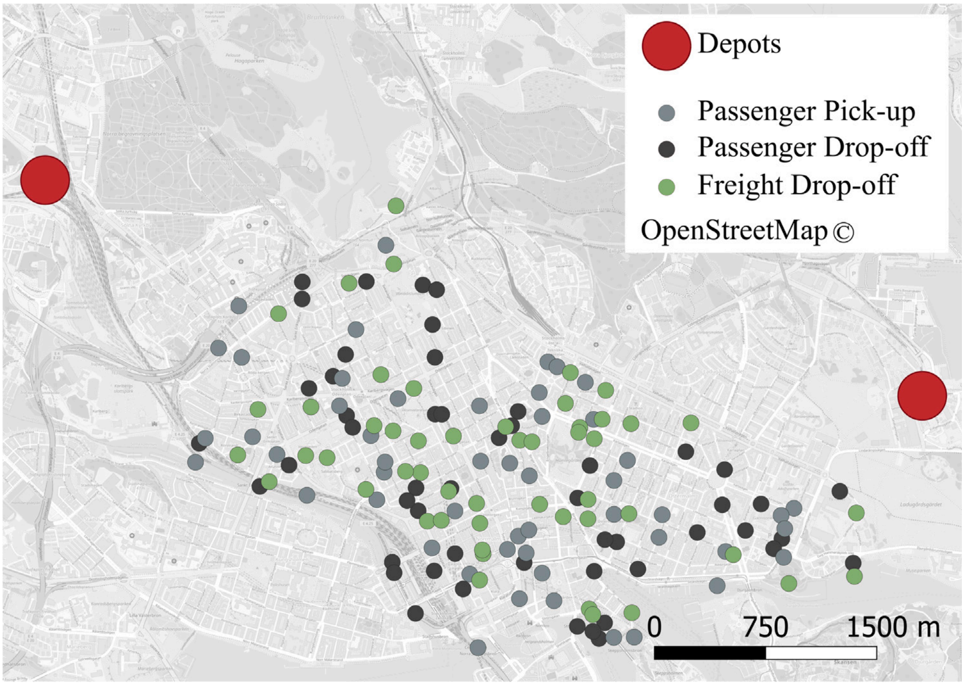


Fig. 11. Spatial representation of the case study in Stockholm, Sweden.

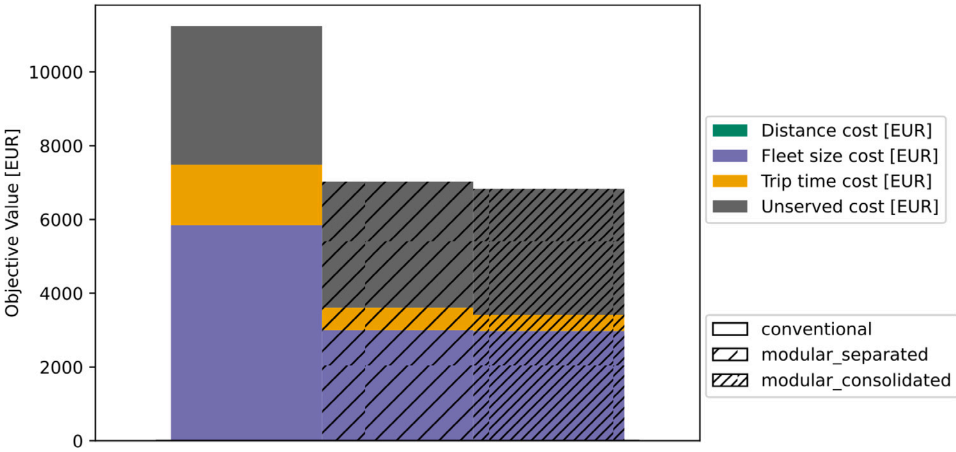
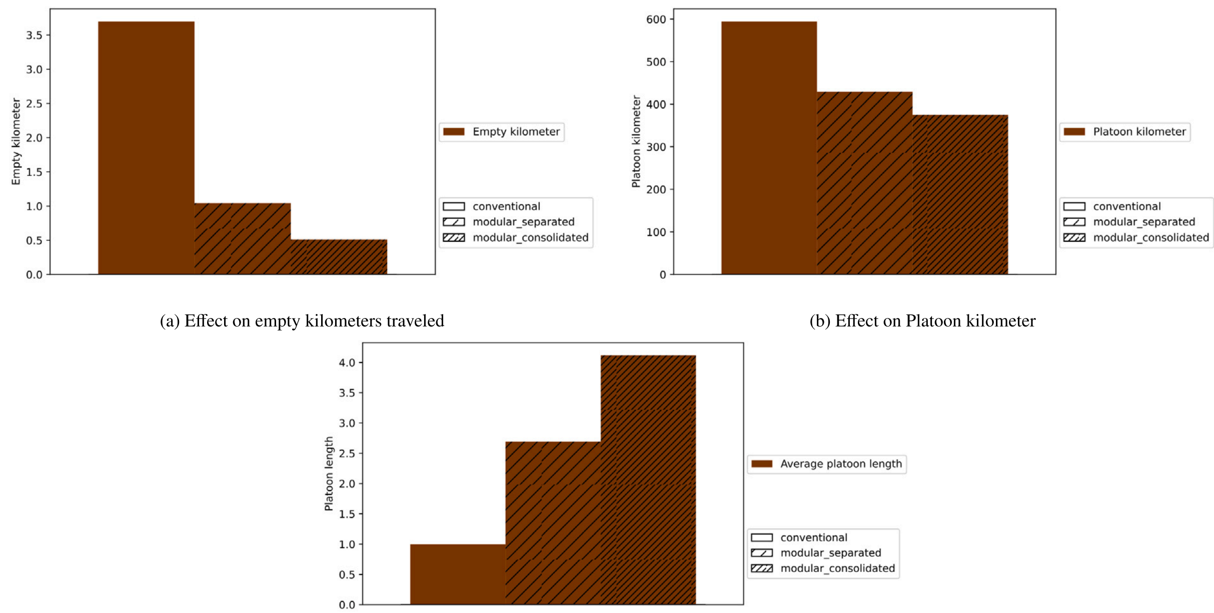


Fig. 12. Cost terms and objective value for conventional, modular and consolidated operations in the Stockholm case study.

The results of the case study highlight the overall system improvement when using modular vehicles compared to conventional vehicles. Similar effects as in the generated scenarios can be identified for the empty kilometers, platoon kilometers and platoon length. The realistic travel distances and travel times based on the underlying road network increase the practical relevance of the study and suggests significant cost savings for a variety of transport scenarios.

It should be noted that the results are achieved assuming $\beta_p = 1$. Since the level of unserved demand in this case study is approximately 10%, the results are expected to be affected by different values of β_p. For very uneven cost penalties for unserved passenger/freight demand, the benefits are expected to diminish. In such cases the transportation system resembles the conventional system, since higher priority is set to either freight or passenger requests. This leads to less opportunities for consolidation, which is the underlying reason for cost reductions. This effect is expected to be stronger for high passenger penalties, i.e. $\beta_p \gg 1$. As described in Section 5 the main cost savings stem from serving passenger requests while delivering freight. If passenger delivery is preferred fewer such combinations are possible and hence lower cost savings can be expected.



(a) Effect on empty kilometers traveled

(b) Effect on Platoon kilometer

(c) Effect on Platoon length

Fig. 13. Results for the case study in Stockholm, Sweden. The vehicle operations are indicated by different patterns, conventional (▭), modular without consolidation (▱), and modular with consolidation (▱).

7. Conclusion

In this work, we study the potential of modular vehicle operations in urban freight and passenger transport systems. The modular vehicle concept allows the operator to form platoons by connecting multiple modules, where different modules can be used to transport different types of commodities. Furthermore, we analyze potential efficiency gains by consolidating multiple commodity flows in the transport system. A novel modular multi-purpose pickup and delivery problem (MMP-PDP) is proposed to analyze the effects of each vehicle operation. The model extends existing research work by adding the necessary modularity and consolidation formulations to existing PDP formulations. The MMP-PDP is solved using an ALNS algorithm which is validated using CPLEX. We perform a series of experiments to study the impacts of the different vehicle operation types under different temporal and spatial demand configurations. Additionally, a parameter study is performed in order to understand the sensitivity of the results to vehicle capacity, vehicle range, and cost savings due to platoon formation.

The results suggest that cost savings of 48% can be achieved with modular vehicles, while an additional 9% can be saved when consolidating passenger and freight demand. These cost reductions are mainly due to the reduction in operating costs for the platoons and reduction of trip duration. The cost savings are insensitive to the demand distribution in space (clustered or distributed) and time (even or peak). The average platoon length for modular operations is 1.8 without consolidation and 2.1 with consolidation. In this study we show that the distance traveled as well as the number of modules required for different scenarios are slightly reduced with modular operations. Furthermore, empty vehicle kilometers can be significantly reduced, indicating a more efficient use of available capacity. The total served requests stay unchanged, representing an unchanged level of service. The results do not suggest additional benefits under uneven temporal demand distributions compared to even demand distributions. Finally, the results indicate that the delivery of freight requests contributes significantly more to the costs than passenger requests, and most cost savings can be achieved through the change of delivery order for freight requests. The combination of requiring fewer modules, lower operational costs, fewer platoon kilometers, and the reduction of empty kilometers imply an increase in efficiency for the transport system. Moreover, a reduction in emissions is expected as a direct consequence of the reduced platoon kilometers.

The results of the parameter study show that the range parameter does not have a significant effect on the objective value. However, the total cost can be further reduced when increasing the module capacity. This is mainly due to lower operating costs, while the number of modules, the length of the platoon, and the empty kilometers all decrease. Furthermore, the fill rate is reduced for high platoon incentives, indicating a less efficient use of the supply. An increase in module capacity leads to a reduction in total costs for all vehicle operations. However, conventional operations are more sensitive to changes in module capacity. For modular operations, the benefits of larger modules are smaller and saturate at a module capacity of approximately 35 units for the scenarios studied. This cost reduction stems mainly from fewer modules required to serve the same number of requests.

The results of the case study show similar results to the scenario analysis. The objective value can be reduced when operating modular vehicles and is further reduced when consolidating passenger and freight demand. This is mainly due to the reduced fleet size cost and the reduction in trip times. Due to the higher demand and larger scale of this scenario modules form larger platoons and the platoon kilometer are higher however this does not affect the improved performance of the modular transport system.

The proposed model and the adjusted ALNS algorithm are generally applicable to similar transportation systems. By adjusting the input data and parameter settings, the proposed approach can be used by practitioners and decision makers to evaluate the impact of modular vehicle operations on a specific transportation system. Additionally, the general formulation makes it possible to evaluate different policies by adjusting the cost parameter. Furthermore, the proposed model can easily be adjusted to consider other forms of consolidation, e.g. to study forward and reverse flows, including goods, parcels, bulk deliveries and recycling.

The study identifies several relevant findings and challenges for practitioners resulting from modular vehicle operations. First, the study shows that the cost savings of modular vehicle operations are less sensitive to vehicle capacity than conventional vehicle operations. This indicates that there is a reduced need for heterogeneous fleet sizes, which in turn simplifies the maintenance, operation, and procurement of such vehicle fleets. Second, due to the result that most cost savings are achieved with short platoon length, the impacts on traffic and changes in the road infrastructure (e.g. larger curbside space when un-/loading) are minimal when switching the transportation operations to modular vehicles. Third, since the main benefits of modular operations stem from the reduction of platoon kilometers by transporting passengers along with freight requests, a successful transition towards modular vehicle operations can be achieved by selectively adding passenger requests to the already served freight demand.

The challenges of modular vehicle operations are mainly related to the formation of platoons at depots, which requires more time and space compared to conventional vehicle operations. Hence potentially larger and more central depots are needed, which requires approval from city authorities. Additionally, long platoons might obstruct surrounding traffic and pedestrian movements when maneuvered in dense urban environments. Additional consequences of the proposed transportation system affect the design and operation of conventional public transport stops, which could be combined with freight delivery stations to multimodal terminals. Furthermore, the combined optimization of freight and passenger demand requires the integrated planning by multiple stakeholders, such as public transport operators and freight carriers, which traditionally have not been synchronized. All in all, there is a need for planning processes and policies that support the benefits of the modular vehicles while minimizing the obstructions of ongoing traffic.

The modularity of the proposed model is limited mainly by one assumption made in the problem description. The platoon configuration is required to stay unchanged throughout the route; hence, en-route platoon reconfiguration is not modeled. On one hand this limits the theoretical benefits of the vehicle technology, one the other hand this assumption increases the practical relevance of the model. Since en-route reconfiguration is not possible with current technologies such systems cannot be operated in a real-world scenarios. Vehicle operations as proposed in this study can be operated in a real-world scenario since there is the technology needed and no changes in infrastructure or legislation are required.

The results presented in this study should be considered as scenario specific. Mainly, the values for the optimal capacity and range-related impacts can be different for different input data. Another limitation of the current study is that no distinction is made between passenger and freight travel time costs. Such an extension is straightforward when evidence supports it. Furthermore, the presented results depend on the assumed cost terms for vehicle fixed costs and operational costs, which are uncertain and likely to change in the future, thereby affecting cost savings and efficiency gains. If vehicle fixed costs can be reduced in the future, then cost savings might be further increased. However, if we assume a reduction in vehicle operating costs, the benefits of the proposed modular system will be reduced and will mainly stem from the consolidation of demand.

Analyzing data from pilot projects which utilize the assumed modules and operations will allow to refine these parameter settings in the future. Furthermore, we assume that all demand is known in advance. For the practical application of the model additional online planning steps which consider e.g. stochastic demand or new requests during the day would have to be added to the framework. Since most of freight and passenger demand is known or can be estimated before the planning period, we expect the effect of considering such unknown demand on the benefits of the modular system to be minimal. Nevertheless this assumption does influence the solutions generated and reduce their practical relevance. An extension of the proposed framework considering online planning as a post-processing step should therefore be considered in future studies.

Future research may extend the proposed model by allowing for dynamic transshipments of requests between different modules and by allowing for en-route platoon reconfiguration. Both of these features would allow further exploiting the flexibility and modularity of the modules, which is likely to further reduce the total costs. The proposed model can be further extended by adding multi-echelon delivery. By replicating consolidation centers and the possibility to model the interaction between different shipping operators. Another future research direction is to formulate the problem using time-expanded networks (e.g., Tong et al., 2017; Vu et al., 2020; Neumann-Saavedra et al., 2021). Such a formulation might be helpful in solving larger instances with complex time window constraints and enable the modeling of time-dependent travel times. Finally, in future studies demand patterns including return flows, and additional types of demand (e.g. waste, recycling) can be added to the scenario definitions to study other future urban transport systems.

CRediT authorship contribution statement

Jonas Hatzenbühler: Formal analysis, Writing – review & editing, Conceptualization, Methodology, Data curation, Writing – original draft. **Erik Jenelius:** Formal analysis, Conceptualization, Methodology, Supervision, Project administration, Writing – review & editing. **Gyözö Gidófalvi:** Formal analysis, Conceptualization, Methodology, Supervision, Project administration, Writing – review & editing. **Oded Cats:** Formal analysis, Conceptualization, Methodology, Supervision, Writing – review & editing.

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.

Acknowledgments

This work was in part supported by the Sustainable and Integrated Transport Systems (Hållbara och Integrerade Transport System HITS) project under grant no. 2020-00565 (Vinnova) through KTH Integrated Transport Research Lab (ITRL). The work was in part supported by Strategic Research Area TRENoP. The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at HPC2N partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

References

Aggoune-Mtalaa, W., Habbas, Z., Ouahmed, A.A., Khadraoui, D., 2015. Solving new Urban freight distribution problems involving modular electric vehicles. IET Intell. Transp. Syst. 9 (6), 654–661. http://dx.doi.org/10.1049/iet-its.2014.0212.

Aloui, A., Hamani, N., Derrouiche, R., Delahoche, L., 2021. Systematic literature review on collaborative sustainable transportation: Overview, analysis and perspectives. Transp. Res. Interdiscip. Perspect. 9, 100291. http://dx.doi.org/10.1016/j.trip.2020.100291.

Behiri, W., Belmokhtar-Berraf, S., Chu, C., 2018. Urban freight transport using passenger rail network: Scientific issues and quantitative analysis. Transp. Res. E Logist. Transp. Rev. 115, 227–245. http://dx.doi.org/10.1016/j.tre.2018.05.002.

Bhoopalam, A.K., Agatz, N., Zuidwijk, R., 2018. Planning of truck platoons: A literature review and directions for future research. Transp. Res. B 107, 212–228. http://dx.doi.org/10.1016/j.trb.2017.10.016.

Börjesson, M., Eliasson, J., 2014. Experiences from the Swedish Value of Time study. Transp. Res. A Policy Pract. 59, 144–158. http://dx.doi.org/10.1016/j.tra.2013.10.022.

Chen, Z., Li, X., 2021. Designing corridor systems with modular autonomous vehicles enabling station-wise docking: Discrete modeling method. Transp. Res. E Logist. Transp. Rev. 152, 102388. http://dx.doi.org/10.1016/j.tre.2021.102388.

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. http://dx.doi.org/10.1287/trsc.2021.1085, URL: https://pubsonline.informs.org/doi/10.1287/trsc.2021.1085.

Cleophas, C., Cottrill, C., Ehmke, J.F., Tierney, K., 2019. Collaborative urban transportation: Recent advances in theory and practice. European J. Oper. Res. 273 (3), 801–816. http://dx.doi.org/10.1016/j.ejor.2018.04.037.

Cochrane, K., Saxe, S., Roorda, M.J., Shalaby, A., 2017. Moving freight on public transit: Best practices, challenges, and opportunities. Int. J. Sustain. Transp. 11 (2), 120–132. http://dx.doi.org/10.1080/15568318.2016.1197349.

Dakic, I., Yang, K., Menendez, M., Chow, J.Y.J., 2021. On the design of an optimal flexible bus dispatching system with modular bus units: Using the three-dimensional macroscopic fundamental diagram. Transp. Res. B 148, 38–59. http://dx.doi.org/10.1016/j.trb.2021.04.005.

Dantzig, G.B., Ramser, J.H., 1959. The truck dispatching problem. Manage. Sci. 6 (1), 80–91. http://dx.doi.org/10.1287/mnsc.6.1.80.

Derigs, U., Pullmann, M., Vogel, U., 2013. Truck and trailer routing—Problems, heuristics and computational experience. Comput. Oper. Res. 40 (2), 536–546. http://dx.doi.org/10.1016/j.cor.2012.08.007.

Desrochers, M., Lenstra, J., Savelsbergh, M., Soumis, F., 1988. Vehicle routing with time windows: Optimization and approximation. Veh. Rout. Methods Stud. 16.

Dündar, H., Ömürgönülşen, M., Soysal, M., 2021. A review on sustainable urban vehicle routing. J. Clean. Prod. 285, 125444. http://dx.doi.org/10.1016/j.jclepro.2020.125444.

EU, 2022. E-commerce statistics for individuals. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics_for_individuals.

Ferrell, W., Ellis, K., Kaminsky, P., Rainwater, C., 2020. Horizontal collaboration: Opportunities for improved logistics planning. Int. J. Prod. Res. 58 (14), 4267–4284. http://dx.doi.org/10.1080/00207543.2019.1651457.

Fu, Z., Chow, J.Y.J., 2022. The pickup and delivery problem with synchronized en-route transfers for microtransit planning. Transp. Res. E Logist. Transp. Rev. 157, 102562. http://dx.doi.org/10.1016/j.tre.2021.102562.

Gansterer, M., Hartl, R.F., 2018. Collaborative vehicle routing: A survey. European J. Oper. Res. 268 (1), 1–12. http://dx.doi.org/10.1016/j.ejor.2017.10.023.

Ghilas, V., Demir, E., Van Woensel, T., 2016. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows and scheduled lines. Comput. Oper. Res. 72, 12–30. http://dx.doi.org/10.1016/j.cor.2016.01.018.

Gong, S., Du, L., 2018. Cooperative platoon control for a mixed traffic flow including human drive vehicles and connected and autonomous vehicles. Transp. Res. B 116, 25–61. http://dx.doi.org/10.1016/j.trb.2018.07.005.

Gong, M., Hu, Y., Chen, Z., Li, X., 2021. Transfer-based customized modular bus system design with passenger-route assignment optimization. Transp. Res. E Logist. Transp. Rev. 153, 102422. http://dx.doi.org/10.1016/j.tre.2021.102422.

Guajardo, M., Rönnqvist, M., 2016. A review on cost allocation methods in collaborative transportation. Int. Trans. Oper. Res. 23 (3), 371–392. http://dx.doi.org/10.1111/itor.12205, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/itor.12205.

Hatzenbühler, J., Jenelius, E., Gidófalvi, G., Cats, O., 2022. Multi-purpose pickup and delivery problem for combined passenger and freight transport. Transportation (Under Review).

Koç, Ç., Bektaş, T., Jabali, O., Laporte, G., 2016. Thirty years of heterogeneous vehicle routing. European J. Oper. Res. 249 (1), 1–21. http://dx.doi.org/10.1016/j.ejor.2015.07.020.

Lenstra, J.K., Kan, A.H.G.R., 1981. Complexity of vehicle routing and scheduling problems. Networks 11 (2), 221–227. http://dx.doi.org/10.1002/net.3230110211.

Li, B., Krushinsky, D., Van Woensel, T., Reijers, H.A., 2016. An adaptive large neighborhood search heuristic for the share-a-ride problem. Comput. Oper. Res. 66, 170–180. http://dx.doi.org/10.1016/j.cor.2015.08.008.

Li, Z., Shalaby, A., Roorda, M.J., Mao, B., 2021. Urban rail service design for collaborative passenger and freight transport. Transp. Res. E Logist. Transp. Rev. 147, 102205. http://dx.doi.org/10.1016/j.tre.2020.102205.

Liang, K.-Y., Mårtensson, J., Johansson, K.H., 2014. Fuel-saving potentials of platooning evaluated through sparse heavy-duty vehicle position data. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings. pp. 1061–1068. http://dx.doi.org/10.1109/IVS.2014.6856540.

Lin, J., Nie, Y., Kawamura, K., 2022. An autonomous modular mobility paradigm. IEEE Intell. Transp. Syst. Mag. 2–10. http://dx.doi.org/10.1109/MITS.2022.3159484.

Liu, X., Qu, X., Ma, X., 2021. Improving flex-route transit services with modular autonomous vehicles. Transp. Res. E Logist. Transp. Rev. 149, 102331. http://dx.doi.org/10.1016/j.tre.2021.102331.

Los, J., Schulte, F., Spaan, M.T.J., Negenborn, R.R., 2020. The value of information sharing for platform-based collaborative vehicle routing. Transp. Res. E Logist. Transp. Rev. 141, 102011. http://dx.doi.org/10.1016/j.tre.2020.102011.

Lozano, S., Moreno, P., Adenso-Díaz, B., Algaba, E., 2013. Cooperative game theory approach to allocating benefits of horizontal cooperation. European J. Oper. Res. 229 (2), 444–452. http://dx.doi.org/10.1016/j.ejor.2013.02.034.

Masson, R., Lehuédé, F., Péton, O., 2013. An adaptive large neighborhood search for the pickup and delivery problem with transfers. Transp. Sci. 47 (3), 344–355. http://dx.doi.org/10.1287/trsc.1120.0432.

Mello Bandeira, R.A., Goes, G.V., Schmitz Gonçalves, D.N., D'Agosto, M.d.A., de Oliveira, C.M., 2019. Electric vehicles in the last mile of urban freight transportation: A sustainability assessment of postal deliveries in Rio de Janeiro-Brazil. Transp. Res. D Transp. Environ. 67, 491–502. http://dx.doi.org/10.1016/j.trd.2018.12.017.

Militão, A.M., Tirachini, A., 2021. Optimal fleet size for a shared demand-responsive transport system with human-driven vs automated vehicles: A total cost minimization approach. Transp. Res. A Policy Pract. 151, 52–80. http://dx.doi.org/10.1016/j.tra.2021.07.004.

Mourad, A., Puchinger, J., Woensel, T.V., 2020. Integrating autonomous delivery service into a passenger transportation system. Int. J. Prod. Res. 1–24. http://dx.doi.org/10.1080/00207543.2020.1746850.

Münster, M., Kopp, G., Friedrich, H.E., Siefkes, T., 2020. Autonomous vehicle concept for the urban transport of the future. ATZ Worldwide 122 (3), 26–31. http://dx.doi.org/10.1007/s38311-020-0202-4.

Neumann-Saavedra, B.A., Mattfeld, D.C., Hewitt, M., 2021. Assessing the operational impact of tactical planning models for bike-sharing redistribution. Transp. Res. A 150, 216–235. http://dx.doi.org/10.1016/j.tra.2021.06.003.

NEXT, 2022. HOME. NextFutureTransport. https://www.next-future-mobility.com.

Nourmohammadzadeh, A., Hartmann, S., 2016. The fuel-efficient platooning of heavy duty vehicles by mathematical programming and genetic algorithm. In: Martín-Vide, C., Mizuki, T., Vega-Rodríguez, M.A. (Eds.), Theory and Practice of Natural Computing. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 46–57. http://dx.doi.org/10.1007/978-3-319-49001-4_4.

OpenStreetMap contributors, 2017. Planet dump retrieved from. https://planet.osm.org. https://www.openstreetmap.org.

Ozturk, O., Patrick, J., 2018. An optimization model for freight transport using urban rail transit. European J. Oper. Res. 267 (3), 1110–1121. http://dx.doi.org/10.1016/j.ejor.2017.12.010.

Pei, M., Lin, P., Du, J., Li, X., Chen, Z., 2021. Vehicle dispatching in modular transit networks: A mixed-integer nonlinear programming model. Transp. Res. E Logist. Transp. Rev. 147, 102240. http://dx.doi.org/10.1016/j.tre.2021.102240.

Pisinger, D., Røpke, S., 2010. Large neighborhood search. In: Gendreau, M. (Ed.), Handbook of Metaheuristics, second ed. Springer, 399420.

Ronald, N., Yang, J., Thompson, R.G., 2016. Exploring co-modality using on-demand transport systems. Tenth International Conference on City Logistics 17-19 June 2015, Tenerife, Spain, Transp. Res. Procedia Tenth International Conference on City Logistics 17-19 June 2015, Tenerife, Spain, 12, 203–212.http://dx.doi.org/10.1016/j.trpro.2016.02.059,

Ropke, S., Pisinger, D., 2006. An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows. Transp. Sci. 40 (4), 455–472. http://dx.doi.org/10.1287/trsc.1050.0135.

Sachs, J., Kroll, C., Lafortune, G., Fuller, G., Woelm, F., 2021. Sustainable Development Report 2021, first ed. Cambridge University Press, http://dx.doi.org/10.1017/9781009106559.

Savelsbergh, M.W.P., Sol, M., 1995. The general pickup and delivery problem. Transp. Sci. 29 (1), 17–29. http://dx.doi.org/10.1287/trsc.29.1.17.

Scania, 2020. NXT concept vehicle represents a vision of the future. Scania. https://www.scania.com/uk/en/home/experience-scania/features/nxt-concept-vehicle-represents-a-vision-of-the-future.html.

Schlenther, T., Martins-Turner, K., Bischoff, J.F., Nagel, K., 2020. Potential of private autonomous vehicles for parcel delivery. Transp. Res. Rec. 2674 (11), 520–531. http://dx.doi.org/10.1177/0361198120949878.

Shaw, P., 1997. A new local search algorithm providing high quality solutions to vehicle routing problems.

Sokolov, V., Larson, J., Munson, T., Auld, J., Karbowski, D., 2017. Platoon formation maximization through centralized routing and departure time coordination. [Cs]. arXiv:1701.01391.

SteadieSeifi, M., Dellaert, N.P., Nuijten, W., Van Woensel, T., Raoufi, R., 2014. Multimodal freight transportation planning: A literature review. European J. Oper. Res. 233 (1), 1–15. http://dx.doi.org/10.1016/j.ejor.2013.06.055.

Tong, L., Zhou, L., Liu, J., Zhou, X., 2017. Customized bus service design for jointly optimizing passenger-to-vehicle assignment and vehicle routing. Transp. Res. C 85, 451–475. http://dx.doi.org/10.1016/j.trc.2017.09.022.

Toth, P., Vigo, D. (Eds.), 2014. Vehicle routing: Problems, methods, and applications, second ed. In: MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics; Mathematical Optimization Society, Philadelphia.

Vargas, A., Fuster, C., Corne, D., 2020. Towards sustainable collaborative logistics using specialist planning algorithms and a gain-sharing business model: A UK case study. Sustainability 12 (16), 6627. http://dx.doi.org/10.3390/su12166627.

Vu, D.M., Hewitt, M., Boland, N., Savelsbergh, M., 2020. Dynamic discretization discovery for solving the time-dependent traveling salesman problem with time windows. Transp. Sci. 54 (3), 703–720. http://dx.doi.org/10.1287/trsc.2019.0911.

Wang, J., Yu, Y., Tang, J., 2018a. Compensation and profit distribution for cooperative green pickup and delivery problem. Transp. Res. B 113, 54–69. http://dx.doi.org/10.1016/j.trb.2018.05.003.

Wang, Y., Zhang, J., Assogba, K., Liu, Y., Xu, M., Wang, Y., 2018b. Collaboration and transportation resource sharing in multiple centers vehicle routing optimization with delivery and pickup. Knowl.-Based Syst. 160, 296–310. http://dx.doi.org/10.1016/j.knosys.2018.07.024.

Wen, J., Nassir, N., Zhao, J., 2019. Value of demand information in autonomous mobility-on-demand systems. Transp. Res. A Policy Pract. 121, 346–359. http://dx.doi.org/10.1016/j.tra.2019.01.018.

Winter, K., Cats, O., Correia, G., van Arem, B., 2018. Performance analysis and fleet requirements of automated demand-responsive transport systems as an urban public transport service. Int. J. Transp. Sci. Technol. http://dx.doi.org/10.1016/j.ijtst.2018.04.004.

Wolfinger, D., 2021. A large neighborhood search for the pickup and delivery problem with time windows, split loads and transshipments. Comput. Oper. Res. 126, 105110. http://dx.doi.org/10.1016/j.cor.2020.105110.

Zhang, W., Jenelius, E., Badia, H., 2019. Efficiency of semi-autonomous and fully autonomous bus services in trunk-and-branches networks. J. Adv. Transp. 2019 (5), 1–17. http://dx.doi.org/10.1155/2019/7648735.

Zhang, Z., Tafreshian, A., Masoud, N., 2020. Modular transit: Using autonomy and modularity to improve performance in public transportation. Transp. Res. E Logist. Transp. Rev. 141, 102033. http://dx.doi.org/10.1016/j.tre.2020.102033.