Data-driven distributionally robust timetabling and dynamic-capacity allocation for automated bus systems with modular vehicles

Xia, Dongyang; Ma, Jihui; Sharif Azadeh, Sh.; Zhang, Wenyi

DOI
10.1016/j.trc.2023.104314

Publication date
2023

Document Version
Final published version

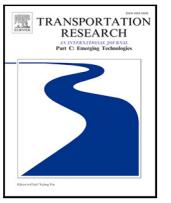
Published in
Transportation Research Part C: Emerging Technologies

Transportation Research Part C 155 (2023) 104314

Contents lists available at ScienceDirect

Transportation Research Part C

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



Data-driven distributionally robust timetabling and dynamic-capacity allocation for automated bus systems with modular vehicles

Dongyang Xia [a], Jihui Ma [a], Sh. Sharif Azadeh [b], Wenyi Zhang [a,*]

[a] Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China

[b] Department of Transport & Planning, Delft University of Technology, Netherlands

ARTICLE INFO

Keywords:
Bus timetabling
Modular vehicles
Uncertainty and time-dependency
Distributionally robust optimization
Integer L-shaped method

ABSTRACT

The collaborative design of the timetable and dynamic-capacity allocation plan of emerging modular vehicles (MVs) is a promising solution to the mismatch between supply and demand in public transportation studies; however, such efforts are subject to high-level dynamics and uncertainty inherent in operating environments. In this study, we focus on the timetabling and dynamic-capacity allocation problem of MVs within the context of distributionally robust optimization under time-dependent demand uncertainty. The dynamic capacity refers to the number of modular units (MUs) comprising an MV can be potentially changed at different times and stops. A Wasserstein distance-based ambiguity set with a time-dependent and station-wise perturbation parameter is adopted to incorporate all potential distributions within a 1-Wasserstein distance for addressing the uncertainty of passenger demand. Further, a data-driven distributionally robust optimization model that considers time-varying capacity is formulated to minimize passenger waiting costs and dispatching costs of operators over all possible demand distributions within the ambiguity set. Subsequently, an expansion that allows for flexible formations of MVs assigned to each trip at each stop is proposed, and this results in more customized operational plans driven by the passenger demand. To improve the computational efficiency of realistic problems, we design a customized integer L-shaped method to exactly solve the models, which incorporates a class of valid equalities to further speed up the computation. The effectiveness of the proposed approaches in reducing the costs for both passengers and operators compared with the practical fixed-capacity operations is verified by real-world case studies based on the operating data of Beijing Bus Line 468. Furthermore, the superiority of the distributionally robust optimization method in comparison to the stochastic programming and the robust optimization approaches is demonstrated.

1. Introduction

Current metropolitan public transportation systems employ fixed-capacity vehicles; however, this makes it challenging to match uneven passenger demand over time and space. This imbalance between supply and demand has led to an excessive number of vehicles operating during off-peak periods and a shortage of vehicles during peak periods. Fortunately, automated public transport systems have emerged recently, and they are highly favored owing to their potential to overcome the above-mentioned problems and

* Corresponding author.
E-mail address: wyzhang@bjtu.edu.cn (W. Zhang).

https://doi.org/10.1016/j.trc.2023.104314
Received 26 October 2022; Received in revised form 4 July 2023; Accepted 19 August 2023
Available online 30 August 2023

(a) Next Future Transportation Inc. (b) Ohmio LIFT.

Fig. 1. Modular vehicles designed by the Next Future Transportation Inc. and Ohmio LIFT.

improve the level of service, while minimizing the social costs of public transportation services (Sadrani et al., 2022b). In addition, the modular vehicle (MV) is an emerging technology that has been proven to be promising by many companies, such as Next Future Transportation Inc. (NEXT Future Transportation Inc., 2022) and Ohmio LIFT (Ohmio, 2023), and tested in Dubai (Spera, 2016; NEXT Future Transportation Inc., 2023). The MVs developed by Next Future Transportation Inc. have the following characteristics. (1) Each MV is comprised of multiple MUs, and the number of MUs contained in an MV is called the formation of that MV in this study. For example, MVs with formations 2 and 3 are shown in Fig. 1. (2) Each MU is capable of docking with other MUs at speed on public roads and can be unlocked smoothly at full speed. Additionally, these operations can be performed relying solely on MU's own device (NEXT Future Transportation Inc., 2022). (3) It enables passengers to stand and walk from one MU to another during travel (NEXT Future Transportation Inc., 2018). Since MVs possess vehicle autonomy, they can dynamically adjust the capacity through the direct physical docking and undocking of MUs at the terminals or at each stop, which is referred to as *time-varying capacity* and *station-wise docking* in Shi and Li (2021) and Chen and Li (2021), respectively. Specifically, if vehicles can only change formations at terminals, then the capacity is dynamically allocated over time; the capacity becomes dynamic in spatial and temporal terms if formations can be modified flexibly at each stop. Both these problems are collectively referred to as the *dynamic-capacity allocation problem of modular vehicles* in this study.

In the field of mobility and transport, technologies are being emerged at a much faster pace than the methodological adaptation. These newly deployed tools of transport come with their own assumptions and complexities, which call for new methods to tackle them. The timetabling and dynamic-capacity allocation (TT-DCA) problem of modular vehicles has naturally become a popular research area. A majority of the relevant research focuses on optimization methods under deterministic conditions; however, as an open and complex mega-system, urban public transport (especially bus systems operating on the road) suffers from a considerable number of uncertainties. Various factors can trigger uncertainties, e.g., weather, signal lights, operating dates, development of other transportation modes (Ma et al., 2019). Operational plans are drafted a priori (i.e., before passenger demand is realized), and therefore, it is necessary to consider uncertainties when designing operational plans (e.g., timetables and dynamic-capacity allocation plans). To the best of our knowledge, Sadrani et al. (2022b) is the most recent research that aims to optimize the service frequency and vehicle size for automated bus systems under uncertainty by using a stochastic programming (SP) optimization method. Stochastic programming is a popular method to address problems with uncertainty; it assumes that the probability distribution of the uncertain parameters is completely known in advance (Liu et al., 2022). The robust optimization (RO) approach is another well-recognized methodology, which has been commonly used in robust timetabling problem (e.g., Goerigk and Schöbel, 2014; Polinder et al., 2019). The RO approach restricts the uncertain parameters to a given support set without any distributional assumptions and performs optimizations for the worst-case scenario over all realizations of that support set. A problem of the RO approach is that it is too conservative because it focuses only on the worst-case scenario. The most reliable strategy can be generated by applying the RO approach to the TT-DCA problem of MVs; however, this will simultaneously, and inevitably, lead to wasted capacity because the demand scenario on which the optimal solution is generated rarely occurs in practice. To sum up, in such systems, we deal with a great deal of uncertainties. However, classic methods are very restrictive as they heavily rely on the assumptions for the sake of tractability.

An emerging optimization approach, namely distributionally robust optimization (DRO), is one of the first attempts to release the already existing problems from restrictive assumptions and shine light on how these methods can be adapted for a larger network of such autonomous and modular vehicles. Even though the technology of modular vehicles is on the verge of deployment at this "proof of concept" stage, more realistic models make operationalization of this technology more efficient. DRO lies at the intersection of the RO and SP methods (Agra and Rodrigues, 2022). The DRO method assumes that information about the probability distribution of uncertain parameters is partially known, and it portrays all potential distributions by constructing an ambiguity set where the true distribution is assumed to belong to the ambiguity set. The worst-case expected cost is minimized over the ambiguity set (Delage and Ye, 2010). Among all types of ambiguity sets, the Wasserstein distance-based ambiguity set has recently become a concern. This ambiguity set consists of all probability distributions whose Wasserstein distance, representing the distance between any two distributions, from the nominal distribution, is constrained by a parameter. There are two reasons behind the growing significance of the Wasserstein distance-based ambiguity set: (i) the Wasserstein distance-based ambiguity set has high practicability, which allows

decision makers to flexibly control the conservativeness of the optimal solutions by adjusting only a single parameter (Agra and Rodrigues, 2022), and (ii) it shows superior out-of-sample performance. The DRO models based on the Wasserstein distance-based ambiguity set can integrate numerous sampled data into the optimization model directly, and therefore, they are often referred to as data-driven DRO methods (Mevissen et al., 2013). The DRO method can mitigate over-conservatism while enhancing the robustness of the optimal solutions because the precise probability distribution of uncertain parameters is difficult to obtain in real-world operations; this can help create a trade-off between operators and passengers. Nevertheless, one of its limitations is that, since a set of probability distributions is considered when employing this method, its models for large-scale problems can be quite challenging to solve, especially when the uncertainty is of high dimension (Cheramin et al., 2022).

To bridge these gaps, this study explores the application of data-driven DRO to the TT-DCA problem of MVs with time-dependent travel times and passenger demands. We focus on uncertain passenger flows and aim to design a robust timetable and dynamic-capacity allocation plan by optimizing the shifting time of each MV at the first stop, the holding time at each stop, and the time-varying and station-wise capacity adjustment strategies. In addition, we provide managerial insights through comparisons between the fixed-capacity and dynamic-capacity operations, as well as establish a comparison between SP and RO to explore the superiority of DRO. Further, we investigate a decomposition algorithm that can efficiently derive high-quality solutions. The tight coupling between the timetable, and the station-wise and time-varying capacity adjustment strategies of MVs, are decomposed, which helps reduce computational intensity.

1.1. Literature review

We briefly review studies that are most relevant to our work, including the timetabling and vehicle scheduling with variable capacity in a heterogeneous fleet, timetabling and dynamic-capacity allocation of MVs in automated public transport systems, and the DRO method.

Over the past few decades, a considerable amount of research has focused on the conventional demand-oriented timetabling problem with a fixed bus capacity to better match the available supply with passenger demand (e.g., Newell and Potts, 1964; Sánchez-Martíneza et al., 2016; Robenek et al., 2016; Wang et al., 2017; Gkiotsalitis and Alesiani, 2019; Wu et al., 2019; Liu et al., 2021; Zhang et al., 2021a). However, most of these studies base their conclusions on the ideal assumption that travel times and/or passenger demand are static and/or deterministic. Recently, several studies have focused on the timetabling and vehicle scheduling problem with variable capacity in a heterogeneous fleet, aiming to enhance the matching of supply and demand and decrease the waste of capacity. The variable capacity in a heterogeneous fleet refers to a fleet of vehicles with multiple sizes. The capacity of each vehicle is fixed but can be different from the others. The potential of operating various sizes of vehicles during different periods to suit the passenger demand at that time has already been demonstrated by Liu and Ceder (2016), Peña et al. (2019), Visentini et al. (2019), and Tang et al. (2023). For example, Liu and Ceder (2016) formulated a two-objective mathematical model for the collaborative optimization of timetables and vehicle sizes, where one objective is to minimize the load discrepancy from a desired occupancy level on vehicles and the other is to minimize the total waiting time of passengers. The results reveal that, through the proposed collaborative optimization method, the two objective values can both be reduced compared with those of the current timetables. Nevertheless, passengers are also typically unevenly distributed spatially. Matching spatially unbalanced passenger flows is challenging because the inability to alter the formations of en-route vehicles in the aforementioned heterogeneous fleet with multiple vehicle sizes during operations.

Fortunately, the emerging MV technology show considerable potential towards achieving station-wise and time-varying capacities by allowing vehicles to be disassembled and assembled at specific stations. Compared to the fleet with multiple vehicle sizes, MVs provide a greater flexibility for adjusting the capacity of individual vehicle by adding or removing MUs quickly. With MVs, it is possible to optimize the formation of each vehicle based on the unevenly distributed passenger flows in spatial and temporal terms. Moreover, MVs offer the advantage of flexible utilization of MUs across different lines due to the convenience of docking and undocking at stops and during traveling. For example, one MU in an MV can be undocked at a transfer stop, and then utilized by another line with higher passenger demand that serves this stop. These flexibilities allow for the efficient allocation of vehicles based on demand and reduce operating costs. The advanced advantages of MV technology have garnered significant attention from researchers. Recent studies have started concentrating on the demand-oriented TT-DCA problem of MVs to strike balance between costs incurred by both passengers and operators. For example, Chen et al. (2019) proposed a mixed-integer linear programming (MILP) model based on the discrete modeling method, wherein the decisions include whether to dispatch a train at each discrete time point and what formation of MVs to dispatch. Moreover, the authors designed a customized dynamic programming (DP) algorithm with an efficient set of inequalities to speed up computation. Furthermore, Chen et al. (2020) developed a continuum approximation (CA) model based on the continuous modeling method for determining the formation and departure times of each MV considering that the discrete modeling approach will definitely increase computational difficulty. Dai et al. (2020) proposed an integer nonlinear programming model to balance the trade-off between the operating costs and costs of increased passenger waiting times by considering the joint design problem of the bus dispatch headway and capacity for transit systems operating with mixed human-driven and autonomous buses. With the aim of increasing the number of served trip requests and decreasing the total vehicles-miles-traveled, Zhang et al. (2020) formulated two ILP models for the problem by utilizing MVs to serve the first- and last-mile trips. Further, Shi and Li (2021) established an MILP model for the joint design problem of departure frequency and formation of MVs, which can be solved efficiently using the DP algorithm. Considering stochastic travel times and 15-min-dependent demand, Sadrani et al. (2022b) proposed a nonlinear programming (NLP) model to determine the service frequency and vehicle size

in automated bus systems. The authors designed a full enumeration (FE) algorithm that can incorporate the Monte Carlo simulation (MCS) approach to obtain a high-quality globally optimal solution.

All the aforementioned studies permit the assembly and disassembly of MVs only at the terminals. Furthermore, some recent studies have enabled the station-wise docking of MVs: vehicles can now change their formations at specific stations along the line to better meet passenger demand with temporally dynamic and spatially uneven characteristics. Chen and Li (2021) formulated a discrete-time-based MILP model that permitted modular vehicles to adjust their formations flexibly by docking and undocking at each station. The theoretical properties were investigated, and a tailored branch-and-bound (B & B) algorithm was proposed to speed up the computation. Chen et al. (2022) employed the CA approach to model this problem for obtaining a dynamic-capacity operating plan more efficiently. A realistic passenger dynamic evolution process was integrated into their CA approach. Further, Tian et al. (2022) proposed a mixed-integer nonlinear programming formulation to determine the location and maximum capacity of these stations, as well as the optimal operational strategy of station-wise dynamic formations over space and time by assuming that the formations of MVs can be adjusted flexibly at intermediate special stations that can perform assembly/disassembly operations. Table 1 summarizes the detailed characteristics of the most related studies with respect to the TT-DCA problem.

In the DRO approach, one key aspect is the design of an ambiguity set appropriate to the problem. In the last decade, theoretical research on the DRO approach, which includes ambiguity sets, has progressed at a remarkable pace. We recommend readers follow Rahimian and Mehrotra (2019) for an up-to-date and thorough review of the theoretical aspects of DRO. Further, the DRO approach has been successfully applied in certain engineering fields to solve the inventory problem (Ardestani-Jaafari and Delage, 2016), hub location problem (Liu et al., 2019), disaster relief problem (Shehadeh and Tucker, 2020), and so on. Earlier research focused on moment-based ambiguity sets, which have poor convergence. Therefore, the recent research has actively focused on discrepancy-based ambiguity sets, particularly on the Wasserstein distance-based ambiguity set. For example, Zhang et al. (2021b) formulated a DRO model with the Wasserstein distance-based ambiguity set in response to the vehicle routing problem with time windows. They designed two algorithms (branch-and-cut algorithm and a meta-heuristic algorithm) for large-scale cases. Agra and Rodrigues (2022) proposed a two-stage DRO formulation to address the berth allocation problem under uncertain handling times; this was integrated with the Wasserstein distance-based ambiguity set. To the best of our knowledge, studies on the application of the advanced DRO method to the urban public transport timetabling problem remain limited.

In summary, the timetabling problem of traditional buses is a popular research topic, and many mathematical formulations and solutions have been explored. Given the emergence of MVs in automated bus systems, the TT-DCA problem has become a new research topic in recent years. However, an in-depth analysis of the collaborative optimization of multiple dynamic operational strategies remains lacking because of the complexity of formulating coupling relations between stochastic and/or time-dependent parameters. For example, time-varying connections between travel times and passenger demand at 1-min time granularity, station-wise capacity, and station-varying headway need to be considered to make precise decisions on the flexible-capacity allocation plan. Furthermore, the relevant mathematical models are typically solved using simulation-based or heuristic algorithms because of the complexity of data-driven DRO problems. Recent studies (e.g., Gkiotsalitis and Alesiani, 2019; Sadrani et al., 2022b) have considered the timetabling problem under stochastic conditions. However, these studies consider uncertain passenger demand as an invariant value within 15 min or 1 h, and the capacity is fixed among all stops along the line. Therefore, extended methodologies that can help take decisions regarding the capacity of each MV at each stop are subject to further research. In addition, studies on the TT-DCA problem of MVs have primarily focused on the transit corridor, and thus, the departure frequency and dynamic vehicle formations are key decision variables. Other strategies that have been tested to effectively mitigate the "bus-bunching" (too small headway) phenomenon, such as holding at stops, have rarely been decided simultaneously.

1.2. Contributions of this paper

This paper contributes to the literature by formulating data-driven DRO models for the TT-DCA problem of MVs, which consider real-world operating characteristics of bus systems (i.e., time-dependent travel times, uncertain and time-dependent passenger demand), to search for optimal robust (1) time-varying and station-wise capacity allocation plans of MVs, and (2) the corresponding timetables. The decisions involve employing strategies in the tactical phase that can provide a scientific basis for subsequent decisions (e.g., the acquisitions of vehicles or MUs) to offer high-quality and high-reliability services. The objective is minimizing the worst-case expectation with respect to the costs of passengers and operators over all distributions residing within an ambiguity set, while improving the dynamic match between capacity and demand. Further, this study provides an efficient decomposition-based branch-and-cut algorithm for solving the proposed formulations to the optimum. The main contributions of this study to the literature are summarized as follows.

- We integrate the timetabling and dynamic-capacity allocation problem of MVs in automated bus systems into data-driven DRO models. These models employ the Wasserstein distance-based ambiguity set, allowing us to more accurately account for the uncertain and time-dependent passenger demand. By incorporating these factors, our approach contributes to a more realistic depiction of the investigated problem.
- We design a customized integer L-shaped method based on theoretical findings to solve our formulated models. Further, a family of problem-based valid equalities is proposed to further speed up the solution process.
- Extensive numerical experiments are conducted based on the operational data of Beijing Bus Line 468 to demonstrate the effectiveness and superiority of our proposed models and the exact algorithm. Computational results show that our methodologies offer considerable improvements compared with the state-of-the-art solver and practical operational plans with the fixed capacity.

Table 1
Summary of the most related recent studies.

Studies	System type	Uncertain passenger demand	Dynamic capacity adjustment strategy	Solution method	Decision variables
Sánchez-Martíneza et al. (2016)	Bus corridor	✗	✗	Rolling-horizon optimization algorithm	Holding times
Wang et al. (2017)	Bus corridor	✗	✗	Simulation-based algorithm	Departure frequency
Chen et al. (2019)	Transit shuttle	✗	MUs docked/undocked at terminals	Customized DP algorithm	(i) Time-varying formations and (ii) departure frequency
Gkiotsalitis and Alesiani (2019)	Bus corridor	✓	✗	Genetic algorithm	(i) Headway and (ii) dwell time
Wu et al. (2019)	Bus corridor	✗	✗	Heuristic algorithm	(i) Headway and (ii) slack time
Chen et al. (2020)	Transit shuttle	✗	MUs docked/undocked at terminals	Analytical solution	(i) Time-varying formations and (ii) departure frequency
Dai et al. (2020)	Bus corridor	✗	MUs docked/undocked at terminals and fixed-capacity buses	Analytical solution	(i) Time-varying formations, (ii) departure frequency, and (iii) number of dispatches
Zhang et al. (2021a)	Bus corridor	✗	✗	Constrained compass search algorithm	(i) Departure frequency and (ii) holding times
Shi and Li (2021)	Transit corridor	✗	MUs docked/undocked at terminals	Customized DP algorithm	(i) Time-dependent formations, (ii) departure frequency, and (iii) number of dispatches
Chen and Li (2021)	Transit corridor	✗	MUs docked/undocked at each stop	Customized B & B algorithm	(i) Departure frequency and (ii) station-wise and time-varying formations
Chen et al. (2022)	Transit corridor	✗	MUs docked/undocked at each stop	Analytical solution	(i) Departure frequency and (ii) station-wise and time-varying formations
Sadrani et al. (2022a)	Bus Corridor	✗	Multiple vehicle sizes	Simulation-based algorithm	(i) Departure frequency and (ii) vehicle types
Sadrani et al. (2022b)	Bus Corridor	✗	Multiple vehicle sizes	FE algorithm	(i) Departure frequency and (ii) vehicle types
Tian et al. (2022)	Transit corridor	✗	MUs docked/undocked at specially designed stops	Surrogate optimization algorithm	(i) Locations and capacities of the stations capable of assembly/disassembly operations and (ii) station-wise and time-varying formations
This paper	Bus corridor	✓	MUs docked/undocked at each stop	Customized integer L-shaped algorithm	(i) Shifting time of each MV at the first stop, (ii) station-wise and time-varying formations, and (iii) holding times

[a] The check mark (✓) indicates that the parameter or the dynamic capacity adjustment strategy is considered in the corresponding study.
[b] The cross (✗) represents that the parameter or the dynamic capacity adjustment strategy is not considered in the corresponding study.

The remainder of this paper is organized as follows. Section 2 provides a specific description of the robust TT-DCA problem of MVs with time-dependent travel times and passenger demand. In Section 3, we formally propose different mathematical formulations for the problem of interest, which enable MUs to be docked/undocked at terminals or at each stop, respectively. In Section 4, we present a customized integer-L-shaped algorithm for solving these models. Section 5 describes extensive real-world case studies. Finally, we briefly summarize the approaches and findings in Section 6 and provide suggestions for future research.

2. Problem description

Fig. 2 illustrates the investigated problem. An automated bus line with $|\mathcal{I}|$ stops is considered in this study. We denote the set of stops as $\mathcal{I} := \{1, 2, \ldots, |\mathcal{I}|\}$, and we use i, i', and i'' to index each passing stop. Let $\mathcal{K} := \{1, 2, \ldots, |\mathcal{K}|\}$ denote a set of trips that needs to be scheduled during the study time horizon. Each trip would depart from stop 1 to stop $|\mathcal{I}|$ and stop at each stop. To facilitate the scheduling process, a single MV is assigned to undertake each trip, equipped with q MUs. Here, $q \in \mathcal{Q} := \{1, 2, \ldots, |\mathcal{Q}|\}$ represents the

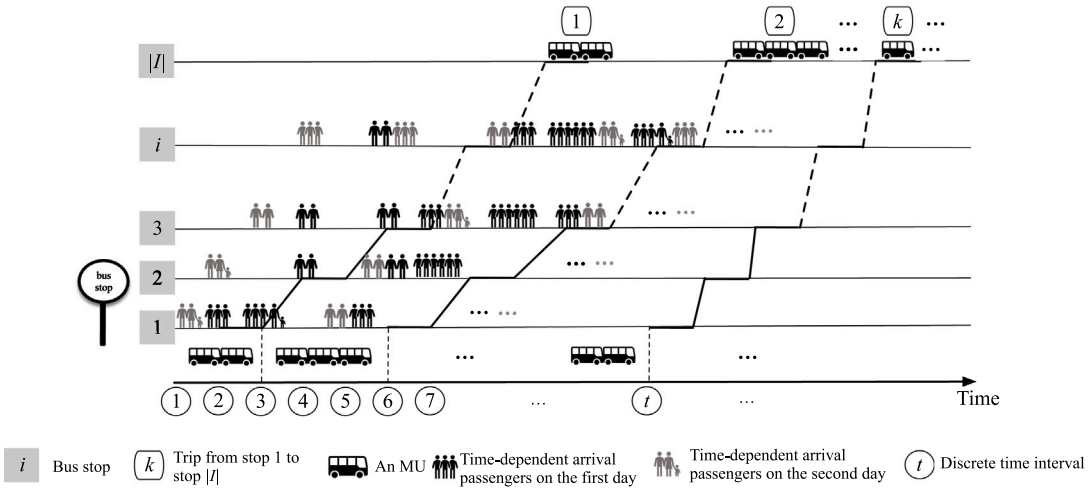
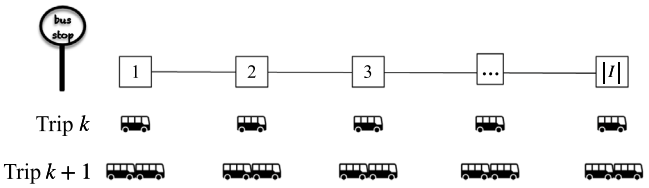


Fig. 2. Problem illustration.

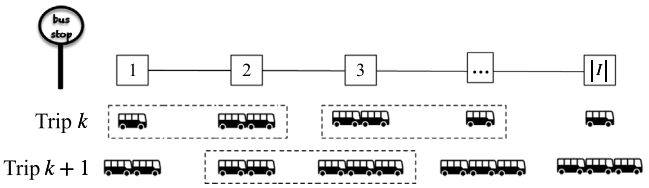
number of MUs that composes an MV, i.e., the formation of the corresponding MV. Additionally, the capacity of the MV equipped with q MUs is denoted as C_q. We assume that the minimum headway at each stop is h_{min} to ensure operational safety. Travel times between two adjacent stops are considered to be time-dependent for accurately depicting the characteristics of real-world operations. To facilitate modeling, we discretize the study time horizon into a finite number of time intervals $\mathcal{T} := \{1, 2, \dots, |\mathcal{T}|\}$, each with duration Δ. Dynamic passenger demand is also taken into account for accurate decision-making on the time-varying and station-wise capacity of MVs to reduce costs (e.g., purchase costs, operating costs, and maintenance costs). It is well known that passenger demand in urban transit systems is time-dependent and uncertain. The time-dependent characteristic refers to the different arrival rates at different time intervals on the same day; the uncertain feature means that arrival rates at the same time interval can be stochastic and may vary from day to day. For example, the black and gray personal markers are used in Fig. 2 to represent arrival rates on the first day and the second day to indicate the stochastic feature, which are time-dependent on each day. To depict these features, two time-dependent parameters, i.e., the stochastic arrival rate at stop i and time interval t and the proportion of passengers arriving at stop i, who aim to travel from stop i to stop j at time interval t, are introduced in this study, denoted as $\widetilde{\mu}_{i,t}$ and $\delta_{i,j,t}$, respectively.

Different from conventional bus systems, in the investigated automated bus systems with MVs, vehicles are able to change their formations, i.e., to add or remove MUs, with little additional time. The setup time is quite short and can be disregarded. Besides, each MU can rely on its own device for decoupling and coupling (NEXT Future Transportation Inc., 2022). For the dynamic-capacity allocation of MVs, we first consider that there is only one unique formation of the MV assigned to trip k at stops along the line, whereas the formations of MVs among various trips can differ. Next, we extend it to a more general approach that allows the MUs to be decoupled or/and coupled at each stop, i.e., the formation of an MV can be changed at each stop in accordance with the spatio-temporal imbalanced passenger demand when it undertakes a trip. Let $x_{k,q}$ denote a binary variable that takes the value of 1 when the MV assigned to trip k consists of q MUs. As shown in Fig. 3(a), the formation of the MV assigned to trip k differs from the formation of the MV assigned to trip $k+1$. However, it remains the same at any stop along the line during the trip, i.e., $\sum_{q\in Q} x_{k,q} = 1, \forall k \in \mathcal{K}$. In contrast, in Fig. 3(b), the formation of the MV assigned to trip k (i.e., $x_{k,i,q}$, for any $k \in \mathcal{K}, i \in \mathcal{I}, q \in Q$) could be different not only from $k+1$ but also from any stop $i \in \mathcal{I}$ to further match the space–time uneven passengers, as highlighted in the rectangles. Thus, an integrated optimization formulation of timetables and dynamic-capacity allocation plans, providing reliable services under uncertain passenger demand, with the objective of achieving a trade-off between the waiting costs of passengers and the operating costs of operators is proposed in this paper. Core decisions are the shifting time of each trip at the first stop, holding time of each trip at each stop, and dynamic vehicle formations assigned to each trip. Station-varying headway varies to better reflect the actual operations of buses on the road. Further, we adopt the data-driven DRO approach with the Wasserstein distance-based ambiguity set, which enhances the quality of decisions (mitigating the conservativeness of RO) of the robust operational plans of MVs by utilizing more distributional information of uncertain parameters mined from massive historical data (e.g., AFC data), while improving the reliability of decisions (enabling a more adequate use of distributional information compared with SP). We introduce the time-dependent stochastic perturbation parameter $\widetilde{\theta}_{i,t}$ for any stop $i \in \mathcal{I}$ and time interval $t \in \mathcal{T}$ into the Wasserstein distance-based ambiguity set, in order to capture the time-varying statistical properties of the uncertain passenger demand in massive amounts of data. Through this designed ambiguity set in conjunction with the problem characteristics, our formulations can provide a solid foundation for developing operational strategies with high reliability and the most suitable operational plans among diverse time periods with different passenger volumes.

Next, we make the following assumptions to reflect the general operations of automated bus systems and rigorously formulate the described problem into mathematical models.



(a) Time-varying capacity allocation.



(b) Time-varying and station-wise capacity allocation.

Fig. 3. An illustration of the investigated dynamic capacity allocation.

Assumption 1. We assume that in-service MVs are required to dwell at each stop. Similar assumptions have been adopted in previous studies (e.g., Sánchez-Martíneza et al., 2016; Zhang et al., 2021a).

Assumption 2. Following Chen and Li (2021), we assume that emerging automation technologies will enable MVs to change their formations with little extra time.

Assumption 3. Following Newell (1971) and Chen and Li (2021), we assume the number of MUs is sufficient in the depot. The fleet size can be determined after solving the investigated TT-DCA problem in the planning stage, and MUs can be procured based on the optimal solutions.

Assumption 4. We assume that passengers are permitted to redistribute within MUs to complete their journeys, as in previous studies related to the optimization problem of MVs (e.g., Tian et al., 2022; Hannoun and Menéndez, 2022). This is consistent with the service provided by NEXT Future Transportation Inc., which enables passengers to stand and walk from one MU to another during travel (NEXT Future Transportation Inc., 2018). Specifically, before the MV assigned to each trip arrives at every stop, in-vehicle passengers will walk to the corresponding MU, such as the MU that is about to be decoupled, or the MU that is continuing to move forward, in order to exit or continue to take this MV.

3. Mathematical formulations

In this section, we first introduce all the used notations in Section 3.1. Next, we introduce two SP models for the TT-DCA problem, which will be expanded by adopting a data-driven DRO approach. More specifically, in Section 3.2, an SP formulation is developed with the time-varying capacity. The extended form is then presented, which allows the MVs assigned to trips to change their formations at each stop. Finally, two novel data-driven DRO formulations are proposed and their connections with the RO and SP models are described in Section 3.3.

3.1. Notations

Throughout this paper, vectors are denoted by bold letters and $\mathbb{R}^n$ represents the set of real vectors with n components, e.g., $\mathbf{x} \in \mathbb{R}^n$. For clarity, a tilde is employed to represent the random parameter $\tilde{\mathbf{x}}$, which has a probability distribution $\mathbb{P}$, namely $\tilde{\mathbf{x}} \sim \mathbb{P}$. $\mathbb{E}_{\mathbb{P}}[\mathbf{f}(\tilde{\mathbf{x}})]$ represents the expectation of the random variable $\mathbf{f}(\tilde{\mathbf{x}})$ under the probability distribution $\mathbb{P}$, where $\mathbf{f}(\tilde{\mathbf{x}})$ is the function: $\mathbb{R}^n \mapsto \mathbb{R}^m$. Furthermore, distributional ambiguity is modeled by an ambiguity set, which is denoted as $\mathcal{P}$. In addition, we denote the Dirac distribution as $\delta_{\boldsymbol{\theta}}$, which concentrates the unit mass at $\boldsymbol{\theta} \in \mathbb{R}^m$. To formulate the model rigorously, five sets of decision variables are defined as follows:

$r_{k,i,t} = 1$, if MV k has not departed from stop i at time interval t; $= 0$, otherwise.

$x_{k,q} = 1$, if the number of MUs in the MV assigned to trip k is q; $= 0$, otherwise.

$x_{k,i,q} = 1$, if the number of MUs in the MV assigned to trip k at stop i is q; $= 0$, otherwise.

α_k : Shifting time of the MV assigned to trip k at the first stop.

$t^h_{k,i}$: Holding time of the MV assigned to trip k at stop i.

Besides, Table 2 summarizes the symbols, notations, and the other variables used in this study.

Table 2
List of notations.

Notation	Detailed definition
Set	
$\mathcal{I}$	Set of stops, $\mathcal{I} = \{1, 2, \ldots, \lvert\mathcal{I}\rvert\}$
$\mathcal{K}$	Set of trips, $\mathcal{K} = \{1, 2, \ldots, \lvert\mathcal{K}\rvert\}$
$\mathcal{T}$	Set of time intervals, $\mathcal{T} = \{1, 2, \ldots, \lvert\mathcal{T}\rvert\}$
$\mathcal{T}_t$	Set of time intervals before and at time interval t, $\mathcal{T}_t = \{1, 2, \ldots, t\}, \mathcal{T}_t \subset \mathcal{T}$, let $\mathcal{T}_0 = \varnothing$
$\mathcal{Q}$	Set of the number of MUs that can be contained in an MV, $\mathcal{Q} = \{1, 2, \ldots, \lvert\mathcal{Q}\rvert\}$
Indices	
i, i', i''	Indices of stops, $i, i', i'' \in \mathcal{I}$
k	Index of trips, $k \in \mathcal{K}$
t	Index of time intervals, $t \in \mathcal{T}$
q	Index of the number of MUs in an MV, $q \in \mathcal{Q}$
Parameters	
Δ	Duration of each time interval
$\overline{\alpha}_k$	The maximum forward shifting time of the MV assigned to trip k at the first stop
$\underline{\alpha}_k$	The maximum backward shifting time of the MV assigned to trip k at the first stop
h_{min}	The minimum headway between any two adjacent trips
$\bar{t}^a_{k,1}$	Original scheduled arrival time of the MV assigned to trip k at the first stop, i.e., stop 1
$t^r_{k,i}$	Travel time of the MV assigned to trip k from stop i to stop $i+1$
$t^p_{k,i}$	Dwell time of the MV assigned to trip k at stop i
$\overline{\beta}_{k,i}$	Maximum holding time of the MV assigned to trip k at stop i
$\widetilde{\mu}_{i,t}$	Uncertain passenger arrival rate at stop i and time interval t
$\overline{\mu}_{i,t}$	Nominal passenger demand arriving at stop i on time interval t
$\check{\mu}_{i,t}$	Deviation from the nominal passenger demand arriving at stop i on time interval t
$\widetilde{\theta}_{i,t}$	Stochastic perturbation parameter related to passenger demand arriving at stop i on time interval t
$\delta_{i,i',t}$	Time-dependent ratio of passengers arriving at stop i with destination i' at time interval t
$\tau_{i,t}$	Time-dependent travel times from stops i to $i+1$ when MVs depart stop i at time interval t
ϑ_q	Operating costs of dispatching an MV equipped with q MUs to undertake a whole trip
$\hat{\vartheta}_q$	Operating costs of utilizing an MV equipped with q MUs at each stop
C_q	Capacity of the MV equipped with q MUs
ϑ^T	The equivalent monetary value of unit waiting time of passengers (unit:$)
ϑ^F	Fixed energy cost of dispatching MVs (unit:$)
ϑ^V	Positive coefficient related to the operating costs of MVs (unit:$)
w_1, w_2	Weights in the objective function
Variables	
$p^w_{k,i,t}$	Number of passengers arriving at stop i on time interval t and waiting for the MV assigned to trip k
$p^{wc}_{k,i,t}$	Total number of passengers waiting for the MV assigned to trip k at stop i and time interval t
$p^b_{k,i}$	Number of passengers boarding the MV assigned to trip k at stop i
$p^a_{k,i}$	Number of passengers alighting the MV assigned to trip k at stop i
$p^{on}_{k,i}$	Number of passengers on the MV assigned to trip k when it leaves stop i
$t^a_{k,i}$	Arrival time of the MV assigned to trip k at stop i
$t^d_{k,i}$	Departure time of the MV assigned to trip k from stop i
$e_{k,i,t}$	Binary indicator, $e_{k,i,t} = 1$ if time interval t belongs to the headway between MVs assigned to trips $k-1$ and k at stop i; $e_{k,i,t} = 0$, otherwise
$l_{k,i,t}$	Binary indicator, $l_{k,i,t} = 1$ if the MV assigned to trip k leaves stop i exactly at time interval t; $l_{k,i,t} = 0$, otherwise
Z_{wait}	Total waiting costs of passengers
Z_{MV}	Total operating costs of operators
Z	Weighted sum of total waiting and operating costs

3.2. Stochastic programming models

We assume that the uncertain demand arriving at each stop i on each time interval t depends on nominal passenger flows and a random perturbation $\widetilde{\mu}_{i,t} = \overline{\mu}_{i,t} + \widetilde{\theta}_{i,t} \cdot \check{\mu}_{i,t}, \forall i \in \mathcal{I}, t \in \mathcal{T}$. To be specific, $\overline{\mu}_{i,t}$ represents the nominal passenger demand arriving at stop i on time interval t, whose value is the minimum demand at the corresponding stop and time interval among all historical data. $\check{\mu}_{i,t}$ represents the deviation from the nominal demand. $\widetilde{\theta}_{i,t}$ represents the stochastic perturbation parameter. Based on the aforementioned foundations, we first formulate an SP model considering the time-varying capacity, for the case in which the information of the probability distribution is completely known. Most constraints of this model are the basis of the subsequent DRO formulations.

3.2.1. Stochastic programming formulation considering the time-varying capacity

(1) ***Three classes of binary variables associated with MV operations.*** Motivated by Niu and Zhou (2013), three binary variables ($(r_{k,i,t})|_{k\in\mathcal{K},i\in\mathcal{I},t\in\mathcal{T}}$, $(e_{k,i,t})|_{k\in\mathcal{K},i\in\mathcal{I},t\in\mathcal{T}}$ and $(l_{k,i,t})|_{k\in\mathcal{K},i\in\mathcal{I},t\in\mathcal{T}}$) are introduced in this study to facilitate the linear modeling of dynamic

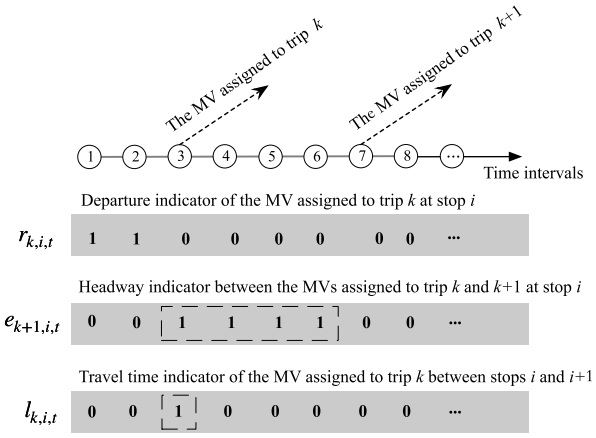


Fig. 4. Illustration of binary stage variables.

connections between the time-dependent travel times and departure times of MVs assigned to trips, as well as the couplings between these MVs and passengers. More specifically, $r_{k,i,t}$ represents a non-increasing departure indicator, whose time intervals related to the "1" indicate that the MV assigned to trip k has not departed from stop i. $e_{k,i,t}$ represents the headway indicator, and the time intervals associated with "1" represent the headway between the MVs assigned to trips $k-1$ and k at stop i, $\forall k \in \mathcal{K}\backslash\{1\}, i \in \mathcal{I}$. For the first trip, i.e., $k = 1$, the values of $e_{k,i,t}$ equal to those of $r_{k,i,t}$ at each stop $i \in \mathcal{I}$ on each time interval $t \in \mathcal{T}$. Travel times between stop i and stop $i+1$ are time-dependent and related to the departure time of the MV assigned to trip k at stop i. $l_{k,i,t}$ is the indicator of the departure time, and the value equals 1 if the MV assigned to trip k leaves stop i at time interval t. For clarity, an intuitive example is shown in Fig. 4. It can be seen that, the MV assigned to trip k departs from stop i at time interval 3, which is highlighted by a rectangle in $l_{k,i,t}$. Moreover, the headway between the MVs assigned to trip k and $k+1$ is at time intervals 3, 4, 5, and 6, as marked by the rectangle in $e_{k,i,t}$. Next, the properties of these three variables are further discussed, along with the applications in modeling MV operations and the evolution of passenger dynamics.

Property 1. *Binary variables are subject to the following constraints:*

$$r_{k,i,t+1} \le r_{k,i,t}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}\backslash\{|\mathcal{T}|\}. \tag{1}$$

$$t^d_{k,i} = \Delta(1 + \sum_{t\in\mathcal{T}} r_{k,i,t}), \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{2}$$

$$e_{k,i,t} = \begin{cases} r_{k,i,t}, k = 1 \\ r_{k,i,t} - r_{k-1,i,t}, k \in \mathcal{K}\backslash\{1\}, \end{cases} \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}. \tag{3}$$

$$l_{k,i,t} = \begin{cases} 1 - r_{k,i,t}, t = 1 \\ r_{k,i,t-1} - r_{k,i,t}, t \in \mathcal{T}\backslash\{1\}, \end{cases} \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{4}$$

$$r_{k,i,t}, e_{k,i,t}, l_{k,i,t} \in \{0,1\}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}. \tag{5}$$

Constraints (1) guarantee the non-increasing property of the binary variable $r_{k,i,t}$. The connection between this binary variable and real-valued departure time is represented by constraints (2). Constraints (3) describes the relationship between binary variables $r_{k,i,t}$ and $e_{k,i,t}$. In addition, we consider the dynamic travel times between two adjacent stops, whose values depend on the departure times of the MVs assigned to trips at each stop. Thus, the exact time interval of the departure time is crucial, and it cannot be characterized by $r_{k,i,t}$. We only need to evaluate the "change point" of $r_{k,i,t}$ to obtain the exact departure time because of the non-increasing property of $r_{k,i,t}$, i.e., the change point is the one where the difference between two adjacent $r_{k,i,t}$ is equal to 1, formulated as constraints (4). These binary variables provide considerable convenience for constructing linear models, the domains of which are defined in constraints (5).

Example 1. As illustrated in Fig. 5, the MVs assigned to trips 1 and 2 depart from stop 1 at time intervals 2 and 4, so the values of $(r_{k,i,t})|_{k=1,i=1,t\in\mathcal{T}}$ and $(r_{k,i,t})|_{k=2,i=1,t\in\mathcal{T}}$ are (1, 0, 0, …, 0) and (1, 1, 1, 0, 0, …, 0), respectively. On this basis, the values of $(e_{k,i,t})|_{k=1,i=1,t\in\mathcal{T}}$ and $(e_{k,i,t})|_{k=2,i=1,t\in\mathcal{T}}$ are (1, 0, 0, …, 0) and (0, 1, 1, …, 0) through constraints (3), which indicate the original headway of the first vehicle (i.e., the MV assigned to trip 1) is time interval 1, and the headway between the MVs assigned to trips 1 and 2 is time intervals 2 and 3. Moreover, the values of $(l_{k,i,t})|_{k=1,i=1,t\in\mathcal{T}}$ and $(l_{k,i,t})|_{k=2,i=1,t\in\mathcal{T}}$ can be calculated according to constraints (4), as highlighted by rectangles in $l_{k,i,t}$, which will later be used to identify the corresponding time-dependent travel times.

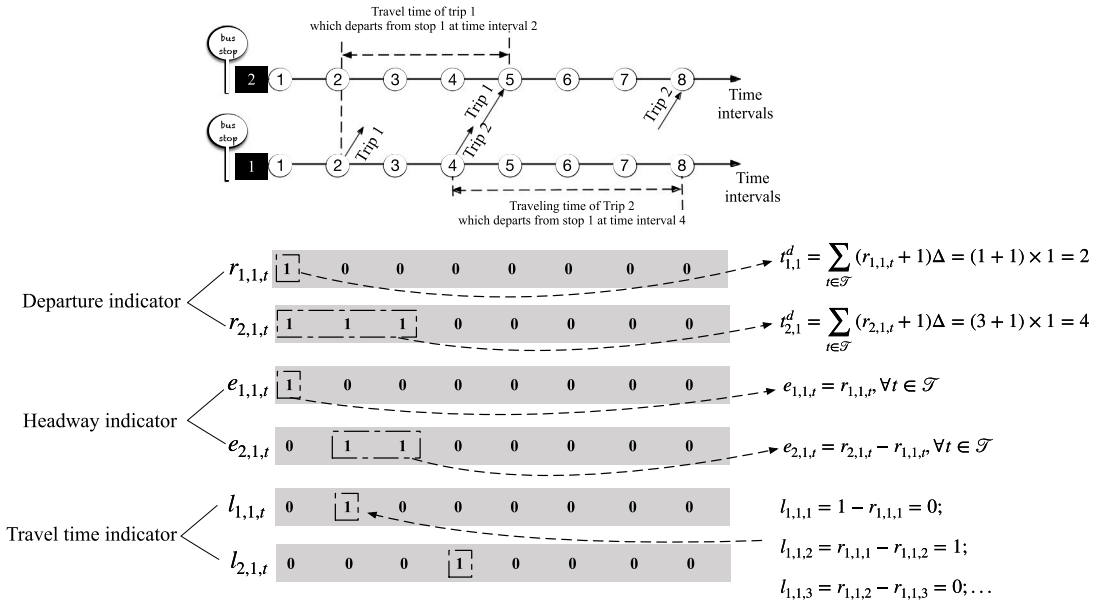


Fig. 5. Illustration of the relations among binary variables.

Property 2. *Coupling between the dynamic travel and departure times of MVs assigned to trips can be expressed as follows:*

$$t^r_{k,i} = \sum_{t\in\mathcal{T}} \tau_{i,t} \cdot l_{k,i,t}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}\backslash\{|\mathcal{I}|\}. \tag{6}$$

Example 2. For the sake of clarity, we still use the above example for illustrative purposes. Suppose that when MVs assigned to trips leave stop 1 at different time intervals, the corresponding travel times between stops 1 and 2 are 2, 3, 2, 4, 2, 3, 4, 2, 2, respectively. As illustrated in Fig. 6, when MVs assigned to trips 1 and 2 depart from stop 1 at time intervals 2 and 4, the travel times can be calculated according to constraints (6). For example, the travel time from stops 1 to 2 of MV assigned to trip 1, namely $t^r_{1,1}$, equals to $\sum_{t\in\mathcal{T}} \tau_{1,t} \cdot l_{1,1,t} = \tau_{1,2} \cdot l_{1,1,2} + \sum_{t\in\mathcal{T}\backslash\{2\}} \tau_{1,t} \cdot l_{1,1,t} = 3 \cdot 1 + 0 = 3$.

Property 3. *Coupling between MVs assigned to trips and passenger dynamics can be formulated as follows:*

$$p^w_{k,i,t} = \widetilde{\mu}_{i,t} \cdot e_{k,i,t}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}. \tag{7}$$

$$p^{wc}_{k,i,t} = e_{k,i,t} \cdot \sum_{\varsigma\in\mathcal{T}_t} p^w_{k,i,\varsigma} = e_{k,i,t} \sum_{\varsigma\in\mathcal{T}_t} \widetilde{\mu}_{i,\varsigma} \cdot e_{k,i,\varsigma}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}. \tag{8}$$

Stochastic passenger demand is a two-dimensional input parameter related to time intervals and stops, i.e., $\widetilde{\mu}_{i,t} \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{T}|}$. Key constraints (7) and (8) are proposed to facilitate the tracking of the cooperative dynamic evolution of MVs assigned to trips and passenger demand. The number of passengers who newly arrive at stop i at time interval t and wait for the MV assigned to trip k, namely $\mathbf{p^w} = \{p^w_{k,i,t}, \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}\} \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{I}|\times|\mathcal{T}|}$, can be obtained by constraints (7). In addition, constraints (8) describe the total number of passengers waiting for the MV assigned to trip k at stop i on time interval t, i.e., $\mathbf{p^{wc}} = \{p^{wc}_{k,i,t}, \forall k \in \mathcal{K}, i \in \mathcal{I}, t \in \mathcal{T}\} \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{I}|\times|\mathcal{T}|}$.

Example 3. As illustrated in Fig. 7, the realization of passenger demand is 1, 2, 3 at the first three time intervals, respectively. Following the previous examples, we still assume that the MV assigned to trip 1 departs from stop 1 at time interval 2, while the MV assigned to trip 2 leaves at time interval 4. Consequently, passengers arriving at time interval 1 can board the MV assigned to trip 1, whereas passengers arriving at time intervals 2 and 3 have to board the later one. In other words, the passengers waiting for an MV assigned to a specific trip arrive during the intervals between its departure and the departure of the MV assigned to the previous trip. Additionally, the number of passengers waiting for this MV will turn to zero once it departs.

For instance, as shown in Fig. 7, the passengers waiting for the MV assigned to trip 1 are those who arrive at the first time interval, because this MV will depart at the second time interval. After the departure of this MV, passengers will no longer wait for this trip; instead, they will await the MVs assigned to the other trips. Based on constraints (7) and (8), the number of passengers waiting for the MV assigned to trip 1 at the first time interval is calculated by $p^{wc}_{1,1,1} = e_{1,1,1} \cdot e_{1,1,1} \cdot \widetilde{\mu}_{1,1} = 1 \cdot 1 \cdot 1 = 1$. However, for subsequent time

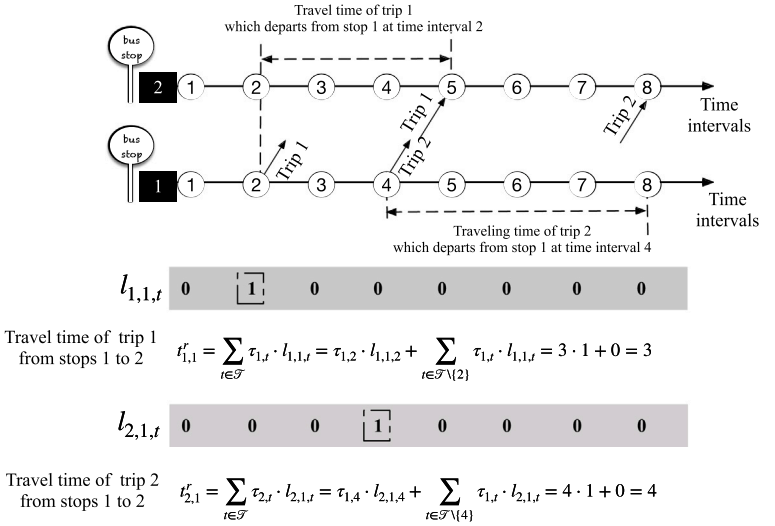


Fig. 6. Illustration of the coupling between the time-dependent travel times and the departure times of trips.

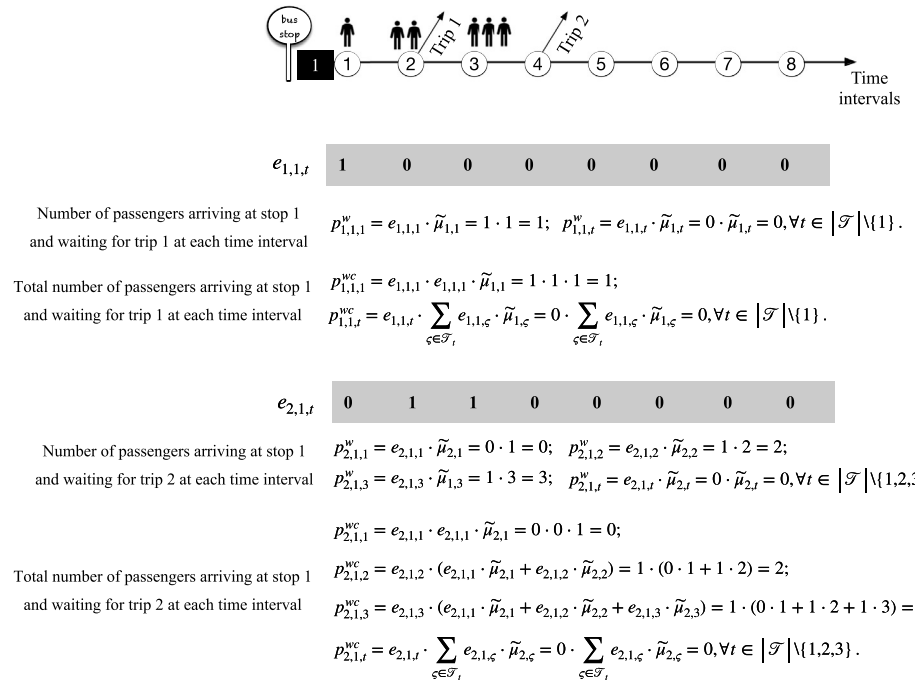


Fig. 7. Illustration of the coupling between trips and passengers.

intervals, this value is 0 due to $e_{1,1,t} = 0$ for each t in $\mathcal{T}\setminus\{1\}$, e.g., $p^{wc}_{1,1,2} = e_{1,1,2}\cdot(e_{1,1,1}\cdot\widetilde{\mu}_{1,1} + e_{1,1,2}\cdot\widetilde{\mu}_{1,2}) = 0\cdot(1\cdot 1 + 0\cdot 2) = 0$. Besides, the MV assigned to trip 2 departs at time interval 4. Consequently, the passengers waiting for this MV are those arriving at the second and third time intervals. Therefore, at the first time interval, the number of passengers waiting for the MV assigned to trip 2 is 0 (i.e., $p^{wc}_{2,1,1} = e_{2,1,1}\cdot e_{2,1,1}\cdot\widetilde{\mu}_{2,1} = 0\cdot 0\cdot 1 = 0$). At the time interval 2, this number is calculated by $p^{wc}_{2,1,2} = e_{2,1,2}\cdot(e_{2,1,1}\cdot\widetilde{\mu}_{2,1} + e_{2,1,2}\cdot\widetilde{\mu}_{2,2}) = 1\cdot(0\cdot 1 + 1\cdot 2) = 2$. Then, this number at time interval 3 turns to $p^{wc}_{2,1,3} = e_{2,1,3}\cdot(e_{2,1,1}\cdot\widetilde{\mu}_{2,1} + e_{2,1,2}\cdot\widetilde{\mu}_{2,2} + e_{2,1,3}\cdot\widetilde{\mu}_{2,3}) = 1\cdot(0\cdot 1 + 1\cdot 2 + 1\cdot 3) = 5$. For subsequent time intervals, the number of passengers waiting for the MV assigned to trip 2 is 0 since $p^{wc}_{2,1,t} = 0, \forall t \in \{4, 5, 6, 7, \ldots, |\mathcal{T}|\}$.

(2) ***Constraints associated with MV operations.*** The selection of vehicle formations (i.e., the number of MUs in each MV), arrival at stops, departure from the stops and running between stops are the basic processes of the dynamic evolution of the vehicle status; they can be formulated as follows:

$$t^a_{k,i} = \begin{cases} \bar{t}^a_{k,1} + \alpha_k, & i = 1 \\ t^d_{k,i-1} + t^r_{k,i-1}, & \forall i \in \mathcal{I}\backslash\{1\}, \end{cases} \quad \forall k \in \mathcal{K}. \tag{9}$$

$$t^d_{k,i} = t^a_{k,i} + t^p_{k,i} + t^h_{k,i}, \quad \forall i \in \mathcal{I}, k \in \mathcal{K}. \tag{10}$$

$$t^d_{k+1,i} - t^d_{k,i} \geq h_{min}, \quad \forall k \in \mathcal{K}\backslash\{|\mathcal{K}|\}, i \in \mathcal{I}. \tag{11}$$

$$\underline{\alpha}_k \leq \alpha_k \leq \overline{\alpha}_k, \quad \forall k \in \mathcal{K}. \tag{12}$$

$$0 \leq t^h_{k,i} \leq \overline{\beta}_{k,i}, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{13}$$

Constraints (9) and (10) are used to track the real-valued arrival and departure times of the MV assigned to each trip, respectively. Further, the arrival time at the first stop of the MV assigned to each trip is the pre-given original scheduled arrival time plus the shifting time; the arrival time at the subsequent stops is the sum of the departure time at the previous stop and the corresponding travel time between stops. The time-dependent travel time between two adjacent stops i and $i+1$ of the MV assigned to each trip depends on its departure time at stop i, as discussed and formulated in Properties 1 and 2. The departure time of the MV assigned to each trip at each stop is related to the arrival, dwell, and holding times at this stop. The holding strategy is imposed to adjust the dwell time in a flexible manner. Further, constraints (11) are formulated to ensure that the minimum headway of two continuous MVs assigned to trips at the same stop can guarantee the operational safety and avoid the "bus bunching" phenomenon. Constraints (12) and (13) specify the limitations of the shifting and holding times.

(3) ***Constraints related to passenger movements.*** The set of constraints related to passenger movements, i.e., waiting, boarding and alighting behaviors, are formulated as constraints (7)–(8) in Property 3, and the following constraints (14)–(15). Constraints (7) and (8) describe the waiting passengers at each time interval, and the cumulative waiting passengers. Constraints (14) indicate that, the number of alighting passengers at stop i is the number of boarding passengers multiplied by the time-dependent ratio of passengers arriving at stop i' to destination i at time interval t. Constraints (15) define the number of passengers boarding the MV assigned to trip k at stop i.

$$p^a_{k,i} = \sum_{i' \leq i, i' \in \mathcal{I}} \sum_{t \in \mathcal{T}} \delta_{i',i,t} \cdot p^w_{k,i',t}, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{14}$$

$$p^b_{k,i} = \sum_{t \in \mathcal{T}} \widetilde{\mu}_{i,t} \cdot e_{k,i,t} \quad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{15}$$

(4) ***Constraints associated with vehicle capacity.*** The set of constraints associated with vehicle capacity and the number of in-vehicle passengers are formulated as constraints (16)–(19). Constraints (16) describe the dynamics of the occupied capacity of the MV assigned to trip k when it leaves stop i, i.e., the number of in-vehicle passengers. Constraints (17) are imposed to ensure that the number of in-vehicle passengers does not exceed the maximum capacity to guarantee operational safety. Constraints (18) ensure that the formation of the MV assigned to each trip is unique. Furthermore, constraints (19) set the binary domain of the variables related to vehicle formation.

$$p^{on}_{k,i} = \begin{cases} p^b_{k,i}, & i = 1 \\ p^{on}_{k,i-1} - p^a_{k,i} + p^b_{k,i}, & \forall i \in \mathcal{I}\backslash\{1\}, \end{cases} \quad \forall k \in \mathcal{K}. \tag{16}$$

$$p^{on}_{k,i} \leq \sum_{q \in \mathcal{Q}} C_q \cdot x_{k,q}, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{17}$$

$$\sum_{q \in \mathcal{Q}} x_{k,q} = 1, \quad \forall k \in \mathcal{K}. \tag{18}$$

$$x_{k,q} \in \{0,1\}, \quad \forall k \in \mathcal{K}, q \in \mathcal{Q}. \tag{19}$$

(5) ***Objective function.*** The study aims to achieve a balance between the costs of passengers and operators from a system perspective by adjusting the timetables to match the time-dependent passenger demand and by dispatching the appropriate numbers of MUs in each MV to reduce operating costs. Therefore, the objective function can be formulated as follows:

$$Z = w_1 \cdot Z_{wait} + w_2 \cdot Z_{MV}, \tag{20}$$

where w_1 and w_2 represent the weight coefficients, Z_{wait} represents the total waiting costs of passengers, and Z_{MV} represents the total operating costs of the operators.

• **Total waiting costs of passengers**

The number of passengers waiting for the MV assigned to trip $k \in \mathcal{K}$ at stop $i \in \mathcal{I}$ in time interval $t \in \mathcal{T}$ is derived from constraints (8), i.e., $p^{wc}_{k,i,t}$. Thus, the corresponding waiting time of passengers can be obtained by multiplying the number of waiting passengers by the unit duration of the time interval (i.e., $\Delta \cdot p^{wc}_{k,i,t}$). Subsequently, by multiplying the equivalent monetary value of unit waiting time (denoted as ϑ^T), the total waiting costs of passengers can be expressed as follows:

$$Z_{wait} = \Delta \cdot \vartheta^T \cdot \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} p^{wc}_{k,i,t}. \tag{21}$$

Remark 1. In this study, we evaluate passengers' costs by considering the total waiting costs, which can be calculated based on the waiting time of passengers. This metric is widely utilized in the single-line timetabling field for assessing service quality. However, it is worth noting that transfer times are not taken into account in this study, since we specifically focus on the single-line problem.

Lemma 1. *Constraint (21) exhibits nonlinearity because of the presence of the nonlinear term $p^{wc}_{k,i,t}$ derived from constraints (8), which can be equivalently reformulated as follows:*

$$Z_{wait} = \Delta \cdot \vartheta^T \cdot \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} \sum_{\varsigma\in\mathcal{T}_t} \widetilde{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma}, \tag{22}$$

where $\mathbf{v} \in \mathbb{R}^{|\mathcal{K}|\times|\mathcal{I}|\times|\mathcal{T}|\times|\mathcal{T}|}$ is an auxiliary binary vector. The motivations behind the linearization are twofold. Firstly, by linearizing this nonlinear constraint, the proposed model can be transformed into an MILP formulation, which has the potential for enhanced computational efficiency and enables more efficient solution methods in theory. Secondly, the dual information of this constraint is essential for formulating the subsequent DRO model.

Proof. First, by integrating constraints (8) and (21), we have

$$Z_{wait} = \Delta \cdot \vartheta^T \cdot \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} e_{k,i,t} \sum_{\varsigma\in\mathcal{T}_t} \widetilde{\mu}_{i,\varsigma} \cdot e_{k,i,\varsigma}. \tag{23}$$

We note that constraint (23) is nonlinear, leading to mathematical models that are difficult to be solved by commercial solvers or exact algorithms. Thus, an auxiliary variable, i.e., $v_{k,i,t_1,t_2} = e_{k,i,t_1} \cdot e_{k,i,t_2}$, is then introduced here to linearize constraint (23):

$$\begin{cases} v_{k,i,t_1,t_2} \le e_{k,i,t1} \\ v_{k,i,t_1,t_2} \le e_{k,i,t2} \\ v_{k,i,t_1,t_2} \ge e_{k,i,t1} + e_{k,i,t2} - 1 \\ v_{k,i,t_1,t_2} \in \{0,1\} \end{cases}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, t_1 \in \mathcal{T}, t_2 \in \mathcal{T}_{t_1}, \tag{24}$$

then, by replacing $e_{k,i,t}$ with $v_{k,i,t,\varsigma}$ for all $\varsigma \in \mathcal{T}_t$, the nonlinear constraint (21) can be equivalently transformed into the following linear form

$$Z_{wait} = \Delta \cdot \vartheta^T \cdot \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} \sum_{\varsigma\in\mathcal{T}_t} \widetilde{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma},$$

which completes the proof. □

• **Total operating costs of operators**

Referring to Chen et al. (2019), the total operating costs can be formulated as follows

$$Z_{MV} = \sum_{q\in\mathcal{Q}} \sum_{k\in\mathcal{K}} \vartheta_q \cdot x_{k,q}, \tag{25}$$

where $\vartheta_q = \vartheta^F + \vartheta^V \cdot (C_q)^\epsilon$, ϑ^F represents the fixed operating costs, and ϑ^V is the positive coefficient related to the operating costs, and power index $\epsilon \le 1$. The first component of ϑ_q (i.e., ϑ^F) represents the fixed operating costs incurred once an MV is dispatched. In addition, given that $\epsilon \le 1$, the second component of ϑ_q have a marginally reducing effect even though they increase monotonically with the number of MUs in the MV.

(6) ***MILP formulation for the stochastic TT-DCA problem with time-varying capacity.*** Based on the above constraints and assuming that the distribution of the stochastic perturbation parameter $\boldsymbol{\theta} := [\theta_{1,1}, \ldots, \theta_{|\mathcal{I}|,|\mathcal{T}|}]^{\mathrm{T}}$ (denoted by $\mathbb{P}_\theta$) is completely known, we can construct the following SP formulation:

$$\begin{cases} \min\limits_{\mathbf{y}\in\mathcal{Y}} & \mathbb{E}_{\mathbb{P}}[Z(\mathbf{y},\boldsymbol{\theta})] \\ \text{s.t.} & \text{Constraints (1)–(7), (9)–(20), (22), (24)–(25),} \end{cases} \tag{26}$$

where the decision vector $\mathbf{y} := (\alpha, \mathbf{t}^h, \mathbf{r}, \mathbf{x})$, and $\mathcal{Y}$ represents the feasible region of the decision variables. Sample average approximation (SAA) is a natural method for handling SP models. Past realizations of random variables at each sample n can be obtained (i.e., $\hat{\boldsymbol{\theta}}_n|_{n\le N}$) by employing a finite set of N samples, and the empirical probability distribution of each sample is $1/N$.

The MILP formulation for the *TT-DCA problem of MVs under stochastic demand with time-varying capacity* (SP-TVC) can be expressed as follows:

$$[\text{SP}-\text{TVC}] \quad \begin{cases} \min\limits_{\mathbf{y}\in\mathcal{Y}} & \frac{1}{N}\sum\limits_{n=1}^{N} Z(\mathbf{y},\hat{\theta}_n) \\ \text{s.t.} & \text{Constraints (1)–(7), (9)–(20), (22), (24)–(25).} \end{cases} \tag{27}$$

3.2.2. Extensions of the proposed model enabling the time-varying and station-wise capacity

In reality, passenger demand is not only time-dependent but also spatially uneven. If we dynamically adjust the formations of MVs assigned to trips based on the demand at each stop in each time interval, the operating costs can be reduced further. Besides, considering that the technology of docking and undocking at each stop has already been realized in the MVs developed by NEXT Future Transportation Inc., we extend the aforementioned SP-TVC model to a more general formulation, which additionally allows the MUs to be decoupled or/and coupled at each stop. In the SP-TVC model without considering the station-wise capacity, the vehicle-related decision variable is $x_{k,q}$ for any $k \in \mathcal{K}, q \in \mathcal{Q}$; this represents that the number of MUs in MV k equal to q. However, once we consider that the formation of the MV assigned to trip k can be changed at each stop $i \in \mathcal{I}$ flexibly, the above-mentioned vehicle-related decision variable should be revised to a new notation $x_{k,i,q}$, which is called the vehicle-station-related formation variable. The new decision variable $x_{k,i,q}$ indicates that the number of MUs in the MV assigned to trip k at stop i is q. Then, the operating costs defined in constraint (25) should be reformulated as

$$Z_{MV} = \sum_{i\in\mathcal{I}}\sum_{q\in\mathcal{Q}}\sum_{k\in\mathcal{K}} \hat{\vartheta}_q \cdot x_{k,i,q}, \tag{28}$$

where $\hat{\vartheta}_q$ refers to the station-wise operating cost, i.e., the operating costs of utilizing an MV equipped with q MUs at each stop, and $\hat{\vartheta}_q = \frac{\vartheta_q}{|\mathcal{I}|}$.

In this new formulation, we also need to revise the limitations on the number of in-vehicle passengers and the formations for ensuring that the number of in-vehicle passengers does not exceed the capacity and that the formation of the MV assigned to each trip is unique at every stop. Subsequently, the related constraints (17)–(19) should be replaced by the following constraints:

$$p^{on}_{k,i} \le \sum_{q\in\mathcal{Q}} C_q \cdot x_{k,i,q}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{29}$$

$$\sum_{q\in\mathcal{Q}} x_{k,i,q} = 1, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{30}$$

$$x_{k,i,q} \in \{0,1\}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, q \in \mathcal{Q}. \tag{31}$$

The SP model for the TT-DCA problem (SP-TVSWC) when considering the *time-varying and station-wise capacity* based on the SP-TVC model can be expressed as

$$[\text{SP}-\text{TVSWC}] \quad \begin{cases} \min\limits_{\mathbf{y}\in\mathcal{Y}} & \frac{1}{N}\sum\limits_{n=1}^{N} Z(\mathbf{y},\hat{\theta}_n) \\ \text{s.t.} & \text{Constraints (1)–(7), (9)–(16), (20), (22), (24), (28)–(31).} \end{cases} \tag{32}$$

Remark 2. The proposed model can be easily reformulated in various ways to suit the real-world operational needs of public transport systems. For example, we can employ $t^r_{k,i} = t^r_{k+1,i}, \forall k \in \mathcal{K}\backslash\{|\mathcal{K}|\}, i \in \mathcal{I}\backslash\{|\mathcal{I}|\}$ to simplify the dynamics of the running time on each section.

3.3. Distributionally robust formulations

The proposed SP models for the TT-DCA problem assume exact prior knowledge of the probability distribution of uncertain passenger demand. However, this assumption does not hold in practical problems because passenger demand may be uncertain because of many unpredictable factors such as weather conditions. We further propose data-driven DRO formulations that aim to optimize the expected value of a function for the worst probability distribution in an ambiguity set. We assume that the distribution of the random perturbation variables belongs to an ambiguity set based on the Wasserstein distance, and this is one of the most widely used distance metrics for constructing statistical distance-based ambiguity sets. We employ this ambiguity set because of two reasons: (1) It facilitates the translation of a computationally intractable DRO model into a computable reformulation in various real-world applications. (2) This set is straightforward for decision making, and it enables the decision maker to control the robustness of the optimal solutions by simply adjusting a single non-negative parameter.

In the following analyses, we first design a customized Wasserstein distance-based ambiguity set with time-dependent and station-wise perturbation variables to characterize the dynamics of the passenger demand at each stop at each time interval and its probability distributions. Subsequently, new linear forms of the objective functions in SP-TVC (27) and SP-TVSWC (32) are obtained, and we deal with the constraints with a right-hand uncertainty.

(1) ***Wasserstein distance-based ambiguity set with time-dependent and station-wise perturbation parameters.*** The Wasserstein distance between the two probability distributions represents the optimal transport cost of moving from one to the other. The definition of the Wasserstein metric d_{W}, first proposed by Kantorovich and Rubinshtein (1958), can be expressed as follows

$$d_{\mathrm{W}}\left(\mathbb{P}_1,\mathbb{P}_2\right) := \inf\left\{\int_{\Xi^2}\left\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\right\| \Pi\left(\mathrm{d}\boldsymbol{\theta}_1,\ \mathrm{d}\boldsymbol{\theta}_2\right) : \begin{array}{l}\Pi \text{ is a joint distribution of } \boldsymbol{\theta}_1 \text{ and } \boldsymbol{\theta}_2\\ \text{with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2, \text{ respectively}\end{array}\right\},$$

in which $\|\cdot\|$ denotes an 1-norm on $\mathbb{R}^m$ in this study.

Accordingly, we consider the following Wasserstein ambiguity set to exploit the information embedded in historical data

$$\mathcal{D}_\rho(\widehat{\mathbb{P}}_N) := \left\{\mathbb{P}\in\mathcal{M}(\Xi)|d_{\mathrm{W}}(\mathbb{P},\widehat{\mathbb{P}}_N)\le\rho\right\},$$

where Ξ represents the support set, and space $\mathcal{M}(\Xi)$ includes all possible probability distributions supported on Ξ. $\mathcal{D}_\rho(\widehat{\mathbb{P}}_N)$ can be considered a Wasserstein ball centered at the empirical distribution that contains all probability distributions $\mathbb{P}$ within radius ρ. Radius ρ reflects the difference between the empirical distribution of the uncertainty and the true distribution.

Remark 3. When ρ is set to a sufficiently large number, the ambiguity set can be considered to contain all probability distributions on support Ξ, and then, DRO coincides with RO. In contrast, the ambiguity set contains only one distribution when ρ equals zero, and DRO coincides with SP.

Moreover, empirical distribution $\hat{\mathbb{P}}_N$ can be constructed using the historical data set $\{\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2,\dots,\hat{\boldsymbol{\theta}}_N\}$, i.e.,

$$\widehat{\mathbb{P}}_N := \frac{1}{N}\sum_{n=1}^{N}\delta_{\hat{\boldsymbol{\theta}}_n}.$$

The perturbation parameter $\boldsymbol{\theta}$ in each sample is both time-dependent and station-wise to portray the dynamic and spatially unbalanced nature of this parameter.

(2) ***Reformulation of objective functions.*** We propose two DRO models with the objective of minimizing the *worst-case* expected value, which considers time-varying capacity and time-varying and station-wise capacity, respectively, to obtain a robust timetable and a robust dynamic-capacity allocation plan to mitigate the impact of uncertain passenger demand. The objective functions of these two models can be expressed in a unified manner as

$$\min_{\mathbf{y}\in\mathcal{Y}}\ \sup_{\mathbb{P}\in\mathcal{D}_\rho(\widehat{\mathbb{P}}_N)}\ \mathbb{E}_{\mathbb{P}}[Z(\mathbf{x},\boldsymbol{\theta})], \tag{33}$$

where the inner $\sup\{\cdot\}$ determines the worst-case evaluation of the weighted objective over all distributions contained in the Wasserstein ball $\mathcal{D}_\rho(\widehat{\mathbb{P}}_N)$. The DRO formulations aim to find the optimal timetables and dynamic-capacity allocation plans to ensure acceptable performance, even in the worst-case scenario. However, the inner $\sup\{\cdot\}$ problem is an infinite-dimensional optimization problem because of the existence of infinite probability distributions in the Wasserstein ball, and it is therefore computationally intractable. Next, the equivalent computationally tractable form of the problem (33) is obtained.

Proposition 1. *First, we construct the objective function of the DRO model by considering the time-varying capacity. In this study, we adopted the polyhedral support set proposed by Bertsimas and Sim (2004) to control the conservativeness of our model. In other words, suppose that the support set of random variable* $\boldsymbol{\theta}$ *is a bounded polyhedron:* $\Xi=\{\boldsymbol{\theta}\mid\mathbf{A}\boldsymbol{\theta}\le\mathbf{b}\}=\{\sum_{i=1}^{|\mathcal{I}|}\theta_{i,t}\le\Gamma_t,\sum_{t=1}^{|\mathcal{T}|}\theta_{i,t}\le\Lambda_i,0\le\theta_{i,t}\le1\}$. *We denote the vector* $\boldsymbol{\theta}=[\theta_{1,1},\theta_{1,2},\dots,\theta_{i,t},\dots,\theta_{|\mathcal{I}|,|\mathcal{T}|}]^{\mathrm{T}}$, *which captures the values of* θ *corresponding to different combinations of indices* i *and* t. *In the* n*th sample of the historical data, the vector* $\hat{\boldsymbol{\theta}}_n$ *is formed as* $[\hat{\theta}_{n,1,1},\hat{\theta}_{n,1,2},\dots,\hat{\theta}_{n,i,t},\dots\hat{\theta}_{n,|\mathcal{I}|,|\mathcal{T}|}]^{\mathrm{T}}$. *To simplify the subsequent derivation of the dual form, we redefine the indices of* $\theta_{i,t}$ *using a new index* j *that ranges from* 1 *to* $|\mathcal{I}|\times|\mathcal{T}|$. *Specifically, we define* $\mathcal{J}=\{1,2,\dots,j,\dots,|\mathcal{J}|\}$, *where* $j=(i-1)|\mathcal{T}|+t$, $|\mathcal{J}|=|\mathcal{I}|\times|\mathcal{T}|$. *By doing so,* $\boldsymbol{\theta}$ *and* $\hat{\boldsymbol{\theta}}_n$ *can be reshaped as* $\boldsymbol{\theta}=[\theta_1,\theta_2,\dots,\theta_j,\dots\theta_{|\mathcal{J}|}]^{\mathrm{T}}$ *and* $\hat{\boldsymbol{\theta}}_n=[\hat{\theta}_{n,1},\hat{\theta}_{n,2},\dots,\hat{\theta}_{n,j},\dots\hat{\theta}_{n,|\mathcal{J}|}]^{\mathrm{T}}$, *respectively. Based on this representation, we have* $\mathbf{A}\in\mathbb{R}^{D\times|\mathcal{J}|}$, *where* D *refers to the number of rows of* $\mathbf{A}$.

Decision makers can use the aforementioned support set to balance solution efficiency and robustness by adjusting the values of Γ_t *and* Λ_i. *The value of* Γ_t *determines the number of time intervals for a stop to reach the maximum possible passenger demand in the study horizon; the value of* Λ_i *determines the maximum number of stops that meet the maximum possible passenger demand at each time interval. Both values can be estimated by analyzing historical data.*

Under the Wasserstein distance-based ambiguity set $\mathcal{D}_\rho(\widehat{\mathbb{P}}_N)$ *and polyhedral support set* Ξ, *the equivalent linear form of the formulation (33) for the TT-DAC problem with time-varying capacity can be expressed as follows:*

$$\inf_{\mathbf{v},\mathbf{x},\sigma,\mathbf{c},\lambda}\left\{w_1\cdot\Delta\cdot\vartheta^T\cdot\Big(\sum_{k\in\mathcal{K}}\sum_{i\in\mathcal{I}}\sum_{t\in\mathcal{T}}\sum_{\varsigma\in\mathcal{T}_t}\overline{\mu}_{i,\varsigma}\cdot v_{k,i,t,\varsigma}+\rho\sigma+\frac{1}{N}\sum_{n=1}^{N}c_n\Big)+w_2\cdot\sum_{q\in Q}\sum_{k\in\mathcal{K}}\vartheta_q\cdot x_{k,q}\right\} \tag{34}$$

$$\text{s.t.} \quad \sum_{j=1}^{|\mathcal{J}|} \pi_j \cdot \hat{\theta}_{n,j} + \sum_{d=1}^{D} (b_d - \sum_{j=1}^{|\mathcal{J}|} A_{d,j} \cdot \hat{\theta}_{n,j}) \lambda_{n,d} \le c_n, \qquad \forall n \le N, \tag{35}$$

$$\sum_{d=1}^{D} A_{d,j} \cdot \lambda_{n,d} - \pi_j \le \sigma, \qquad \forall n \le N, j \in \mathcal{J}, \tag{36}$$

$$\pi_j - \sum_{d=1}^{D} A_{d,j} \cdot \lambda_{n,d} \le \sigma, \qquad \forall n \le N, j \in \mathcal{J}, \tag{37}$$

$$\pi_j = \breve{\mu}_{i,t} \sum_{k \in \mathcal{K}} \sum_{t_0 \in \mathcal{T} \setminus \mathcal{T}_{t-1}} v_{k,i,t_0,t}, \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}, j = (i-1)|\mathcal{T}| + t, \tag{38}$$

$$\lambda_{n,d} \ge 0, \qquad \forall n \le N, d \le D. \tag{39}$$

Proof. First, constraint (22) can be converted to

$$\begin{aligned} Z_{wait} &= \Delta \cdot \vartheta^T \cdot \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \widetilde{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} \\ &= \Delta \cdot \vartheta^T \cdot \left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} (\overline{\mu}_{i,\varsigma} + \theta_{i,\varsigma} \cdot \breve{\mu}_{i,\varsigma}) \cdot v_{k,i,t,\varsigma} \right] \\ &= \Delta \cdot \vartheta^T \cdot \left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \theta_{i,t} \cdot \breve{\mu}_{i,t} \sum_{k \in \mathcal{K}} \sum_{t_0 \in \mathcal{T} \setminus \mathcal{T}_{t-1}} v_{k,i,t_0,t} \right] \\ &= \Delta \cdot \vartheta^T \cdot \left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \theta_{i,t} \cdot \pi_{i,t} \right] \\ &= \Delta \cdot \vartheta^T \cdot \left[\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + \sum_{j \in \mathcal{J}} \theta_j \cdot \pi_j \right], \end{aligned} \tag{40}$$

where $\pi_j = \pi_{i,t} = \breve{\mu}_{i,t} \sum_{k \in \mathcal{K}} \sum_{t_0 \in \mathcal{T} \setminus \mathcal{T}_{t-1}} v_{k,i,t_0,t}, \forall j \in \mathcal{J}$.

Then, by substituting constraint (40) into formulation (33), we have

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{Y}} \sup_{\mathbb{P} \in \mathcal{D}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[Z(\mathbf{x}, \boldsymbol{\theta})] &= \min_{\mathbf{y} \in \mathbb{Y}} \sup_{\mathbb{P} \in \mathcal{D}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[w_1 \cdot Z_{wait} + w_2 \cdot Z_{MV}] \\ &= \min_{\mathbf{y} \in \mathbb{Y}} \Bigg\{ w_1 \cdot \Delta \cdot \vartheta^T \cdot \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + w_2 \cdot \sum_{q \in \mathcal{Q}} \sum_{k \in \mathcal{K}} \vartheta_q \cdot x_{k,q} \\ &\quad + w_1 \cdot \Delta \cdot \vartheta^T \cdot \sup_{\mathbb{P} \in \mathcal{D}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}} \left[\sum_{j \in \mathcal{J}} \theta_j \cdot \pi_j \right] \Bigg\}. \end{aligned} \tag{41}$$

All terms in constraint (41) are in the linear form except for the inner problem, i.e., $\sup_{\mathbb{P} \in \mathcal{D}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}} \left[\sum_{j \in \mathcal{J}} \theta_j \cdot \pi_j \right]$. Further, according to the derivations in Appendix A, this term can be equivalently transformed into

$$\inf_{\sigma, \mathbf{c}, \lambda} \quad \rho\sigma + \frac{1}{N} \sum_{n=1}^{N} c_n \tag{42}$$

$$\text{s.t.} \quad \sum_{j=1}^{|\mathcal{J}|} \hat{\theta}_{n,j} \cdot \pi_j + \sum_{d=1}^{D} (b_d - \sum_{j=1}^{|\mathcal{J}|} A_{d,j} \cdot \hat{\theta}_{n,j}) \lambda_{n,d} \le c_n, \qquad \forall n \le N, \tag{43}$$

$$\left\| \mathbf{A}^{\mathrm{T}} \lambda_n - \boldsymbol{\pi} \right\|_* \le \sigma, \qquad \forall n \le N, \tag{44}$$

$$\lambda_{n,d} \ge 0, \qquad \forall n \le N, d \le D. \tag{45}$$

Given that 1-norm is used as the metric in the definition of the Wasserstein distance, its dual norm $\|\cdot\|_*$ is an ∞-norm. Therefore, constraints (44) can be reduced to

$$\left| \sum_{d=1}^{D} A_{d,j} \cdot \lambda_{n,d} - \pi_j \right| \le \sigma, \qquad \forall n \le N, j \in \mathcal{J}. \tag{46}$$

We can obtain the formulations (34)–(39) by integrating constraints (40)–(46), and the proof of Proposition 1 is thus complete. □

Proposition 2. *We then formulate the objective function of the DRO model by considering time-varying and station-wise capacity. The use of constraint (28) instead of the constraint (25) distinguishes this extended formulation from the aforementioned DRO model in terms of the objective function. Then, using a similar derivation, we have*

$$\inf_{\mathbf{r}, \mathbf{x}, \sigma, \mathbf{c}, \lambda} \quad \Bigg\{ w_1 \cdot \Delta \cdot \vartheta^T \cdot \Big(\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} \sum_{\varsigma \in \mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + \rho\sigma +$$

$$\frac{1}{N}\sum_{n=1}^{N} c_n) + w_2 \cdot \sum_{i\in\mathcal{I}}\sum_{q\in Q}\sum_{k\in\mathcal{K}} \widehat{\vartheta}_q \cdot x_{k,i,q} \Bigg\}$$

s.t. Constraints (35)–(39). (47)

(3) ***Reformulation of constraints with right-hand uncertainties.*** The following robust constraints are further reformulated to replace the constraints (17) and (29) to ensure that all passengers can board successfully regardless of the realizations of uncertain demand.

$$\sum_{q\in Q} C_q \cdot x_{k,q} \geq \sup_{\boldsymbol{\theta}\in\Xi}\{p^{on}_{k,i}\}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{48}$$

$$\sum_{q\in Q} C_q \cdot x_{k,i,q} \geq \sup_{\boldsymbol{\theta}\in\Xi}\{p^{on}_{k,i}\}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{49}$$

Notice that the polyhedral support set Ξ contains infinite realization of $\boldsymbol{\theta}$, thus, constraints (48) and (49) are still computationally intractable. In the following, we further explore their linear equivalent forms.

Proposition 3. *We first reformulate the related constraints (48) of the DRO model considering the time-varying capacity. Under the polyhedral support set Ξ, the equivalent linear form of constraints (48) can be expressed as the following set of constraints:*

$$\sum_{q\in Q} C_q \cdot x_{k,q} \geq \sum_{t\in\mathcal{T}}\sum_{i''\leq i,i''\in\mathcal{I}} \gamma_{i,i'',t} \cdot \bar{\mu}_{i'',t} \cdot e_{k,i'',t} + \sum_{d=1}^{D} b_d \cdot g_{k,i,d}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{50}$$

$$\sum_{d=1}^{D} A_{d,(i''-1)|\mathcal{T}|+t} \cdot g_{k,i,d} = \gamma_{i,i'',t} \cdot \check{\mu}_{i'',t} \cdot e_{k,i'',t}, \qquad \forall k \in \mathcal{K}, i'' \leq i, i, i'' \in \mathcal{I}, t \in \mathcal{T}. \tag{51}$$

$$\sum_{d=1}^{D} A_{d,m} \cdot g_{k,i,d} = 0, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, m > i\,|\mathcal{T}|\,. \tag{52}$$

$$g_{k,i,d} \geq 0, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, d \in D. \tag{53}$$

$$\gamma_{i,i'',t} = 1 - \sum_{i'=i''}^{i} \delta_{i'',i',t}, \qquad \forall i'' \leq i, i, i'' \in \mathcal{I}, t \in \mathcal{T}. \tag{54}$$

Proof. First, by combining the constraints (7) and (15), we have

$$p^{b}_{k,i} = \sum_{t\in\mathcal{T}} \widetilde{\mu}_{i,t} \cdot e_{k,i,t} = \sum_{t\in\mathcal{T}} p^{w}_{k,i,t}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}.$$

Then, considering the constraints (14) and (16), the number of passengers on the MV assigned to trip $k \in \mathcal{K}$ when it leaves stop $i \in \mathcal{I}$ can be recursively calculated as follows:

$$p^{on}_{k,1} = p^{b}_{k,1} - p^{a}_{k,1} = \sum_{t\in\mathcal{T}} p^{w}_{k,1,t} - \sum_{i'\leq 1,i'\in\mathcal{I}}\sum_{t\in\mathcal{T}} \delta_{i',1,t} \cdot p^{w}_{k,i',t} = \sum_{t\in\mathcal{T}} p^{w}_{k,1,t} - \sum_{t\in\mathcal{T}} \delta_{1,1,t} \cdot p^{w}_{k,1,t} = \sum_{t\in\mathcal{T}} (1-\delta_{1,1,t}) p^{w}_{k,1,t},$$

$$p^{on}_{k,2} = p^{on}_{k,1} + p^{b}_{k,2} - p^{a}_{k,2} = \sum_{t\in\mathcal{T}} \left[(1-\delta_{1,1,t}-\delta_{1,2,t}) p^{w}_{k,1,t} + (1-\delta_{2,2,t}) p^{w}_{k,2,t}\right],$$

$$p^{on}_{k,3} = p^{on}_{k,2} + p^{b}_{k,3} - p^{a}_{k,3} = \sum_{t\in\mathcal{T}} \left[(1-\delta_{1,1,t}-\delta_{1,2,t}-\delta_{1,3,t}) p^{w}_{k,1,t} + (1-\delta_{2,2,t}-\delta_{2,3,t}) p^{w}_{k,2,t} + (1-\delta_{3,3,t}) p^{w}_{k,3,t}\right],$$

$\cdots$

$$\begin{aligned} p^{on}_{k,i} &= p^{on}_{k,i-1} + p^{b}_{k,i} - p^{a}_{k,i} \\ &= \sum_{t\in\mathcal{T}} \left[(1 - \sum_{i'=1}^{i} \delta_{1,i',t}) p^{w}_{k,1,t} + (1 - \sum_{i'=2}^{i} \delta_{2,i',t}) p^{w}_{k,2,t} + \cdots + (1-\delta_{i,i,t}) p^{w}_{k,i,t}\right] \\ &= \sum_{t\in\mathcal{T}}\sum_{i''\leq i,i''\in\mathcal{I}} (1 - \sum_{i'=i''}^{i} \delta_{i'',i',t}) p^{w}_{k,i'',t} \\ &= \sum_{t\in\mathcal{T}}\sum_{i''\leq i,i''\in\mathcal{I}} (1 - \sum_{i'=i''}^{i} \delta_{i'',i',t}) \cdot \widetilde{\mu}_{i'',t} \cdot e_{k,i'',t}. \end{aligned}$$

For clarity, we introduce a new variable $\gamma_{i,i'',t}$, $\forall t \in \mathcal{T}, i'' \leq i, i, i'' \in \mathcal{I}$, and let $\gamma_{i,i'',t} = 1 - \sum_{i'=i''}^{i} \delta_{i'',i',t}$. Hence, the number of in-vehicle passengers when the MV assigned to trip $k \in \mathcal{K}$ leaves stop $i \in \mathcal{I}$ can be expressed as follows:

$$\begin{aligned} p^{on}_{k,i} &= \sum_{t\in\mathcal{T}}\sum_{i''\leq i,i''\in\mathcal{I}} \gamma_{i,i'',t} \cdot \widetilde{\mu}_{i'',t} \cdot e_{k,i'',t} \\ &= \sum_{t\in\mathcal{T}}\sum_{i''\leq i,i''\in\mathcal{I}} \left(\gamma_{i,i'',t} \cdot \bar{\mu}_{i'',t} \cdot e_{k,i'',t} + \gamma_{i,i'',t} \cdot \theta_{i'',t} \cdot \check{\mu}_{i'',t} \cdot e_{k,i'',t}\right), \forall k \in \mathcal{K}, i \in \mathcal{I}. \end{aligned} \tag{55}$$

Combining constraints (55) with constraints (48), the right hand side of the constraints (48) for each possible combination of a stop and an MV assigned to a trip can be transformed into

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Xi}\left\{\sum_{t\in\mathcal{T}}\sum_{i''\le i,i''\in\mathcal{I}}\left(\gamma_{i,i'',t}\cdot\bar{\mu}_{i'',t}\cdot e_{k,i'',t}+\gamma_{i,i'',t}\cdot\theta_{i'',t}\cdot\check{\mu}_{i'',t}\cdot e_{k,i'',t}\right)\right\}\\&=\sum_{t\in\mathcal{T}}\sum_{i''\le i,i''\in\mathcal{I}}\gamma_{i,i'',t}\cdot\bar{\mu}_{i'',t}\cdot e_{k,i'',t}+\sup_{\boldsymbol{\theta}\in\Xi}\left\{\sum_{t\in\mathcal{T}}\sum_{i''\le i,i''\in\mathcal{I}}\theta_{i'',t}\cdot\gamma_{i,i'',t}\cdot\check{\mu}_{i'',t}\cdot e_{k,i'',t}\right\}.\end{aligned}\tag{56}$$

Further, based on the strong duality theory, the $sup\{\cdot\}$ problem in constraints (56) can be reformulated as

$$\begin{aligned}&\inf_{g_{k,i,d}\in\mathbb{R}+}\sum_{d=1}^{D}b_d\cdot g_{k,i,d}, && \forall k\in\mathcal{K},i\in\mathcal{I},\\&\text{s.t.}\quad\sum_{d=1}^{D}A_{d,(i''-1)|\mathcal{T}|+t}\cdot g_{k,i,d}=\gamma_{i,i'',t}\cdot e_{k,i'',t}\cdot\check{\mu}_{i'',t}, && \forall k\in\mathcal{K},i''\le i,i,i''\in\mathcal{I},t\in\mathcal{T},\\&\sum_{d=1}^{D}A_{d,m}\cdot g_{k,i,d}=0, && \forall k\in\mathcal{K},i\in\mathcal{I},m>i\,|\mathcal{T}|,\\&g_{k,i,d}\ge 0, && \forall k\in\mathcal{K},i\in\mathcal{I},d\in D.\end{aligned}\tag{57}$$

Combining with constraints (55)–(57), Proposition 3 holds. □

Finally, the equivalent MILP form of the DRO model considering the time-varying capacity (denoted as DRO-TVC) over the Wasserstein distance-based ambiguity set can be expressed based on the above derivations as

$$[\text{DRO}-\text{TVC})\quad\begin{cases}\inf\limits_{\mathbf{r},\mathbf{x},\sigma,\mathbf{c},\lambda,\mathbf{g}} & \left\{w_1\cdot\Delta\cdot\vartheta^T\cdot\left(\sum\limits_{k\in\mathcal{K}}\sum\limits_{i\in\mathcal{I}}\sum\limits_{t\in\mathcal{T}}\sum\limits_{\varsigma\in\mathcal{T}_t}\overline{\mu}_{i,\varsigma}\cdot v_{k,i,t,\varsigma}+\rho\sigma+\right.\right.\\ & \left.\left.\frac{1}{N}\sum\limits_{n=1}^{N}c_n\right)+w_2\cdot\sum\limits_{q\in\mathcal{Q}}\sum\limits_{k\in\mathcal{K}}\vartheta_q\cdot x_{k,q}\right\}\\ \text{s.t.} & \text{Constraints (1)–(6), (9)–(13), (18)–(19), (24),}\\ & \text{(35)–(39), (50)–(54).}\end{cases}\tag{58}$$

Proposition 4. *We propose the linear form of constraints (49) in the DRO formulation, which enables time-varying and station-wise capacity. Following a derivation similar to that in Proposition 3, we have*

$$\begin{cases}\sum\limits_{q\in\mathcal{Q}}C_q\cdot x_{k,i,q}\ge\sum\limits_{t\in\mathcal{T}}\sum\limits_{i''\le i,i''\in\mathcal{I}}\gamma_{i,i'',t}\cdot\bar{\mu}_{i'',t}\cdot e_{k,i'',t}+\sum\limits_{d=1}^{D}b_d\cdot g_{k,i,d},\forall k\in\mathcal{K},i\in\mathcal{I}\\ \text{Constraints (51)–(54).}\end{cases}\tag{59}$$

In accordance with the preceding analyses, we can construct the following MILP formulation with time-varying and station-wise capacity over the Wasserstein distance-based ambiguity set (DRO-TVSWC) as

$$[\text{DRO}-\text{TVSWC})\quad\begin{cases}\inf\limits_{\mathbf{r},\mathbf{x},\sigma,\mathbf{c},\lambda,\mathbf{g}} & \left\{w_1\cdot\Delta\cdot\vartheta^T\cdot\left(\sum\limits_{k\in\mathcal{K}}\sum\limits_{i\in\mathcal{I}}\sum\limits_{t\in\mathcal{T}}\sum\limits_{\varsigma\in\mathcal{T}_t}\overline{\mu}_{i,\varsigma}\cdot v_{k,i,t,\varsigma}+\rho\sigma+\right.\right.\\ & \left.\left.\frac{1}{N}\sum\limits_{n=1}^{N}c_n\right)+w_2\cdot\sum\limits_{i\in\mathcal{I}}\sum\limits_{q\in\mathcal{Q}}\sum\limits_{k\in\mathcal{K}}\widehat{\vartheta}_q\cdot x_{k,i,q}\right\}\\ \text{s.t.} & \text{Constraints (1)–(6), (9)–(13), (24), (30)–(31),}\\ & \text{(35)–(39), (59).}\end{cases}\tag{60}$$

Remark 4. In the case of $\rho=0$, the DRO-TVC (58) and DRO-TVSWC (60) are reduced into the SP models, respectively, namely, SP-TVC (27) and SP-TVSWC (32). Besides, when the value of ρ is large enough, the DRO-TVC (58) and DRO-TVSWC (60) are equivalent to the traditional robust optimization models (denoted as RO-TVC and RO-TVSWC).

4. Exact soultion method

In theory, our proposed MILP models can be optimally solved using commercial solvers in the case of small-scale problems. However, the size of the problem increases significantly in real-world applications, i.e., more stops and MVs are considered, and

therefore, the numbers of 0–1 variables grow exponentially, and the complexity of the coupling constraints increases dramatically, which poses a great challenge for realizing the solutions. To overcome this challenge, an exact decomposition algorithm is developed to solve the TT-DCA problem under uncertain passenger demand. Roughly speaking, we divide the original problem into two sub-problems. Following the integer L-sharped algorithm proposed by Laporte and Louveaux (1993), we consider the timetabling variables as the decisions in the first sub-problem. The dynamic-capacity allocation variables are then regarded as decisions in the second sub-problem, which depend on the decisions in the first sub-problem. At each iteration, the first sub-problem is solved, after which the timetabling decisions are fixed in the second sub-problem. Next, the second sub-problem is solved and fed back with a set of optimality cuts that can be added to the first sub-problem to raise the lower bound. This process is performed iteratively until a near-optimal solution is obtained. Further, we propose a set of tailored cuts based on the problem properties and design speed-up procedures to speed up the computation. The framework of our designed algorithm can be applied to the aforementioned models considering the time-varying capacity, as well as the time-varying and station-wise capacity. We illustrate our proposed algorithm in this section by solving the DRO-TVC model as an example. The solution process for the DRO-TVSWC model is detailed in Appendix B.

In Section 4.1, we present the decomposition of DRO-TVC and explore the mathematical theoretical properties to ensure the feasibility and convergence of the algorithm. The customized integer L-shaped method with a set of problem-based cuts is designed in Section 4.2. Finally, in Section 4.3, we describe the speed-up procedures used to improve the performance of the overall algorithmic framework.

4.1. Decomposition of the DRO-TVC model

Theoretically, DRO-TVC can be solved efficiently by GUROBI using the built-in B & B or cutting plane methods, since it is a linear programming formulation. We attempted to use GUROBI to solve the DRO-TVC for a real-world case based on the Beijing bus line. Unfortunately, no feasible solution was obtained after computing for more than 2 h. This can be attributed to two reasons: (1) The numbers of binary decision variables are enormous. (2) The binary decision variables related to the timetable and dynamic dispatching plans are closely coupled. Motivated by these observations, to obtain high-quality solutions within an acceptable computing time, a fundamental idea is to break the tight coupling between the timetable and the dispatching plan. Therefore, we divide the DRO-TVC into two sub-problems: (1) Determining the decision variables related to timetables (denoted by $\mathbf{r}$) with the objective of minimizing the total passenger waiting costs, and (2) using the timetables as an input to optimize the dynamic-capacity allocation plan (i.e., the number of MUs in each MV, denoted by $\mathbf{x}$), that is, $\mathbf{r}$ is the input parameter for the second sub-problem.

Next, we construct the following sub-problems (denoted by SP1 and SP2), i.e.,

$$[\text{SP1}] \quad \begin{cases} \inf\limits_{\mathbf{r},\sigma,\mathbf{c},\lambda,\mathbf{g}} & w_1 \cdot \Delta \cdot \vartheta^T \cdot \Big(\sum\limits_{k\in\mathcal{K}} \sum\limits_{i\in I} \sum\limits_{t\in\mathcal{T}} \sum\limits_{\varsigma\in\mathcal{T}_t} \overline{\mu}_{i,\varsigma} \cdot v_{k,i,t,\varsigma} + \rho\sigma + \dfrac{1}{N} \sum\limits_{n=1}^{N} c_n \Big) \\ \text{s.t.} & \text{Constraints (1)–(6), (9)–(13), (24), (35)–(39), (51)–(54).} \end{cases}$$

$$[\text{SP2}] \quad \begin{cases} \min\limits_{\mathbf{r}^*,\mathbf{x}} & \sum\limits_{q\in\mathcal{Q}} \sum\limits_{k\in\mathcal{K}} \vartheta_q \cdot x_{k,q} \\ \text{s.t.} & \text{Constraints (18)–(19), (50)–(54).} \end{cases}$$

In the following discussions, we explore the mathematical properties of these sub-problems. For clarity, the SP2 is expressed as $U(\mathbf{r}^*,\mathbf{x}) = \min_{\mathbf{x}}\{\sum_{q\in\mathcal{Q}}\sum_{k\in\mathcal{K}}\vartheta_q \cdot x_{k,q} | \text{Constraints (18)–(19), (50)–(54)}\}$, where $\mathbf{r}^*$ refers to the solution of variable $\mathbf{r}$ obtained by solving SP1.

Lemma 2. *There exists a finite lower bound $L = |\mathcal{K}| \cdot \vartheta_q$ with $q = 1$ satisfying $L \le U(\mathbf{r}^*,\mathbf{x})$.*

Proof. In this paper, we do not consider cancelling trips, so that at least $|\mathcal{K}|$ MUs (i.e. each MV only consist of one MU in this case) are required. On this basis, considering a scenario with relatively smaller passenger demand, where one MU is sufficient to ensure that every waiting passenger gets on board, then the lower bound of the total dispatching costs can be calculated as $L = |\mathcal{K}| \cdot \vartheta_q$ with $q = 1$. Hence, we have $U(\mathbf{r}^*,\mathbf{x}) \ge |\mathcal{K}| \cdot \vartheta_q$ with $q = 1$, which derives the proof. □

4.2. Customized integer L-shaped method

Since both of the above subproblems contain integer decision variables, traditional decomposition approaches (e.g. Benders decomposition) cannot be applied. In this case, the integer L-shaped method is an useful approach for some MILP formations, wherein all decomposed sub-problems can all be ILP models. Laporte and Louveaux (1993) is the first work that explored this algorithm. Subsequently, Angulo et al. (2016) has devoted efforts to make some advances for further improving its computational performance. Based on the basic framework of the integer L-shaped method, we develop a customized integer L-shaped method to effectively solve the problem under investigation.

First, we propose the following set of tailored cuts to ensure that SP2 can always find a feasible vehicle dispatching solution, i.e., the dynamic-capacity allocation plan of MVs assigned to trips, with each timetable generated by SP1. This can prevent inefficiencies caused by the impossibility.

$$\sup_{\boldsymbol{\theta}\in\Xi}\{p^{on}_{k,i}\} \le C_{|\mathcal{Q}|}, \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}. \tag{61}$$

For clarify, we present a straightforward demonstration of the validity of inequalities (61) in the following. If the MV assigned to a trip that departs the stop with in-vehicle passengers exceeding the maximum capacity (i.e., the capacity of the vehicle with the maximum number of MUs), there would be no feasible dispatching solution that can satisfy the constraints (48).

Further, like Proposition 3, constraints (61) can be equivalently transformed into the following linear form

$$\sum_{t\in\mathcal{T}}\sum_{i''\le i,i''\in\mathcal{I}} \gamma_{i,i'',t}\cdot\bar{\mu}_{i'',t}\cdot e_{k,i'',t} + \sum_{d=1}^{D} b_d\cdot g_{k,i,d} \le C_{|\mathcal{Q}|}, \qquad \forall k\in\mathcal{K}, i\in\mathcal{I}, \tag{62}$$

which is then employed as one of the constraints of SP1.

Next, we reformulate the first sub-problem (i.e., SP1) using the integer L-shaped method. To incorporate the information of MP2 into MP1, we successively add the optimality cuts proposed by Laporte and Louveaux (1993) and Angulo et al. (2016) into MP1. Specifically, the optimality cut proposed by Laporte and Louveaux (1993) can be formulated for the explored problem as follows

$$\eta \ge (U(\mathbf{r}^*,\mathbf{x}) - L)\left(\sum_{r_{k,i,t}\in H(r^*)} r_{k,i,t} - \sum_{r_{k,i,t}\notin H(r^*)} r_{k,i,t} - \left|H(r^*)\right|\right) + U(\mathbf{r}^*,\mathbf{x}), \tag{63}$$

where η represents the estimated value of $U(\mathbf{r},\mathbf{x})$, and $H(r^*) = \{r_{k,i,t}|r_{k,i,t}=1\}$. Moreover, L represents the lower bound of $U(\mathbf{r},\mathbf{x})$, whose values can be set based on Lemma 2. It is worth clarifying that the optimality cut (63) is added only when $\eta < U(\mathbf{r},\mathbf{x})$. Since the aforementioned optimality cuts are not tight for solutions other than $\mathbf{r}^*$, we further employ the following continuous L-shaped optimality cut proposed by Angulo et al. (2016), that is,

$$\eta \ge \mathbf{s}(\mathbf{r}-\mathbf{r}^*) + U_{LP}(\mathbf{r}^*,\mathbf{x}), \tag{64}$$

where $\mathbf{s}$ represents the subgradient vector of $U(\mathbf{r},\mathbf{x})$ at $\mathbf{r}^*$, and $U_{LP}(\mathbf{r}^*,\mathbf{x})$ represents the continuous relaxation of $U(\mathbf{r}^*,\mathbf{x})$ that resets the binary variable $x_{k,q}$ as a continuous variable. The subgradient L-shaped optimality cut (64) is not necessarily required for the convergence of the L-shaped method; however, it enhances the computing performance of this algorithm.

Finally, we reformulate SP1 based on the above analyses and construct the current problem (CP) at each node of the branch-and-cut search tree as

$$[\text{CP}]\quad \begin{cases} \inf\limits_{\mathbf{r},\mathbf{x},\sigma,\mathbf{c},\boldsymbol{\lambda},\mathbf{g},\eta} w_1\cdot\Delta\cdot\vartheta^T\cdot\left(\sum\limits_{k\in\mathcal{K}}\sum\limits_{i\in\mathcal{I}}\sum\limits_{t\in\mathcal{T}}\sum\limits_{\varsigma\in\mathcal{T}_t}\overline{\mu}_{i,\varsigma}\cdot v_{k,i,t,\varsigma} + \rho\sigma + \dfrac{1}{N}\sum\limits_{n=1}^{N} c_n\right) + \eta \\ \text{s.t.} \quad \eta \ge (U(\mathbf{r}^*,\mathbf{x}) - L)\left(\sum\limits_{(k,i,t)\in H(r^*)} r_{k,i,t} - \sum\limits_{(k,i,t)\notin H(r^*)} r_{k,i,t} - \left|H(r^*)\right|\right) + U(\mathbf{r}^*,\mathbf{x}). \\ \quad\quad \eta \ge \mathbf{s}(\mathbf{r}-\mathbf{r}^*) + U_{LP}(\mathbf{r}^*,\mathbf{x}). \\ \quad\quad \text{Constraints (1)–(6), (9)–(13), (24), (35)–(39), (50)–(54), (62).} \end{cases} \tag{65}$$

In summary, the main procedure of our designed decomposition-based branch-and-cut algorithm for DRO-TVC can be briefly presented as follows: (1) Relax CP (65) by discarding constraints (5) and solving it to the optimum. It is worth noting that no optimality cut or subgradient cut has been generated yet. (2) Determine whether to continue solving. If the stopping criteria are satisfied, the algorithm is stopped. Otherwise, we select a $r_{k,i,t}\notin\{0,1\}$ and then branch at this node. (3) Repeat the previous step until an integer solution $\mathbf{r}^*$ is obtained. In this process, it is necessary to determine whether to add the optimality cuts (63) and (64).

In the specific implementation, we construct a branch-and-cut search tree using the state-of-the-art commercial solver GUROBI. The subgradient cut (64) and optimality cut (63) are added to the CP step-by-step using the lazy constraint callback function. During the search process, the callback function is executed at the node where the relaxed CP is solved to the optimum, or a new MIP incumbent is found. The pseudocode of the proposed decomposition-based branch-and-cut algorithm is presented as Algorithm 1.

Algorithm 1 The integer L-shaped method for the DRO-DCA model.

Initialization: Input the lower bound L of the subproblem, root node of the relaxed current problem. Create an empty list of nodes and put the root node into that list. Set the minimum gap ϵ, upper bound UB = $+\infty$ and lower bound LB = L. Let Gap = $(UB - LB)/UB$.
While Gap $\leq \epsilon$ **do**
  if the node list is empty **then**
    break.
  else
    Select a pendent node from the node list;
    Solve the CP corresponding to the selected node;
    if the CP is infeasible **then**
      Fathom the current node and go to line 3;
    else
      Let $(\mathbf{r}^*, \eta^*)$ be the optimal solution of the CP;
      if $Z_{wait}(\mathbf{r}^*) + \eta^* \geq$ UB **then**
        Fathom the current node and go to line 3;
        if some variables in $\mathbf{r}$ are non-binary **then**
          Select and branch a non-binary variable. Delete the current node and append the new two branch nodes to the pendent node list. go to line 3;
        else
          Compute the linear relaxation of $U(\mathbf{r}, \mathbf{x})$ and get the corresponding optimal objective value U_{LP};
          if $U_{LP} > \eta^*$
            Add the corresponding subgradient cut (64) to CP and go to line 7;
          else
            Call GUROBI to solve the subproblem $U(\mathbf{r}^*, \mathbf{x})$, denote the optimal value as U_0.
            if $Z_{wait}(\mathbf{r}^*) + U_0 <$ UB
              Update UB = $Z_{wait}(\mathbf{r}^*) + U_0$
            end if
            if $U_0 < \eta^*$
              Add the corresponding integer optimality cut (63) to CP and go to line 7.
            end if
          end if
        end if
      end if
    end if
  end if
end while
Output: The optimal solution of DRO-OPMV.

4.3. Family of problem-based valid equalities

To further improve the computational efficiency, we propose the following valid equalities related to "time window" to narrow the searching space of CP considering the restrictions of adjustments related to timetables.

Proposition 5. *The following constraints*

$$r_{k,i,t} = \begin{cases} 1, & 1 \leq t \leq \underline{\zeta}_{k,i} \\ 0, & \overline{\zeta}_{k,i} \leq t \leq |\mathcal{T}|, \end{cases} \qquad \forall k \in \mathcal{K}, i \in \mathcal{I}, \tag{66}$$

are a family of valid equalities for the set of feasible solutions of DRO-TVC, where $\underline{\zeta}_{k,i}$ and $\overline{\zeta}_{k,i}$ represent the earliest and latest departure times of the MV assigned to trip k leaving stop i, respectively.

Proof. First of all, according to constraints (10), we have

$$t^d_{k,1} = t^a_{k,1} + t^p_{k,1} + t^h_{k,1}$$

$$t^d_{k,2} = t^a_{k,2} + t^p_{k,2} + t^h_{k,2} = t^a_{k,1} + t^r_{k,1} + \sum_{i'=1}^{2} (t^p_{k,i'} + t^h_{k,i'})$$

...

$$t^d_{k,i} = t^a_{k,i} + t^p_{k,i} + t^h_{k,i} = t^a_{k,1} + \sum_{i'=1}^{i-1} t^r_{k,1} + \sum_{i'=1}^{i} (t^p_{k,i'} + t^h_{k,i'}).$$

In real-world operations, buses do not travel at speeds lower than the free-flow speed. Moreover, without a speed control strategy, drivers choose speed based on real-time traffic conditions rather than deliberately slowing down. Therefore, based on historical data, bounds on the time-dependent travel times of MV k can be obtained by preprocessing prior to optimization, i.e.,

$$\mathrm{TR}^{min}_i \le t^r_{k,i} \le \mathrm{TR}^{max}_i, \tag{67}$$

where $\mathrm{TR}^{min}_i = \min_{t\in\mathcal{T}}\{\tau_{i,t}\}$, and $\mathrm{TR}^{max}_i = \max_{t\in\mathcal{T}}\{\tau_{i,t}\}$

Further, combining the inequalities (67) with constraints associated with the adjustment of shifting and holding times, i.e., constraints (12) and (13), the bounds of $t^d_{k,i}$ can be further formulated as

$$\underline{\zeta}_{k,i} \le t^d_{k,i} \le \overline{\zeta}_{k,i}, \quad \forall k \in \mathcal{K}, i \in \mathcal{I}, \tag{68}$$

where $\underline{\zeta}_{k,i} = t^0_{k,1} + \underline{\alpha}_k + \sum_{i'=1}^{i-1} \mathrm{TR}^{min}_{i'}$, and $\overline{\zeta}_{k,i} = t^0_{k,1} + \overline{\alpha}_k + \sum_{i'=1}^{i-1} \mathrm{TR}^{max}_{i'} + \sum_{i'=1}^{i} \overline{\beta}_{k,i'}$.

To sum up, on the one hand, constraints (68) indicate that the MV assigned to trip k is unlikely to depart from stop i before time interval $\underline{\zeta}_{k,i}$ or after time interval $\overline{\zeta}_{k,i}$. On the other hand, the binary indicator $r_{k,i,t}$ is defined as equal to 1 if the MV assigned to trip k has not depart from stop i before time interval t, and 0, otherwise. Hence, Proposition 5 holds. The proof is complete. □

Remark 5. We need to mention that, the problem-based valid equalities (66) are also employed in the following benchmark experiments, which are directly solved by GUROBI.

5. Numerical experiments

In this section, we report extensive computational results with the four main goals: (1) to investigate the effectiveness of the customized decomposition algorithm; (2) to explore the benefits of enabling the time-varying and station-wise capacity; (3) to analyze the superiority of the DRO approach; and (4) to evaluate the impact of the Wasserstein radius on the robustness of optimal solutions.

The utilized parameter settings are described in Section 5.1. Section 5.2 compares the performance of the proposed solution methodologies (i.e., GUROBI and our proposed algorithm) for solving the DRO models. Section 5.3 compares the optimal solutions under the two dynamic-capacity allocation strategies in detail, and Section 5.4 contrasts the DRO solutions with stochastic solutions on the out-of-sample dataset. Finally, Section 5.5 compares the performance of the DRO, SP, and RO formulations under various Wasserstein radius to support the operators when they select the parameters in accordance with the practical situations. The experiments were implemented in Python on a Windows 11 personal computer with 12th Gen Intel(R) Core(TM) i7-12700H and 32 GB RAM.

5.1. Case description and parameter settings

In this case study, we tested our models and the solution method using data collected from the Bus Line 468 in Beijing, which has 18 stops with a total length of approximately 12.9 km, as shown in Fig. 8. Bus Line 468 links residential communities and a public transport hub, and it connects Beijing Metro lines 1 and 6; therefore, it has a relatively high passenger demand during peak periods. Operational data on the working days in September 2018 are used in this case study. We adopted the data-processing method proposed in Zhang et al. (2021a) to obtain the dynamic travel times between two adjacent stops and the time-dependent arrival rates of passengers at each stop. Besides, the dynamic passenger demand data can also be obtained using advanced forecasting techniques (Yan et al., 2022). Further, we considered the minimum arrival rate at each stop and each time interval as the nominal value (i.e., $\overline{\mu}_{i,t}$) and the difference between the maximum and minimum values as the deviation at each time interval (i.e., $\breve{\mu}_{i,t}$) to depict the randomness of passenger demand. Thus, the realized values of $\theta_{i,t}$ for $i \in \mathcal{I}, t \in \mathcal{T}$, i.e., the uncertain perturbation of passenger demand at each stop on the time interval in each sample (i.e., one working day) can be derived simultaneously. In Fig. 9, we demonstrate the processing results with respect to the aforementioned parameters in each time interval (with 1-min duration) from 5:30 AM to 23:00 PM. The travel times and passenger arrival rates are significantly dynamic, and the latter is particularly characterized in our proposed Wasserstein distance-based ambiguous set with time-dependent fluctuations. Moreover, demand exhibits a clear uncertainty, particularly during peak hours.

We selected 7:00 AM to 10:00 AM as the study time horizon, which contains the entire morning peak period and a portion of the off-peak hours. The duration of each time interval is 1 min. For the marginal dispatching costs, we follow the mechanism proposed by Zhang et al. (2018) and Chen et al. (2019), where the value is equal to 0.59($). The fixed cost is set as 19.12($) following Dai et al. (2020). In addition, the monetary value of the unit waiting time is 0.8($/minute) according to Chen et al. (2019). For the number of MUs in an MV, the standards provided by the dynamic autonomous road transit (DART) in Singapore (https://www.tum-create.edu.sg/) are referenced. The capacity of each MU is 30 passengers, and one MV can be composed of a maximum of 10 MUs. However, considering the limited road resources on the Bus Line 468 in Beijing, the maximum number of MUs is set to five in this study. The aim of this study is to generate service-oriented data-driven robust timetables and dynamic dispatching plans; therefore, the weight associated with passenger benefits (i.e., w_1) is set to 2, whereas $w_2 = 1$. In addition, the

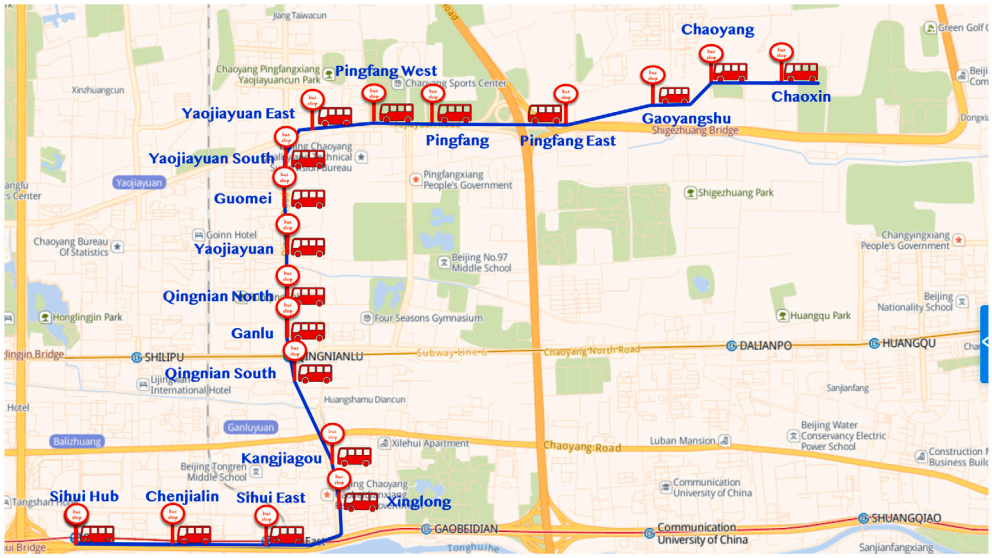
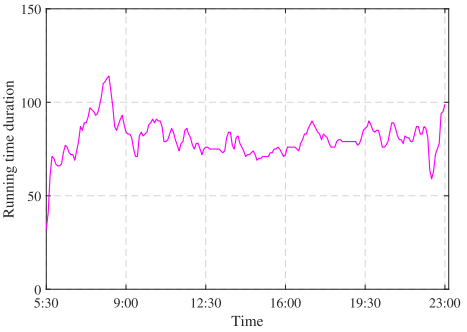
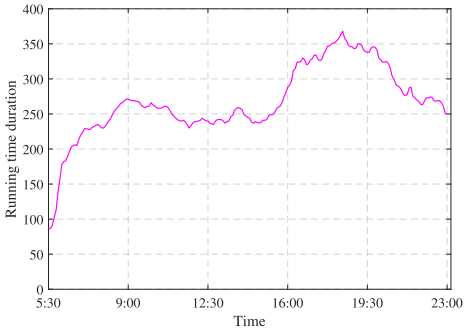


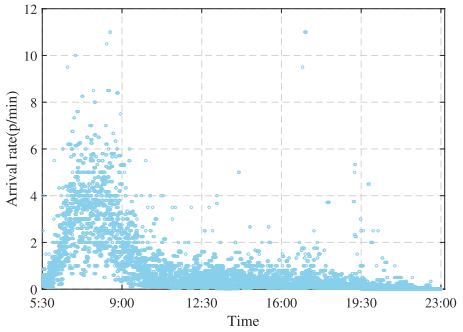
Fig. 8. Illustration of Bus Line 468 in Beijing.
Source: Gaode Maps.



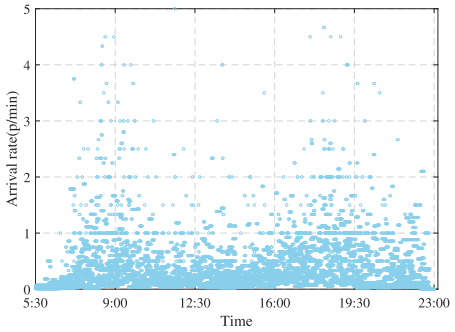
(a) Travel time between stops Chaoxin and Chaoyang



(b) Travel time between stops Qingnian North and Ganlu



(c) Arrival rates at stop Chaoxin among different samples



(d) Arrival rates at stop Qingnian North among different samples

Fig. 9. Illustrations of the average travel times of all samples and arrival rates among different samples.

values of the parameters $\underline{\beta}_k$, Γ_i, and Λ_i are set according to the historical data; $\overline{\alpha}_k$, $\underline{\alpha}_k$, and $\overline{\beta}_k$ are 2, -2, and 2 min, respectively. A termination condition was considered in the following numerical experiments to balance the solution quality and computational time; i.e., the computation could be terminated when the relative gap is no greater than 5%. For the results reported in Sections 5.2, 5.3, and 5.4, the parameter ρ is set as 1, whereas it is varied in Section 5.5.

Table 3
Performance comparison between the Benchmark and the proposed algorithm for solving DRO-TVC.

($\|\mathcal{I}\|, \|\mathcal{K}\|, \|\mathcal{T}\|$)	Solution method	Best upper bound	Best lower bound	Relative gap (%)	Computing time (s)
(5, 3, 50)	GUROBI	472.733	455.840	3.57	1.34
	Integer L-shaped method	470.778	448.030	4.83	2.12
(10, 10, 110)	GUROBI	3398.466	3227.464	4.99	87.28
	Integer L-shaped method	3362.86	3197.531	**4.92**	**60.71**
(10, 12, 120)	GUROBI	4159.263	3988.015	4.12	24.89
	Integer L-shaped method	4193.788	3985.153	4.97	102.31
(10, 15, 140)	GUROBI	5120.827	4890.739	4.49	713.38
	Integer L-shaped method	5032.374	4796.788	4.68	282.28
(15, 10, 130)	GUROBI	4320.044	4263.892	1.30	1995.83
	Integer L-shaped method	4365.543	4205.518	3.67	432.32
(15, 12, 150)	GUROBI	5642.315	5369.172	4.84	3142.47
	Integer L-shaped method	5593.654	5344.061	**4.46**	**630.71**
(15, 15, 160)	GUROBI	6755.381	6426.691	4.87	1860.55
	Integer L-shaped method	6589.074	6302.527	**4.35**	**1115.05**
(18, 10, 150)	GUROBI	4729.405	4411.610	6.72	7200.00
	Integer L-shaped method	4530.301	4306.826	**4.97**	**905.94**
(18, 12, 160)	GUROBI	9421.677	5543.441	41.20	7200.00
	Integer L-shaped method	5709.350	5447.774	**4.58**	**1207.67**
(18, 15, 180)	GUROBI	–	–	–	7200.00
	Integer L-shaped method	6862.426	6534.242	**4.78**	**2512.75**

[1] The bolded metrics indicate that our proposed algorithm outperforms GUROBI in both relative gap and computing time.

Table 4
Performance comparison between the Benchmark and the proposed algorithm for solving DRO-TVSWC.

($\|\mathcal{I}\|, \|\mathcal{K}\|, \|\mathcal{T}\|$)	Solution method	Best upper bound	Best lower bound	Relative gap (%)	Computing time (s)
(5, 3, 50)	GUROBI	457.518	435.787	4.75	1.38
	Integer L-shaped method	455.504	436.233	4.23	3.36
(10, 10, 110)	GUROBI	3302.799	3140.988	4.90	136.38
	Integer L-shaped method	3231.517	3084.098	**4.56**	**76.16**
(10, 12, 120)	GUROBI	4140.01	3942.276	4.78	204.56
	Integer L-shaped method	4051.273	3861.519	**4.68**	**128.31**
(10, 15, 140)	GUROBI	5433.686	5237.488	3.61	379.87
	Integer L-shaped method	5317.386	5106.982	3.96	978.21
(15, 10, 130)	GUROBI	4203.554	4155.274	1.15	1900.37
	Integer L-shaped method	4248.171	4058.048	4.48	400.23
(15, 12, 150)	GUROBI	5473.516	5237.384	4.31	3059.65
	Integer L-shaped method	5342.299	5141.377	**3.76**	**1045.31**
(15, 15, 160)	GUROBI	6400.694	6288.512	1.75	4940
	Integer L-shaped method	6430.039	6110.227	4.97	630.36
(18, 10, 150)	GUROBI	4506.3577	4280.43737	5.00	4442.32
	Integer L-shaped method	4429.398	4218.527	**4.76**	**1374.29**
(18, 12, 160)	GUROBI	9524.614	5396.208	43.30	7200.00
	Integer L-shaped method	5552.899	5276.046	**4.99**	**1663.34**
(18, 15, 180)	GUROBI	–	–	–	7200.00
	Integer L-shaped method	6728.546	6393.676	**4.98**	**5235.25**

[1] The bolded metrics indicate that our proposed algorithm outperforms GUROBI in both relative gap and computing time.

5.2. Performance comparison between different solution methods

We focus on the evaluation of the performance of the proposed integer L-shaped algorithm in this subsection. To this end, we consider the state-of-the-art solver GUROBI as a benchmark for solving the DRO-TVC and DRO-TVSWC models. Based on real-world operating data, we generate a total of 10 instances by presetting the different combinations of parameters (i.e., the numbers of stops $|\mathcal{K}|$, trips $|\mathcal{I}|$, and time intervals $|\mathcal{T}|$). In particular, we take $|\mathcal{K}| = \{3, 10, 12, 15\}$, $|\mathcal{I}| = \{5, 10, 15, 18\}$ and $|\mathcal{T}| = \{50, 110, 120, 130, 140, 150, 160, 180\}$ into account. Considering that it is challenging to solve large-scale cases using GUROBI, we set the maximum computing time based on the scale of the instances.

Tables 3 and 4 provide detailed comparative results for solving the DRO-TVC and DRO-TVSWC models, respectively. We report the specific problem size, solution method, best bounds, relative gap (i.e., the gap between the lower and upper objective bound), and computing time of each instance. From the experimental results in Tables 3 and 4, we conclude that GUROBI is an efficient

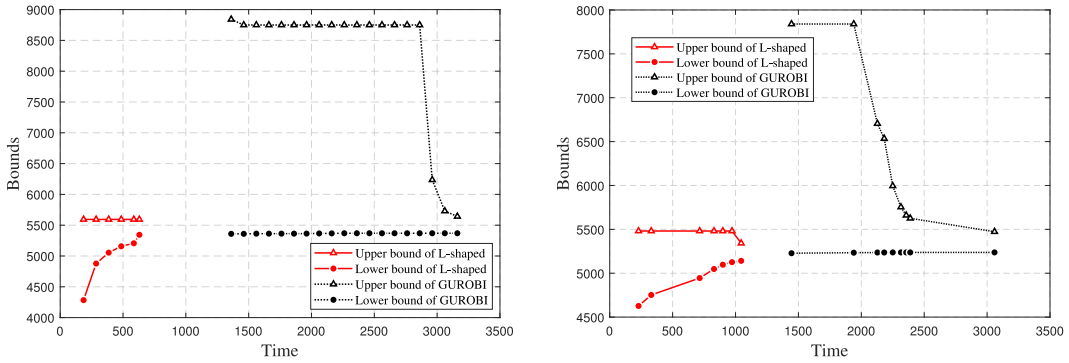


Fig. 10. Convergence trends of the upper bounds and lower bounds with respect to the objective values on instance with $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (15, 12, 150)$.

Table 5
Performance comparison between the proposed algorithm and the combined solution method in solving DRO-TVC with a time limit of 120 min.

$(\lvert\mathcal{I}\rvert, \lvert\mathcal{K}\rvert, \lvert\mathcal{T}\rvert)$	Solution method	Best upper bound	Best lower bound	Relative gap (%)	Computing time (s)
(18, 15, 180)	Integer L-shaped method	6862.426	6534.242	4.78	2512.75
	Integer L-shaped method + GUROBI	6862.426	6534.242	4.78	7200.00

Table 6
Performance comparison between the proposed algorithm and the combined solution method in solving DRO-TVSWC with a time limit of 120 min.

$(\lvert\mathcal{I}\rvert, \lvert\mathcal{K}\rvert, \lvert\mathcal{T}\rvert)$	Solution method	Best upper bound	Best lower bound	Relative gap (%)	Computing time (s)
(18, 15, 180)	Integer L-shaped method	6728.546	6393.676	4.98	5235.25
	Integer L-shaped method + GUROBI	6728.546	6393.676	4.98	7200.00

solver only for small-scale instances (e.g., $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (5, 3, 50)$ and $(10, 12, 120)$). However, the superiority of the integer L-shaped algorithm with respect to GUROBI becomes more clear with an increase in the size of the problem. For example, for $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (10, 10, 110)$, the integer L-shaped algorithm can find an approximate optimal solution with a relative gap of 4.92% within 61 s for DRO-TVC and a near-optimal solution with a gap of 4.56% after 76.16 seconds' computation for DRO-TVSWC. In contrast, GUROBI requires 87.28 and 136.38 s to generate solutions with larger gaps. In this relatively small-scale problem, the difference in performance is not significant. Nevertheless, for $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (18, 15, 180)$, our proposed algorithm can find solutions in 2512.75 and 5235.25 s with gaps below 5%, whereas GUROBI cannot even find a feasible solution after 7200 s of computation for both models; this demonstrates the effectiveness of our decomposition algorithm with problem-based cuts when dealing with large-scale problems that occur frequently in practice. In addition, we observe that the optimal lower bound obtained by the integer L-shaped algorithm is slightly inferior to that of GUROBI; however, the difference between the upper bounds of the two is relatively small, and sometimes, the upper bound of the integer L-shaped algorithm is even better than that of GUROBI.

Next, we present the convergence trends of the upper bounds and lower bounds of the objective function values for GUROBI and the proposed integer L-shaped method. The convergence trends for these cases are similar, and therefore, for brevity, only the results of DRO-TVC and DRO-TVSWC with $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (15, 12, 150)$ are shown in Fig. 10. These results provide the following observations: (1) The proposed integer L-shaped method can find a better first feasible solution than that of GUROBI, which is close to the near-optimal solution obtained at the termination of the computation. (2) Although the lower bound obtained by the integer L-shaped algorithm is not satisfactory at the beginning, the rate of improvement of the lower bound is more desirable. (3) It is difficult to improve the lower bound during the B & B search process used by GUROBI and to find a feasible solution when the problem scale is relatively large. (4) The convergence rate of GUROBI is slower than that of the proposed integer L-shaped method, especially for large-scale instances. In conclusion, the proposed integer L-shaped algorithm clearly outperforms GUROBI in terms of computational efficiency, optimal bounds, and convergence speed, especially for a large problem size.

Besides, we are particularly interested in combining the integer L-shaped algorithm with GUROBI. We have conducted a set of experiments for both two DRO models to evaluate the effectiveness of this combined solution method. Specifically, we have selected solutions with a gap of less than 5% obtained from the integer L-shaped method and fed them into GUROBI. Additionally, we have incorporated the lower bound derived from the integer L-shaped method into GUROBI to speed up computations. Tables 5 and 6 report the results of solving the two DRO models, where "Integer L-shaped method + GUROBI" represents this combined approach. It can be seen that GUROBI cannot find a better solution for both models after two-hour computations. To sum up, the combined solution method does not yield any noticeable enhancements in our study.

Remark 6. To verify the effectiveness of the proposed algorithm in other passenger flow scenarios, the other dataset of passenger flow was randomly generated. The experimental results of DRO-TVC are shown in Table 9 and Fig. 17 (see Appendix C). It can

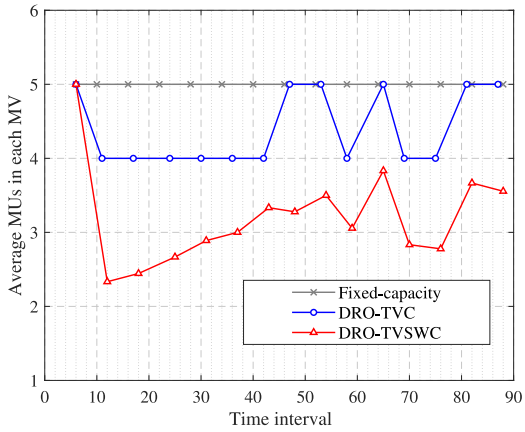


Fig. 11. Comparison of the vehicle dispatching plans.

Table 7
Performance comparison of DRO-TVC and DRO-TVSWC with different dynamic-capacity allocation strategies.

Model	Objective value (unit: $)		Total number of utilized MUs
	Waiting costs of passengers	Operating costs of operators	
DRO-TVC	3707.411	930.566	81
DRO-TVSWC	3736.678	749.858	44
Objective improvement (%)	−0.789	19.419	45.679

be seen that our proposed integer L-shaped algorithm is superior in different sets of passenger flow scenarios, proving its wider applicability.

5.3. Comparison between different dynamic-capacity allocation strategies

In this subsection, we focus on evaluating performance of timetabling using three capacity allocation strategies. The first strategy involves static (fixed) capacity allocation plans, the second strategy involves dynamically adjusting the capacity of MVs assigned to different trips considering the time-varying capacity, and the third strategy involves enabling the time-varying and station-wise capacity. For this set of experiments, we use the practical existing operational plan, which is based on a uniform headway and a fixed formation (i.e, the aforementioned first strategy), as our benchmark. The operating strategies adopted in the benchmark is the most common plans used in practice (Ceder, 2016).

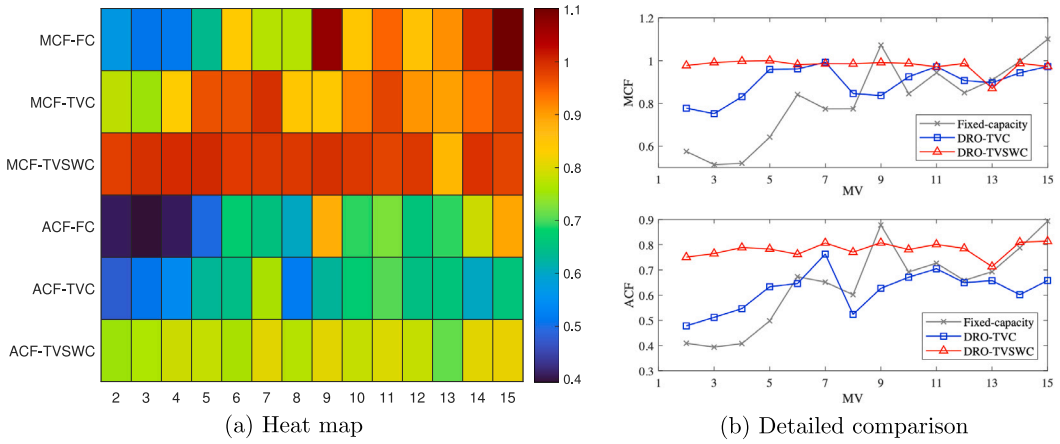
Fig. 11 shows the number of MUs in the MV assigned to each trip and the departure time at the first stop of the MV assigned to each trip. It can be seen that the number of utilized MUs with static (fixed) capacity allocation plans is the largest. This is expected because the train capacity in this case cannot be adjusted to the time-dependent and spatially uneven passenger demand, which leads to a waste of redundant capacity. Further, Table 7 compares the performances of DRO-TVC and DRO-TVSWC in detail. These results yield two observations: (1) With the time-varying and station-wise capacity allocation strategy, the operating costs and the number of utilized MUs are evidently reduced (by approximately 20% and 46%, respectively) compared with the strategy with only the time-varying capacity. The results reveal the benefits of enabling MUs to be docked/undocked at each stop for dispatching vehicles driven by passenger demand. (2) The time-varying and station-wise capacity strategy results in an increase of approximately 0.8% in the objective value with respect to the waiting costs of passengers.

In addition, we are interested in the vehicle congestion during operation. The following indices

$$\mathrm{ACF}_k = \frac{1}{|I|}\sum_{i\in I}\frac{\sup_{\boldsymbol{\theta}\in\Xi}\{p^{on}_{k,i}\}}{\sum_{q\in Q}\vartheta_q\cdot x_{k,q}}, \forall k\in\mathcal{K}, \tag{69}$$

$$\mathrm{MCF}_k = \max_{i\in I}\left\{\frac{\sup_{\boldsymbol{\theta}\in\Xi}\{p^{on}_{k,i}\}}{\sum_{q\in Q}\vartheta_q\cdot x_{k,q}}\right\}, \forall k\in\mathcal{K}, \tag{70}$$

are employed to explore the average and maximum congestion of vehicles during operation, respectively. More precisely, ACF_k represents the average full-loading rate of MV assigned to trip k, which reflects the average level of the in-vehicle congestion in the worst-case scenario, and MCF_k represents the maximum level of in-vehicle congestion of MV k. A heat map is plotted to visualize the average and maximum levels of in-vehicle congestion in Fig. 12(a). A more reddish shade implies a greater level of congestion, whereas a bluer shade indicates little congestion and considerable wasted capacity. An ideal level in terms of passengers would therefore be blue, whereas the desired situation for operators would be on the red side, but not too red, and it may lie between the colors represented by values between 0.8 and 1. The results shown in Fig. 12 help us conclude the following.



(a) Heat map (b) Detailed comparison

Fig. 12. Comparison of the maximum and average congestion among three strategies.

(a) Under the operational plan with fixed capacity, the maximum level of congestion (i.e., MCF-FC) for some MVs would exceed 1, e.g., the MVs assigned to trips 9 and 15. This can result in passengers being stranded and unable to board the expected MVs assigned to trips, especially in the case of regular epidemic prevention and control, which will significantly reduce their satisfaction. Further, the average levels of congestion (i.e., ACF-FC) are all less than 0.5 during the first 5 trips, indicating the serious waste of capacity. This can be primarily attributed to there being lower demand during this period, whereas the capacity is fixed.

(b) Through the robust collaborative optimization of timetables and the dynamic-capacity allocation strategy with time-varying capacity, the maximum levels of congestion (i.e., MCF-TVC) of MVs are all between 0.75 and 1, which ensure that all waiting passengers can leave on the desired MVs. Further, compared with the operational plan with fixed capacity, it has a smaller range of variations in the MCF and ACF with a more stable capacity occupancy. Moreover, this demonstrates that the time-varying capacity can dynamically match time-dependent demand. These facts (e.g., a fewer number of in-service MUs, a more comfortable level of congestion, and a guarantee of being able to board the desired MVs) support that our proposed dynamic-capacity allocation approach can reduce the operating costs for operators while providing passengers with high-quality services.

(c) The results of DRO-TVSWC, which enables the time-varying and station-wise capacity, remain at the top in terms of the maximum and average levels of congestion. The maximum and average levels of congestion of each vehicle largely remained at approximately 1 and 0.8, respectively. These results indicate that, compared with the solutions obtained with fixed capacity and timetables, our proposed method of integrating the timetabling and the time-varying and station-wise capacity can provide more effective and flexible solutions where the majority of MVs assigned to trips can be well matched to the time-dependent and unevenly distributed passenger demand at each stop. In addition, it is worth noting that our method is able to reduce the redundancy of capacity when passenger demand is lower. For example, in the first four trips, our solution yields improved average full-loading rates of MVs assigned to trips. In practice, operators may profit from the observed improvements.

Moreover, we compare the computational results with respect to the congestion of the MV assigned to each trip. Considering the stop Kangjiagou as an example, a detailed comparison of the in-vehicle and waiting passengers in the *worst-case* scenario with the three strategies is displayed in Fig. 13(a) and Fig. 13(b), respectively. Specifically, the first strategy is using the fixed timetable with even headway and fixed capacity; the second one is the integrated optimization of timetabling and time-varying capacity; and the last one is the integration of timetabling and station-wise capacity. In other words, the timetables and allocation plans of capacity in all three solutions are different from each other. From the results depicted in Fig. 13(a), it can be observed that the maximum number of in-vehicle passengers has decreased, which highlights the reduction of the maximum congestion in vehicles and leads to improved travel comfort. Besides, Fig. 13(b) reports the number of passengers waiting for trips at this stop. It can be seen that the maximum number of waiting passengers has also decreased, illustrating a reduction in passenger bunching at this stop, and thereby improving operational safety. Further, this again demonstrates that it is greatly beneficial to consider dynamic capacity in practical operations. In addition, another interesting observation is that the in-vehicle passengers shown in Fig. 13(a) are reduced compared to the fixed-capacity solution. The main reason is that the *worst-case* perturbation parameter with respect to passenger demand is closely related to both the timetables and dynamic-capacity allocation plans. When using different strategies to optimize these two aspects simultaneously, the *worst-case* perturbation parameter at a specific stop and time interval may be different, since the uncertainty set of the random variable is a bounded polyhedron. Furthermore, according to Proposition 3, the number of in-vehicle passengers varies with the realizations of the *worst-case* passenger demand among different strategies.

5.4. Out-of-sample performance of DRO-TVC and SP-TVC

The investigated robust TT-DCA problem is oriented towards long-term planning at the tactical level. After deriving a data-driven robust timetable and dynamic-capacity allocation plan, its robustness should be evaluated in various scenarios that may appear in

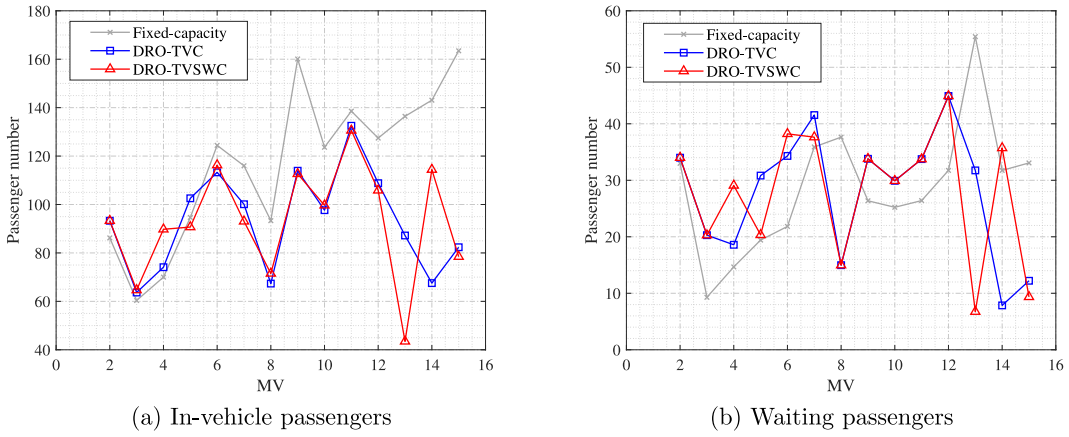


Fig. 13. Comparison of in-vehicle and waiting passengers at stop Kangjiagou among three strategies.

practical operations. Therefore, we now compare the out-of-sample performance of the DRO and SP methods by taking the related models (27) and (58) with the time-varying capacity allocation strategy as an illustration. In this subsection, we adopt the out-of-sample simulation testing procedure proposed by Shehadeh and Tucker (2020). Considering that the degree of disturbance of the external environment on the public transportation system varies at different times in practical operations, we use a time-varying perturbation parameter $\phi(t)$ to vary the parameter $u_{i,t}$ (i.e., the passenger arrival rate) as

$$u'_{i,t} = u_{i,t} \cdot \phi(t), \forall i \in \mathcal{I}, t \in \mathcal{T},$$

where $\phi(t) \sim U[(1-\Pi),(1+\Pi)]$, i.e., $\phi(t)$ obeys a uniform distribution, and Π represents the perturbation of the passenger demand. Thus, new passenger demand scenarios can be conducted to characterize the uncertain and dynamic environments in real-world operations. Specifically, the realization of the passenger arrival rate on a typical day is randomly selected as the benchmark, and then a new dataset including N' out-of-sample scenarios can be constructed by computing $u'_{i,t} = u_{i,t} \cdot \phi(t)$ at each stop and time interval. In our implementation, datasets with different sizes, i.e., $N' \in \{10, 30, 50, 100, 300, 1000\}$, are generated with different perturbations in demand, that is, $\Pi \in \{0.05, 0.10.0.15, 0.20\}$, to evaluate the out-of-sample performance of DRO-OPMV and SP-OPMV among various cases. Then, we solve problems (27) and (58) by using all historical data to obtain the solutions $\hat{\mathbf{y}}_{dro}$ and $\hat{\mathbf{y}}_{so}$, respectively. Finally, the out-of-sample performances $Z_{out}(\hat{\mathbf{y}}_{dro})$ and $Z_{out}(\hat{\mathbf{y}}_{so})$ are computed and compared.

Fig. 14 depicts the median and area between the upper and lower quartiles of the out-of-sample performance. Note that the upper and lower boundary of the shaded area correspond to the upper and lower quartiles of the out-of-sample performance, respectively. Theoretically, the lower the boundary, the better is the model performance. These results suggest that DRO-TVC has better upper and lower quartiles than SP-TVC among all out-of-sample scenarios. In addition, DRO-OPMV has better out-of-sample performance with respect to the median value. For both DRO-TVC and SP-TVC, the performance remains the best when the size of the dataset is small (i.e., $N = 30$ or $N = 50$).

We further compare the out-of-sample performance of the DRO-TVC and the SP-TVC in detail. The comparative results of the statistical indicators are similar under different Π, and therefore, we only present the results of instances when $\Pi = 0.1$. Table 8 illustrates the mean value, standard deviation and median values of the out-of-sample performance for different sample sizes. It can be observed that the average performance of the two methods tends to stabilize with an increase in the sample size, and DRO-TVC outperforms than SP-TVC in all cases. Further, the robustness of the solutions obtained by DRO-TVC and SP-TVC both decreases with an increase in the sample size; however, DRO-TVC still outperforms SP-TVC. Based on the above observations, we conclude that (1) in most cases, the solutions obtained by DRO-TVC are more effective than those obtained by SP-TVC in out-of-sample experiments, and (2) the DRO-TVC model is more robust than the SP-TVC model. These observations demonstrate that DRO-TVC can generate more efficient and reliable solutions than SP-TVC. The superior performance of the DRO-TVC also reflects the value of formulating a data-driven metric-based distributional ambiguity for passenger demand.

5.5. Sensitivity analysis of the Wasserstein radius ρ

Recall that one advantage of the Wasserstein distance-based ambiguity set is its ability to enable decision-makers to control the robustness of solutions by adjusting the Wasserstein radius parameter. Hence, we perform an analysis of the impact of the variation in the parameters on the obtained robust solutions via two sets of experiments. To be specific, we first use statistical methods to analyze a better setting of the parameter ρ, i.e., the radius of the Wasserstein ball, because its value may lead to extensively different out-of-sample performance of DRO-TVC according to Mohajerin Esfahani and Kuhn (2018). The optimal objectives of the DRO, SP, and RO solutions are then compared to demonstrate the superiority of the DRO-TVC model.

There are several methods for estimating the optimal radius of the Wasserstein ball (readers are referred to Mohajerin Esfahani and Kuhn, 2018; Zhou et al., 2020). In this study, we adopted the holdout and k-fold cross-validation approach that can provide

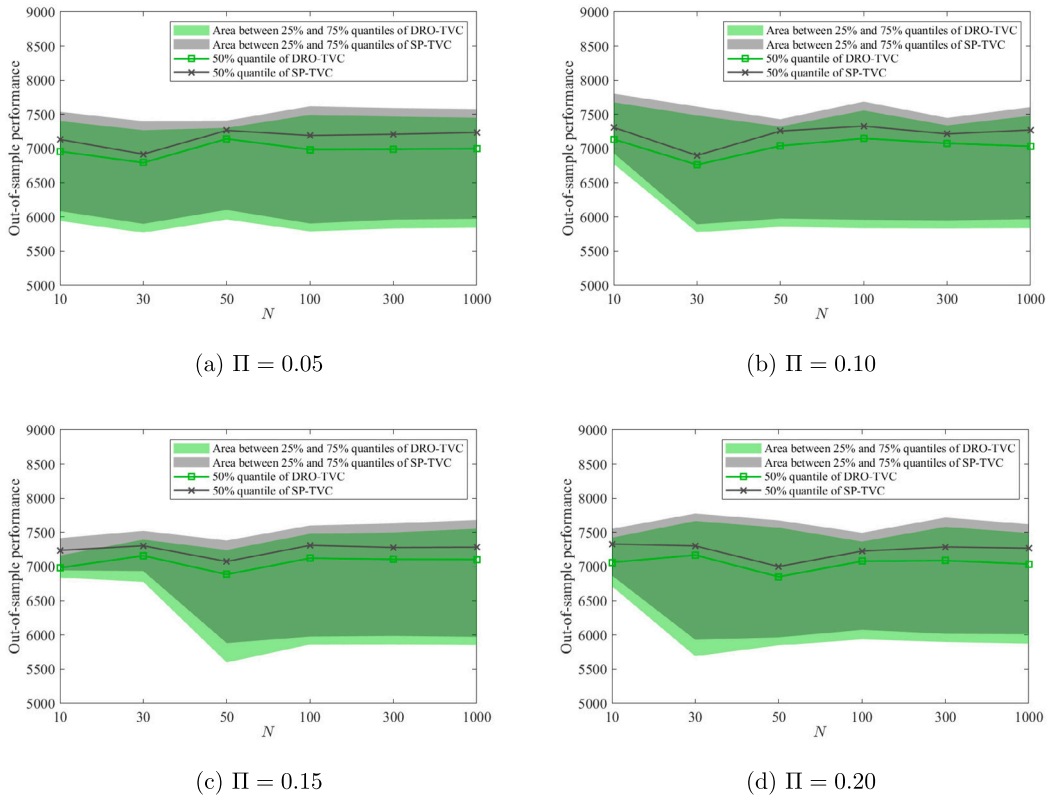


Fig. 14. Out-of-sample performance comparison among various Π of DRO-TVC and SP-TVC.

Table 8
Out-of-sample performance comparison between the DRO-TVC and SP-TVC with $\Pi = 0.1$.

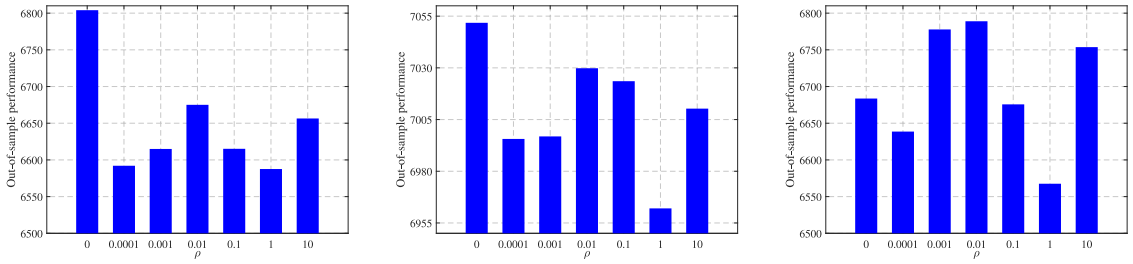
N	Model	Mean value	Standard deviation	Range value
10	DRO-TVC	**7061.724**	862.884	2652.745
	SP-TVC	7248.232	823.311	2560.040
30	DRO-TVC	**6573.188**	**1025.676**	**3773.849**
	SP-TVC	6726.381	1049.201	3862.936
50	DRO-TVC	**6760.523**	**948.894**	**3796.409**
	SP-TVC	6911.596	958.890	3887.237
100	DRO-TVC	**6864.716**	**992.100**	**3838.539**
	SP-TVC	7018.427	1018.084	3917.495
300	DRO-TVC	**6707.349**	**998.745**	**3901.860**
	SP-TVC	6858.051	1008.844	3992.608
1000	DRO-TVC	**6779.612**	**972.498**	**3907.831**
	SP-TVC	6936.462	988.876	3993.457

[1] The bolded metrics indicate that DRO-TVC outperforms SP-TVC.

a more comprehensive analysis and generate a setting of ρ with better out-of-sample performance according to Mohajerin Esfahani and Kuhn (2018). Specifically, first, we denote the historical dataset by $\Theta = \{\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \hat{\boldsymbol{\theta}}_N\}$. The historical dataset is then divided into a testing set Θ_1 of size N_1 and a training set Θ_2 of size $N_0 = N - N_1$; the evaluative index of the out-of-sample performance can be formulated as

$$Z_{out}(\hat{\mathbf{y}}_N(\rho)) = \frac{1}{N_0} \sum_{\hat{\boldsymbol{\theta}}_n \in \Theta_1} Z(\hat{\mathbf{y}}_N(\rho), \hat{\boldsymbol{\theta}}_n), \tag{71}$$

where $\hat{\mathbf{y}}_N(\rho)$ represents the solution obtained using only the training dataset. We provide the pseudocode of the holdout method for DRO-TVC in Algorithm 2 (see Appendix D). The cross-validation method is shown in Algorithm 3 and it is implemented based on Algorithm 2 (see Appendix D). In our implementation, the historical dataset is divided into three subsets, and the candidate value of ρ is selected from the set $P = \{0, 0.0001, 0.001, 0.01, 0.1, 1, 10\}$. Fig. 15 visualizes the out-of-sample performance of the solutions



(a) 1-fold (b) 2-fold (b) 3-fold

Fig. 15. Out-of-sample performance of k-fold cross validation.

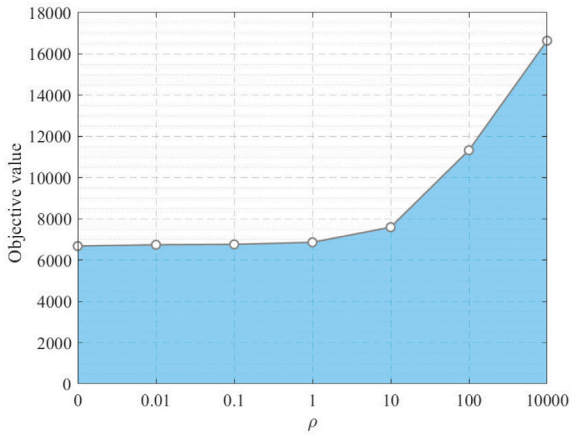


Fig. 16. Computational results under different ρ.

obtained by Algorithm 3 under different values of ρ. Clearly, the DRO-TVC model has the best out-of-sample performance when the value of ρ equals to 1 for the various training and test datasets.

Next, we explore the impact of ρ on the optimal objectives. Fig. 16 shows that the objective value increase with an increase in ρ. This is consistent with the theory because a larger value of ρ indicates that more probability distributions are covered in the Wasserstein ball, which implies that the DRO-TVC is more conservative. Further, as discussed in Remark 4, DRO-TVC (58) with $\rho = 1$ is reduced to SP-TVC (27) when $\rho = 0$; and it is approximately equivalent to the traditional robust optimization model RO-TVC when $\rho = 10000$. As can be observed from Fig. 16, the objective values decreases by 2.71% from $\rho = 1$ to $\rho = 0$. However, the objective value increases by more than 142% as $\rho = 1$ increases to $\rho = 10000$. These results suggest that DRO-TVC can address the challenges of probability distributions that are not precisely available in practice at a smaller cost compared with the stochastic programming method, while obtaining a more reliable solution. Compared with the traditional robust optimization method, DRO-TVC can significantly reduce its over-conservativeness and achieve more efficient solutions. The in-sample and out-of-sample performances of the model are satisfactory when $\rho = 1$.

6. Conclusions and future research

To improve the operational efficiency and reliability of urban public transportation, this paper proposes a data-driven distributionally robust optimization approach to collaboratively optimize the timetable and dynamic-capacity allocation plans of MVs, which takes into account dynamic and uncertain passenger demand, as well as time-dependent travel times. By extracting relevant information from historical data, we construct a Wasserstein distance-based ambiguity set with time-dependent and station-wise perturbation parameter to characterize the uncertain and dynamic operating environment in practice. Subsequently, a data-driven distributionally robust optimization model with time-varying capacity allocation strategy is proposed to minimize the waiting costs of passengers and operating costs of operators. Further, the proposed model is extended to a more general version that enables the time-varying and station-wise capacity. A series of mathematical properties are proven in order to obtain the computationally tractable version of the proposed DRO models. Since real-world instances require a significant number of variables to be described, we designed a customized integer L-shaped algorithm with a set of problem-based valid equalities. Finally, a series of numerical experiments are conducted based on real-world operational data of Beijing Bus Line 468. The computational results indicate the

proposed algorithm can quickly find high-quality solutions of small-scale problems and outperforms the state-of-the-art commercial solver GUROBI in terms of quality and efficiency for solving large-scale problems.

Our computational results demonstrate the practical benefits of enabling dynamic-capacity allocation strategies for reducing operational costs, especially when the time-varying and station-wise capacity of modular vehicles is taken into account. By enabling MUs to be docked/undocked at each stop along the investigated bus line, a good trade-off can be achieved between the costs of passengers and operators, where the number of utilized MUs and the operating costs are significantly reduced by approximately 20% and 46%, compared with the optimized plan with only time-varying capacity. The comparison results with the practical operation strategies considering the static capacity indicate a considerable improvement in the match between supply and demand. Meanwhile, the results obtained show that the in-sample performance of the distributionally robust optimization method is very competitive when compared with the robust optimization approach, whereas its out-of-sample performance outperforms that of the stochastic programming approach.

Future research can be conducted in the following aspects. (1) We constructed a metric-based ambiguity set to describe the uncertainty of passenger demand in this study. Therefore, in the future, other ambiguity sets such as moment-based ambiguity sets can be designed for performance comparison. (2) Storing MUs at stops with limited space and rerouting MUs to the stops where they are needed are two problems that can arise when considering the time-varying and station-wise capacity on a single line. In addition, the number of required MUs may be higher compared to the number of MUs actually used, which can lead to an inaccurate estimation of the corresponding operating costs through the solutions obtained by the proposed approaches on a single line. It would be interesting to explore the problems investigated in this study at the network level, and take passengers' traveling time consisting of waiting time, in-vehicle time and transfer time into account.

CRediT authorship contribution statement

Dongyang Xia: Conceptualization, Methodology, Software, Data curation, Formal analysis, Writing – original draft. **Jihui Ma:** Supervision, Investigation, Validation, Funding acquisition. **Sh. Sharif Azadeh:** Supervision, Conceptualization, Writing – review & editing, Validation. **Wenyi Zhang:** Supervision, Conceptualization, Formal analysis, Funding acquisition, 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 research was supported by the Fundamental Research Funds for the Central Universities (No. 2019JBZ003), the National Natural Science Foundation of China (No. 72288101), and the Fundamental Research Funds for the Central Universities (No. 2022JBMC045). The first author would like to thank the program of China Scholarship Council (No. 202207090003) for the financial support to his visiting Ph.D. research in Delft University of Technology, the Netherlands. The authors are grateful to four anonymous reviewers for their comments that led to a much improved version of this paper.

Appendix A. Proof of Proposition 1

Denote $f(\boldsymbol{\theta}) = \sum_{j\in\mathcal{J}} \theta_j \cdot \pi_j$. Obviously, $f(\boldsymbol{\theta})$ is a convex function. According to Theorem 4.2 proposed in Mohajerin Esfahani and Kuhn (2018), $\sup_{\mathbb{P}\in D_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}\left[\sum_{j\in\mathcal{J}} \theta_j \cdot \pi_j\right]$ equals the optimal value of the following finite convex program:

$$\inf_{\sigma,c,\chi,\varsigma} \quad \rho\sigma + \frac{1}{N}\sum_{n=1}^{N} c_n \tag{72}$$

$$\text{s.t.} \quad [-f] * (\chi_n - \varsigma_n) + g_{\Xi}(\varsigma_n) - \langle \chi_n, \hat{\boldsymbol{\theta}}_n \rangle \le c_n, \qquad \forall n \le N, \tag{73}$$

$$\|\chi_n\|_* \le \sigma, \qquad \forall n \le N, \tag{74}$$

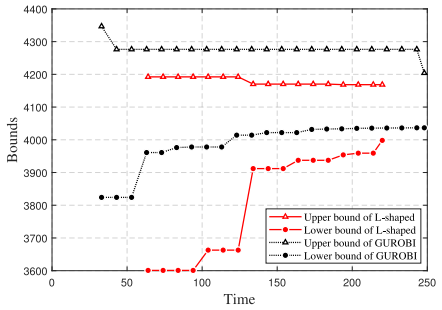
where $[-f] * (\chi_n - \varsigma_n)$ denote the conjugate of $-f$ evaluated at $(\chi_n - \varsigma_n)$, $g_{\Xi}(\varsigma_n)$ is the support function of Ξ, and c_n are the auxiliary variables

According to the definition of the conjugate operator, we can derive

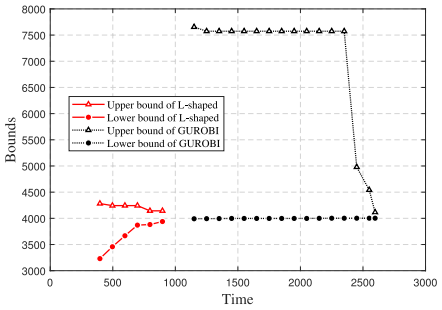
$$[-f] * (\chi_n - \varsigma_n) = \sup_{\boldsymbol{\theta}} \langle \chi_n - \varsigma_n, \boldsymbol{\theta} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\theta} \rangle = \begin{cases} 0, & \text{if } \chi_n - \varsigma_n = -\boldsymbol{\pi}, \\ \infty, & \text{otherwise} \end{cases} \tag{75}$$

Then, based on the definition of the support function, we can derive

$$g_{\Xi}(\varsigma_n) = \begin{cases} \sup\limits_{\boldsymbol{\theta}} \langle \varsigma_n, \boldsymbol{\theta} \rangle \\ \text{s.t. } \mathbf{A}\boldsymbol{\theta} \le \mathbf{b}. \end{cases} = \begin{cases} \inf\limits_{\lambda \ge 0} \langle \lambda_n, \mathbf{b} \rangle \\ \text{s.t. } \mathbf{A}^T \lambda_n = \varsigma_n. \end{cases} \tag{76}$$



(a) Instance with $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (10, 15, 140)$



(b) Instance with $(|\mathcal{I}|, |\mathcal{K}|, |\mathcal{T}|) = (15, 12, 150)$

Fig. 17. Convergence trends of experiments for DRO-TVC with the generated data set.

The left-hand side of constraints (73) can be converted to

$$\begin{aligned}
[-f] * (\chi_n - \varsigma_n) + g_{\Xi}(\varsigma_n) - \langle \chi_n, \hat{\boldsymbol{\theta}}_n \rangle &= \langle \lambda_n, \mathbf{b} \rangle - \langle \varsigma_n - \boldsymbol{\pi}, \hat{\boldsymbol{\theta}}_n \rangle \\
&= \langle \lambda_n, \mathbf{b} \rangle + \langle \boldsymbol{\pi}, \hat{\boldsymbol{\theta}}_n \rangle - \langle \mathbf{A}^T \lambda_n, \hat{\boldsymbol{\theta}}_n \rangle \\
&= \langle \lambda_n, \mathbf{b} \rangle + \langle \boldsymbol{\pi}, \hat{\boldsymbol{\theta}}_n \rangle - \langle \lambda_n, \mathbf{A}\hat{\boldsymbol{\theta}}_n \rangle \\
&= \langle \boldsymbol{\pi}, \hat{\boldsymbol{\theta}}_n \rangle + \langle \lambda_n, \mathbf{b} - \mathbf{A}\hat{\boldsymbol{\theta}}_n \rangle \\
&= \sum_{j=1}^{J} \hat{\theta}_{n,j} \cdot \pi_j + \sum_{d=1}^{D} (b_d - \sum_{j=1}^{J} A_{d,j} \cdot \hat{\theta}_{n,j}) \lambda_{n,d}
\end{aligned}$$

Since $\chi_n = \varsigma_n - \boldsymbol{\pi} = \mathbf{A}^T \lambda_n - \boldsymbol{\pi}$, constraints (74) can be replaced by

$$\left\| \mathbf{A}^{\mathrm{T}} \lambda_n - \boldsymbol{\pi} \right\|_* \leq \sigma, \qquad \forall n \leq N.$$

Appendix B. The integer L-shaped method for DRO-TVSWC

Similarly, we construct the current problem as follows:

$$[\mathrm{CP_S}] \quad \begin{cases} \inf\limits_{\mathbf{r},\mathbf{x},\sigma,\mathbf{c},\lambda,\mathbf{g},\eta} w_1 \cdot \Delta t \cdot \vartheta^T \cdot \Big(\sum\limits_{k\in\mathcal{K}} \sum\limits_{i\in\mathcal{I}} \sum\limits_{t\in\mathcal{T}} \sum\limits_{\varsigma\in\mathcal{T}_t} \upsilon_{k,i,t,\varsigma} \cdot \overline{\mu}_{i,\varsigma} + \rho\sigma + \dfrac{1}{N} \sum\limits_{n=1}^{N} c_n \Big) + \eta \\ \text{s.t.} \quad \eta \geq (U(\mathbf{r}^*, \mathbf{x}) - L) \Big(\sum\limits_{(k,i,t)\in\ H(r^*)} r_{k,i,t} - \sum\limits_{(k,i,t)\notin\ H(r^*)} r_{k,i,t} - \big|H(r^*)\big| \Big) + U(\mathbf{r}^*, \mathbf{x}). \\ \qquad \eta \geq \mathbf{s}(\mathbf{r} - \mathbf{r}^*) + U_{LP}(\mathbf{r}^*, \mathbf{x}). \\ \qquad \text{Constraints (1)–(6), (9)–(13), (24), (35)–(39), (59), (62).} \end{cases}$$

For the sake of simplicity, denote $U_S(\mathbf{r}, \mathbf{x}) = \min_{x_{k,i,q}} \{ \sum_{i\in\mathcal{I}} \sum_{q\in\mathcal{Q}} \sum_{k\in\mathcal{K}} \hat{\vartheta}_q \cdot x_{k,i,q} | \text{Constraints (30)–(31), (59)} \}$. Next, we replace CP and $U(\mathbf{r}, \mathbf{x})$ in Algorithm 1 with CP_S and $U_S(\mathbf{r}, \mathbf{x})$ to obtain the tailored integer L-shaped algorithm for the DRO-TVSWC problem.

Appendix C. Comparison of results between the benchmark and the proposed algorithm based on generated data

We generated a data set to further validate the effectiveness of the proposed algorithm on different data sets. Table 9 provide detailed comparative results for solving the DRO-TVC on the generated data set. Further, Fig. 17 depicts the convergence trends for some selected cases.

Appendix D. *k*-fold cross validation method

By following Mohajerin Esfahani and Kuhn (2018), we first propose the holdout method for DRO-TVC. The detailed procedures are shown as follows

Table 9
Comparison of results between the Benchmark and the proposed algorithm on the generated data set.

Instance[a]	Solution method	Best upper bound	Best lower bound	Relative gap (unit: %)	Computing time (unit: s)
In_5_3_50	Gurobi	365.014	348.599	4.5	1.71
	Integer L-shaped method	358.464	343.465	4.18	4.41
In_10_10_110	Gurobi	2421.878	2340.388	3.36	76.57
	Integer L-shaped method	2428.213	2317.889	4.54	62.71
In_10_12_120	Gurobi	3133.760	3026.326	3.43	93.22
	Integer L-shaped method	3103.182	2963.377	4.51	74.75
In_10_15_140	Gurobi	4204.071	4036.633	3.98	248.00
	Integer L-shaped method	4168.45	3998.195	4.08	220.00
In_15_10_130	Gurobi	3242.484	3093.785	4.59	2227.93
	Integer L-shaped method	3203.693	3043.74	4.99	329.12
In_15_12_150	Gurobi	4114.608	4002.187	2.73	2577.45
	Integer L-shaped method	4143.206	3938.713	4.94	854.32
In_15_15_160	Gurobi	9111.324	4856.397	46.70	3600.00
	Integer L-shaped method	5063.687	4811.038	**4.99**	**1209.76**
In_18_10_150	Gurobi	6655.857	3213.289	51.70	3600.00
	Integer L-shaped method	3332.72	3167.135	**4.97**	**1628.03**
In_18_12_160	Gurobi	8314.553	4151.29	50.10	7200.00
	Integer L-shaped method	4303.769	4105.175	**4.61**	**1644.49**
In_18_15_180	Gurobi	–	–	–	7200.00
	Integer L-shaped method	5247.437	4988.254	**4.94**	**2863.77**

[a]The specific description of each instance is "In_No. Stop_No. Bus_No. time interval", e.g., "In_5_3_50" represents this instance considering 5 stops, 3 buses and 50 time intervals.

Algorithm 2 Holdout method for DRO-TVC.

Step 1: Input testing dataset Θ_1 and training dataset Θ_2 . Create a finite candidate set $P = \{\rho_1, \rho_2, ..., \rho_m, ..., \rho_M\}$. Initial the optimal radius $\rho^{opt} = 0$ and minimum value of out-of sample performance $Z_{out}^{min} = +\infty$, then go to Step 2.
Step 2: If P is not empty, select a ρ_m and remove it from P; go to Step 3. Otherwise, return ρ^{opt} and stop.
Step 3: Run algorithm 1 to solve DRO-TVC by only using the training dataset and get solution $\hat{\mathbf{y}}_N(\rho_m)$, then go to Step 3.
Step 4: Calculate $Z_{out}(\hat{\mathbf{y}}_N(\rho_m))$. If $Z_{out}(\hat{\mathbf{y}}_N(\rho_m)) < Z_{out}^{min}$, let $Z_{out}^{min} = Z_{out}(\hat{\mathbf{y}}_N(\rho_m))$ and let $\rho^{opt} = \rho_m$. Then return to step 2.

Then, k-fold cross validation method is proposed based on the holdout method. The specific procedure of this method is as follows

Algorithm 3 k-fold cross validation.

Step 1: Let $l = 0$. Divide the historical dataset into k subsets $S_1, S_2, .., , S_k$ and go to step 2.
Step 2: If $l \leq k$, let the subset S_l obtained in step 1 as the test dataset, and the remaining subset as the training dataset; go to Step 3. Otherwise, go to Step 4.
Step 3: Run algorithm 2 and get the optimal radius ρ_l of l holdout runs. Let $l = l + 1$, then go to Step 2.
Step 4: Calculate $\rho^{best} = (1/N)\sum_{l \leq k} \rho_l$ and output this value, stop.

References

Agra, A., Rodrigues, F., 2022. Distributionally robust optimization for the berth allocation problem under uncertainty. Transp. Res. B 164, 1–24.
Angulo, G., Ahmed, S., Dey, S., 2016. Improving the integer L-shaped method. INFORMS J. Comput. 28 (3), 483–499.
Ardestani-Jaafari, A., Delage, E., 2016. Robust optimization of sums of piecewise linear functions with application to inventory problems. Oper. Res. 64, 474–494.
Bertsimas, D., Sim, M., 2004. The price of robustness. Oper. Res. 52 (1), 35–53.
Ceder, A., 2016. Public Transit Planning and Operation: Modeling, Practice and Behavior. CRC Press, Boca Raton, FL.
Chen, Z., Li, X., 2021. Designing corridor systems with modular autonomous vehicles enabling station-wise docking: Discrete modeling method. Transp. Res. E 152 (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.
Chen, Z., Li, X., Zhou, X., 2019. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Discrete modeling method. Transp. Res. B 122, 1–19.
Chen, Z., Li, X., Zhou, X., 2020. Operational design for shuttle systems with modular vehicles under oversaturated traffic: Continuous modeling method. Transp. Res. B 132, 76–100.
Cheramin, M., Cheng, J., Jiang, R., Pan, K., 2022. Computationally efficient approximations for distributionally robust optimization under Moment and Wasserstein ambiguity. INFORMS J. Comput. 34 (3), 1768–1794.
Dai, Z., Liu, X.C., Chen, X., Ma, X., 2020. Joint optimization of scheduling and capacity for mixed traffic with autonomous and human-driven buses: A dynamic programming approach. Transp. Res. C 114, 598–619.

Delage, E., Ye, Y., 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. 58 (3), 596–612.
Gkiotsalitis, K., Alesiani, F., 2019. Robust timetable optimization for bus lines subject to resource and regulatory constraints. Transp. Res. E 128, 30–51.
Goerigk, M., Schöbel, A., 2014. Recovery-to-optimality: A new two-stage approach to robustness with an application to aperiodic timetabling. Comput. Oper. Res. 52, 1–15.
Hannoun, G., Menéndez, M., 2022. Modular vehicle technology for emergency medical services. Transp. Res. C 140 (2022), 103694.
Kantorovich, L., Rubinshtein, S., 1958. On a space of totally additive functions. Vestnik St. Petersburg Univ.: Math. 13 (7), 52–59.
Laporte, G., Louveaux, F., 1993. The integer L-shaped method for stochastic integer programs with complete recourse. Oper. Res. Lett. 13 (3), 133–142.
Liu, T., Ceder, A., 2016. Synchronization of public transport timetabling with multiple vehicle types. Transp. Res. Rec. 2539 (1), 84–93.
Liu, K., Li, Q., Zhang, Z., 2019. Distributionally robust optimization of an emergency medical service station location and sizing problem with joint chance constraints. Transp. Res. B 119, 79–101.
Liu, X., Liu, X., Zhang, X., Zhou, Y., Chen, J., Ma, X., 2022. Optimal location planning of electric bus charging stations with integrated photovoltaic and energy storage system. Comput.-Aided Civ. Infrastruct. Eng. http://dx.doi.org/10.1111/mice.12935.
Liu, X., Qu, X., Ma, X., 2021. Optimizing electric bus charging infrastructure considering power matching and seasonality. Transp. Res. D 100, 103057.
Ma, X., Zhang, X., Li, X., Wang, X., Zhao, X., 2019. Impacts of free-floating bikesharing system on public transit ridership. Transp. Res. D 76, 100–110.
Mevissen, M., Ragnoli, E., Yu, J., 2013. Data-driven distributionally robust polynomial optimization. Adv. Neural Inf. Process. Syst. 26, 37–45.
Mohajerin Esfahani, P., Kuhn, D., 2018. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. Math. Program. 171 (1), 115–166.
Newell, G., 1971. Dispatching policies for a transportation route. Transp. Sci. 5 (1), 91–105.
Newell, G., Potts, R., 1964. Maintaining a bus schedule. In: Australian Road Research Board (ARRB) Conference, 2nd. pp. 388–393.
NEXT Future Transportation Inc., 2018. NEXT future transportation - Intro. Accessed May 20, 2023, https://www.youtube.com/watch?v=j6L3DCpGg5w.
NEXT Future Transportation Inc., 2022. NEXT is now. Accessed October 6, 2022, https://www.next-future-mobility.com/.
NEXT Future Transportation Inc., 2023. Dubai experiments the future of transportation, with NEXT. Accessed May 20, 2023, https://www.next-future-mobility.com/post/dubai-with-next-experiments-with-the-future-of-transportation.
Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transp. Res. C 36, 212–230.
Ohmio, 2023. Our vehicles and technology. Accessed June 8, 2023, https://ohmio.com/our-vehicles-and-technology.
Peña, D., Tchernykh, A., Nesmachnow, S., Massobrio, R., Feoktistov, A., Bychkov, I., Radchenko, G., Drozdov, A.Y., Garichev, S.N., 2019. Operating cost and quality of service optimization for multi-vehicle-type timetabling for urban bus systems. J. Parallel Distrib. Comput. 133, 272–285.
Polinder, G., Breugem, T., Dollevoet, T., Maróti, G., 2019. An adjustable robust optimization approach for periodic timetabling. Transp. Res. B 128, 50–68.
Rahimian, H., Mehrotra, S., 2019. Distributionally Robust Optimization: A Review. Technical report, arXiv preprint, arXiv:1908.05659.
Robenek, T., Maknoon, Y., Azadeh, S., Chen, J., Bierlaire, M., 2016. Passenger centric train timetabling problem. Transp. Res. B 89, 107–126.
Sadrani, M., Tirachini, A., Antoniou, C., 2022a. Vehicle dispatching plan for minimizing passenger waiting time in a corridor with buses of different sizes: Model formulation and solution approaches. European J. Oper. Res. 299, 263–282.
Sadrani, M., Tirachini, A., Antoniou, C., 2022b. Optimization of service frequency and vehicle size for automated bus systems with crowding externalities and travel time stochasticity. Transp. Res. C 143, 103793.
Sánchez-Martíneza, G., Koutsopoulosa, H., Wilsona, N., 2016. Real-time holding control for high-frequency transit with dynamics. Transp. Res. B 83, 1–19.
Shehadeh, K., Tucker, E., 2020. A distributionally robust optimization approach for location and inventory prepositioning of disaster relief supplies. arXiv preprint, arXiv:2012.05387.
Shi, X., Li, X., 2021. Operations design of modular vehicles on an oversaturated corridor with first-in, first-out passenger queueing. Transp. Sci. 55 (5), 1187–1205.
Spera, E., 2016. CAREEM to bring driverless transportation solutions to the MENA region through partnership with NEXT future transportation. Accessed June 8, 2023, from https://www.linkedin.com/pulse/careem-bring-driverless-transportation-solutions-mena-emmanuele-spera/.
Tang, C., Ge, Y., Xue, H., Ceder, A., Wang, X., 2023. Optimal selection of vehicle types for an electric bus route with shifting departure times. Int. J. Sustain. Transp. http://dx.doi.org/10.1080/15568318.2022.216107.
Tian, Q., Lin, Y., Wang, D., Liu, Y., 2022. Planning for modular-vehicle transit service system: Model formulation and solution methods. Transp. Res. C 138 (2022), 103627.
Visentini, M., Araújo, O., Borenstein, D., Kummer, A., 2019. Exploiting the timetabling flexibility in the context of the vehicle scheduling problem with heterogeneous fleet. Pesqui. Oper. 39, 85–108.
Wang, Y., Zhang, D., Hu, L., Yang, Y., Lee, L., 2017. A data-driven and optimal bus scheduling model with time-dependent traffic and demand. IEEE Trans. Intell. Transp. Syst. 18 (9), 2443–2452.
Wu, W., Liu, R., Jin, W., Ma, C., 2019. Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. Transp. Res. B 121, 275–303.
Yan, H., Ma, X., Pu, Z., 2022. Learning dynamic and hierarchical traffic spatiotemporal features with transformer. IEEE Trans. Intell. Transp. Syst. 23 (11), 22386–22399.
Zhang, A., Kang, J.E., Axhausen, K., Kwon, C., 2018. Multi-day activity-travel pattern sampling based on single-day data. Transp. Res. C 89, 96–112.
Zhang, Z., Tafreshian, A., Masoud, Neda, 2020. Modular transit: Using autonomy and modularity to improve performance in public transportation. Transp. Res. E 141, 102033.
Zhang, W., Xia, D., Liu, T., Fu, Y., Ma, J., 2021a. Optimization of single-line bus timetables considering time-dependent travel times: A case study of Beijing, China. Comput. Ind. Eng. 158, 107444.
Zhang, Y., Zhang, Z., Lim, A., Sim, M., 2021b. Robust data-driven vehicle routing with time windows. Oper. Res. 69 (2), 469–485.
Zhou, A., Yang, M., Wang, M., Zhang, Y., 2020. A linear programming approximation of distributionally robust chance-constrained dispatch with Wasserstein distance. IEEE Trans. Power Syst. 35 (5), 3366–3377.