An exact method for the integrated optimization of subway lines operation strategies with asymmetric passenger demand and operating costs

Pengli Mo\textsuperscript{a,b}, Andrea D’Ariano\textsuperscript{c}, Lixing Yang\textsuperscript{a}, Lucas P. Veelenturf\textsuperscript{d}, Ziyu Gao\textsuperscript{a}

\textsuperscript{a}State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, 100044 Beijing, China
\textsuperscript{b}School of Industrial Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands
\textsuperscript{c}Department of Engineering, Roma Tre University, via della vasca navale, 79 - 00146 Rome, Italy
\textsuperscript{d}Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, Netherlands

\section*{ARTICLE INFO}

Article history:
Received 29 April 2020
Revised 14 May 2021
Accepted 14 May 2021

Keywords:
Subway train timetabling
Dynamic passenger demand
Rolling stock planning
Operating costs
Energy consumption and regeneration
Forward dynamic programming

\section*{ABSTRACT}

Subway lines connecting different urban functional zones in large cities have direction-dependent and time-variant passenger demand, namely, asymmetry in passenger demand. Most existing studies adopt a symmetric strategy to design operations in both directions and sequentially optimize the different problems associated with operations, thereby failing to meet the asymmetry in passenger demand. This study formulates an asymmetric operation strategy as an integrated mixed-integer non-linear model to optimize the entire operational process of rolling stock from the perspective of service quality and operating costs. Based on the proposed model, an exact algorithm is proposed with speed-up techniques to quickly generate an optimal solution. To this end, the original model is decomposed into several sub-problems that can be exactly solved by using a forward dynamic programming algorithm. Based on actual data from the Beijing subway's Yizhuang line, numerical experiments are conducted to investigate the effectiveness of the asymmetric operation strategy, to identify managerial insights on the integrated optimization, and to evaluate the performance of the proposed methodology.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Motivation

In most modern cities worldwide, urban subway systems have an irreplaceable role in transporting passengers, due to high capacity and reliable operations. Most subway passengers travel to and from their workplace during peak hours in the morning and evening. Under the influence of urban functional zoning, the distribution of residential zones and working zones is spread across different areas in cities. Therefore, passenger demand of subway lines can be asymmetric in different directions during severe time periods \(\text{(i.e., peak hours)}\) during a full day. For instance, we consider four radial lines in the Beijing subway system. As shown in Fig. 1(a), these four lines connect the suburbs with downtown Beijing. It can be clearly seen in Fig. 1(b) that during the morning peak hours, passenger demand toward the downtown area using these four
lines is significantly higher than demand in the direction of the suburban areas, because most working zones in Beijing are mainly located in the downtown area, while residential zones are in the suburbs. Generally, in megacities with a large population and functional zones, passenger demand in different directions of the interzone subway lines presents asymmetric characteristics during peak hours.

In the case of asymmetric passenger demand, executing asymmetric operations (i.e., asymmetric rail plans) across different directions can flexibly guarantee service quality and save operating costs. For example, services can be optimized and scheduled on the direction with more passengers to minimize total passenger waiting time, where each service refers to the task of each train from its start station to its end station. The overall process of planning rail traffic involves multiple procedures and levels (as described, e.g., in Bussieck et al., 1997 and Hansen and Pachl, 2008), leading to complexity when designing asymmetric rail plans. Furthermore, in the traditional approach, the procedures adopted for rail traffic planning are performed sequentially, as shown in Fig. 2. Although each procedure can be easily managed in this way, some problems may still persist if an integrated planning process is not considered. For instance, if we do not consider rolling stock constraints, the optimized train timetable may overestimate the required rolling stock that must be used. Therefore, we aim to investigate a systemwide planning process by considering different procedures required at the tactical level.

To achieve the benefits of system optimization, we treat the optimization of multiple procedures as an integrated optimization problem. As shown in Fig. 2, this study focuses on asymmetric rail plans from the perspective of timetable and rolling stock plans (the parts boxed by solid lines). We further consider some aspects of line planning and optimization of train speed profiles (blue parts).

Note that this study considers the aspect of determining service frequencies (i.e., we only optimize the number of services that serve each direction in the planned time horizon) in line planning, which usually corresponds to the tactical level of subway traffic planning. For descriptive simplicity, we use the operation strategy to represent a combination of the
four component examined in this study (i.e., service frequency planning, timetabling, rolling stock planning, and train speed profile optimization). Since commercial software packages cannot be used to directly solve the problem of integrated optimization of the operation strategy (due to its complexity when dealing with real-world instances), we propose an exact method for obtaining an optimal solution for this problem.

1.2. Literature review

As shown in Fig. 2, train timetabling forms the core bridge of service frequency planning, rolling stock planning, and train speed profile optimization. In this study, the key optimization problem is train timetabling. For this reason, we first review the state-of-the-art techniques used in train timetabling, especially for research on passenger-oriented timetabling and energy-efficient timetabling. We then discuss recent studies that combine train timetabling with the other procedures in the operation strategy. We complete this section with a detailed discussion of the focus of this work.

1.2.1. Train timetabling problem

Train timetabling (also called train scheduling) aims at determining the arrival and departure times of all train services at each station on their respective routes. Early railway operations research can be traced back to the last century (see, e.g., the survey of Cordeau et al., 1998), while train timetabling has also become a popular topic in recent research (see, e.g., the surveys of Cacchiani and Toth, 2012, Caimi et al., 2017). In the literature, formulations for the train timetabling problem are categorized as: periodic event scheduling problem (Serafini and Ukovich, 1989; Zimmermann and Lindner, 2003; Kroon et al., 2015), job shop scheduling problem (Higgins et al., 1996; Zhou and Zhong, 2007; D’Ariano et al., 2007; 2008), and multicommodity flow problem (Caprara et al., 2002; Harrod, 2011; Liu and Dessouky, 2019; Zhang et al., 2019b). Applications of the train timetabling problem are mainly associated with lines and networks in regular railway systems or urban subway systems. Since this study is based on the problem of train timetabling of urban subway lines (referred to as TP-SL), the following descriptions focus on TP-SL aspects such as passenger-oriented timetabling and energy-efficient timetabling.

In passenger-oriented TP-SL, the dynamic (i.e., time-dependent) characteristic of passenger demand is a core aspect of the problem. In the literature, the commonly used pre-determined matrix to address dynamic passenger demand comprises of three dimensions such as origin, destination, and time (ODT matrix). For example, Niu and Zhou (2013), Sun et al. (2014), Barrena et al. (2014a) used the ODT matrix to embed dynamic passenger demand into train timetabling models and designed case studies based on the Guangzhou, Singapore, and Madrid subway systems. These three studies arrive at a similar conclusion whereby under dynamic passenger demand, a timetable with variable headways shows better performance in reducing passenger waiting time than a timetable with fixed headways. However, most works on the TP-SL with variable headways in the existing research still use the same or pre-given service frequency in different directions (Yin et al., 2017; Canca and Zarzo, 2017; Wang et al., 2018; Chen et al., 2019), thus avoiding issues about rolling stock planning and service frequency planning.

Energy-efficient TP-SL has now become a popular topic to protect the environment and to reduce operating costs (see, e.g., the survey of Scheepmaker et al. (2017)). Broadly, the energy-saving TP-SL focuses on two aspects: reducing traction energy and increasing reused energy. The calculation for traction energy consumption relates closely to the execution of train speed profiles. There is a general conclusion that train speed profiles with shorter travel time duration consume more energy (Ghoseiri et al., 2004). Accordingly, traction energy can be considered as a cost function based on train travel times, which makes a passenger travel time trade-off (Canca and Zarzo, 2017). Relatively, energy reuse is an emerging technology that can be briefly described as the braking energy regenerated in trains and used by other trains for traction. For details, refer to Yang et al. (2015). Recent research has been focusing on adjusting timetables to maximize the time overlap between traction and braking phases of train speed profiles (Ramos et al., 2008; Yang et al., 2014) or on calculating reused energy through simulation methods (Su et al., 2014). However, the existing research on the energy-efficient TP-SL ignores the energy savings generated by optimizing the service frequency. In addition, energy consumption is often used as the only indicator of operating costs, which is clearly insufficient, since there are other costs for training operations (Jun and Kim, 2007; Rong et al., 2016), such as maintenance costs.

1.2.2. Integrated optimization problem

Generally, integrated optimization considers multiple problems as a whole. Compared with solving these problems sequentially, integrated optimization often achieves superior performance (Schöbel, 2017). With the development of computing power and theoretical knowledge, research on integrated methods for optimization has emerged in recent years and has been widely used in railway operations management (Gupta et al., 2016; Meng et al., 2018; D’Ariano et al., 2019; Zhang et al., 2019a; Hong et al., 2021). The integrated optimization of train timetabling along with some aspects of service frequency planning and rolling stock planning and the optimization of train speed profiles is also addressed in the following sections.

Line planning aims to determine the number of services and their routes (Schöbel, 2012), which usually belongs to the strategic level. We recall that service frequency planning in this study is a tactical decision, that is, to optimize the number of services for each direction in the planned time horizon. The integrated optimization of service frequencies and train timetables is usually considered in the TP-SL literature, where the routes are predefined (e.g., in Kaspi and Raviv, 2013; Barrena et al., 2014a, Chen et al., 2019).
Rolling stock planning refers to the process of matching rolling stock units with train services, where each rolling stock unit is composed of a number of train carriages that cannot be split in everyday operations. However, the modelling specifications can be quite different under distinctive applications. For example, rolling stock planning in regular railway systems includes train unit assignment (Cacchiani et al., 2010 and Cacchiani et al., 2019) and rolling stock circulation planning (Alfieri et al., 2006; Peeters and Kroon, 2008; Kroon et al., 2015) at the network level. However, the normal situation in many subway systems is that different lines are separated, in this case, rolling stock planning may focus on the effective connection between services on the focused subway line. This closely links the TP-SL to the efficient rolling stock planning process, which generates research topics about the integrated optimization problem of the two procedures (Cadarso and Marín, 2012; Wang et al., 2018; Yue et al., 2017; Mo et al., 2020).

The speed profile of a train determines the speed trajectory during each journey (Howlett et al., 1994; Albrecht et al., 2016; Li et al., 2020), which corresponds to the operation strategy at the microscopic level (Goverde et al., 2016). In recent years, some studies on the integrated optimization of timetables and train speed profiles on a railway line have emerged, which can be categorized into two groups. The first group achieves energy savings by calibrating timetables and optimizing the speed profile of trains. Relevant studies focus on single-train or multi-train operations and are mainly applied in regular railway lines (Wang and Rakha, 2017; Ye and Liu, 2016; Zhou et al., 2017). The second group focuses on the energy-saving TP-SL, and the related studies generate integrated optimization solutions by embedding a set of pre-generated train speed profiles or a train speed profile generation process into the timetabling process (Yin et al., 2017; Canca et al., 2018; Su et al., 2019).

Due to the complexity of the problem, literature on the integrated optimization of three or more procedures is relatively limited. Canca and Zarzo (2017) proposed an optimization model based on cyclic timetabling to optimize service frequency, timetable, rolling stock plan, and train speed profiles simultaneously, with the goal of reducing energy consumption and increasing passenger loading. However, the headway in the cyclic timetable is constant in both directions, which cannot be further adapted to the dynamic passenger demand, especially to the asymmetric passenger demand. Schöbel (2017) proposed an eigenmodel for iteratively optimizing line planning, train timetabling, and rolling stock planning. Accordingly, multiple objectives of these three procedures can be optimized simultaneously through the e-constraint method. Based on the latter work, Fuchs and Corman (2019) extended a method to decrease the complexity of the solution process. However, Schöbel (2017) and Fuchs and Corman (2019) focused on the macroscopic level, without considering the optimization of train speed profiles (at the microscopic level), lacking a detailed description of the operating costs. Based on managerial insights and the practical experience of Dutch railways, Goverde et al. (2016) proposed a three-level timetabling framework to achieve comprehensive optimization of timetables (from both a macroscopic and microscopic perspective) and train speed profiles (at the corridor fine-tuning level). The authors also analyzed the relationship between the various elements (e.g., train travel time and speed) and the objectives (e.g., passengers and energy consumption) of the operation strategy. However, their heuristic framework can only obtain a feasible (rather than optimal) solution and does not directly consider the service frequency optimization.

To address the research gap mentioned above, we propose an innovative research methodology that jointly optimizes service frequency planning, timetabling, rolling stock planning, and train speed profiles for a subway line, with the joint aim of improving passenger service quality and operating costs. Regarding train timetables and service frequency, we aim to optimize the headways and the number of services in the upstream and downstream directions to effectively meet the asymmetric passenger demand. We also intend to reduce operating costs by selecting suitable train speed profiles and optimizing rolling stock planning.

### 1.3. Study focus and contributions

To highlight the focus of our work, some relevant studies are summarized in Table 1 and compared with our study, in terms of objectives (i.e., dynamic passenger demand (D), energy consumption (E)), strategy (symmetric or asymmetric), in-

<table>
<thead>
<tr>
<th>Publications</th>
<th>D</th>
<th>E</th>
<th>Strategy</th>
<th>Integration</th>
<th>Solution Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrena et al. (2014b)</td>
<td>✓</td>
<td>-</td>
<td>symmetric • • •</td>
<td>-</td>
<td>adaptive large neighborhood search algorithm</td>
</tr>
<tr>
<td>Ning et al. (2014)</td>
<td>-</td>
<td>✓</td>
<td>symmetric - • • •</td>
<td>-</td>
<td>integrated control method</td>
</tr>
<tr>
<td>Huang et al. (2016)</td>
<td>✓</td>
<td>✓</td>
<td>symmetric - • • •</td>
<td>-</td>
<td>genetic algorithm</td>
</tr>
<tr>
<td>Canca and Zarzo (2017)</td>
<td>-</td>
<td>✓</td>
<td>symmetric * • • •</td>
<td>-</td>
<td>sequential mixed-integer linear programming approach</td>
</tr>
<tr>
<td>Yin et al. (2017)</td>
<td>✓</td>
<td>✓</td>
<td>symmetric - • • •</td>
<td>-</td>
<td>Lagrangian relaxation based algorithm</td>
</tr>
<tr>
<td>Yue et al. (2017)</td>
<td>✓</td>
<td>-</td>
<td>asymmetric • • •</td>
<td>-</td>
<td>CPLEX, simulated annealing algorithm</td>
</tr>
<tr>
<td>Wang et al. (2018)</td>
<td>✓</td>
<td>✓</td>
<td>symmetric - • • •</td>
<td>-</td>
<td>iterative approach, CPLEX</td>
</tr>
<tr>
<td>Mo et al. (2020)</td>
<td>✓</td>
<td>✓</td>
<td>asymmetric - • • •</td>
<td>-</td>
<td>CPLEX</td>
</tr>
<tr>
<td>Mo et al. (2019b)</td>
<td>✓</td>
<td>✓</td>
<td>asymmetric - • • •</td>
<td>-</td>
<td>modified tabu search</td>
</tr>
<tr>
<td>Yang et al. (2020)</td>
<td>✓</td>
<td>✓</td>
<td>asymmetric - • • •</td>
<td>-</td>
<td>nondominated sorting generic algorithm II</td>
</tr>
<tr>
<td>This study</td>
<td>✓</td>
<td>✓</td>
<td>asymmetric - • • •</td>
<td>-</td>
<td>exact method</td>
</tr>
</tbody>
</table>
tegration (including service frequency (F), timetable (T), rolling stock plan (R), train speed profiles (P)), and specific solution methods.

It follows from Table 1 that most of the existing studies on subway lines consider dynamic passenger demand and energy consumption as objectives, while some of them integrate the TP-SL with other procedures in rail traffic planning. However, recent research on asymmetric operation strategies is quite limited. The work by Yue et al. (2017) is based on a subway loop line, which is essentially different from the operating model of the linear (or radial) subway line. In the research by Yang et al. (2020) and some previous studies (Mo et al., 2020 and Mo et al., 2019b), there is lack of consideration of the studied integration. In addition, the other studies in Table 1 often use heuristic or metaheuristic algorithms, and thus there is lack of research on the exact methods. The main study contributions are highlighted in the following numbered list:

1. We propose a novel methodology that jointly optimizes the following four aspects involved in our version of the TP-SL: service frequency planning, timetabling, rolling stock planning, and train speed profile optimization. An asymmetric operation strategy is considered (i.e., optimizing the four aspects in upstream and downstream directions separately) to effectively meet asymmetric passenger demand. At the same time, rolling stock plans are optimized to ensure the applicability of the asymmetric operation strategy. We formulate the resulting problem as a mixed-integer non-linear programming (MINLP) formulation.

2. This study aims to minimize both passenger waiting and travel times from a service quality perspective and reduce operating costs that comprise costs such as rolling stock utilization, maintenance, and energy. These costs clearly relate to the four TP-SL aspects optimized in this study. Specifically, utilization costs are directly related to the rolling stock planning process, maintenance costs are determined by the service frequency and rolling stock planning, and energy costs are considered in all the four aspects. Furthermore, this study reduces traction energy consumption and increases energy reuse, while optimizing the service frequency.

3. We develop an exact method to solve the proposed MINLP model. First, this method divides the original problem into multiple sub-problems based on the service frequency; subsequently, it accurately solves each sub-problem through a forward dynamic programming algorithm. In addition, we design speed-up techniques, including a pre-processing procedure and two stopping criteria, to end unnecessary calculation processes, thus improving the computational efficiency of our algorithm.

4. The proposed MINLP model and the exact algorithm are evaluated using numerical experiments based on the Beijing subway Yizhuang line data (hereafter referred to as the Yizhuang line). We also present a multi-objective case study, by adjusting the weight coefficients of service quality and operating costs, to generate an approximated Pareto frontier, and to provide more informative solutions to decision makers. Based on different objective preferences, we compare asymmetric operation strategies with symmetric operation strategies to analyze the advantages and limitations of the proposed approach. In addition, we investigate managerial insights by comparing the optimal solutions of the studied integrated problem with the practical operation strategy.

The rest of the paper is organized as follows. Section 2 presents detailed problem statements and the study assumptions. Section 3 describes the MINLP formulation of the integrated optimization problem under asymmetric passenger demand. Section 4 introduces the exact method for generating optimal TP-SL solutions. Section 5 illustrates the performance of various approaches for solving real-world instances of the Beijing subway Yizhuang line. Section 6 provides conclusions on the main findings of our work and recommends future research directions to extend the proposed methodology.

2. Problem statement

2.1. Operation strategy based on asymmetric passenger demand

This study investigates how to best transport an asymmetric passenger demand in a subway line, where there is a significant difference in passenger demand about travel directions. For descriptive convenience, we use (without loss of generality) the upstream direction to represent the side with more passengers. We note that asymmetric passenger demand is a special case of dynamic passenger demand. However, we use the ODT matrix with three dimensions (i.e., origin, destination, and time) to characterize the asymmetric passenger demand.

As shown in Fig. 3, this study considers a bidirectional subway line with 2 depots and 25 stations, denoted by the set $S = \{1, \cdots, 25\}$. The segments are also denoted by the station index (e.g., segment 1 denotes the segment between station 1 and
station 2). The stations are divided into two sets, namely $\mathcal{S} = \{1, \ldots, S\}$ in the upstream direction and $\overline{\mathcal{S}} = \{S + 1, \ldots, 2S\}$ in the downstream direction.

We design operation strategies adapted to the asymmetry in passenger demand during the planned time horizon. We aim to optimize the service frequency, the timetable, and train speed profiles according to the passenger demand for each direction separately and design an efficient rolling stock plan, while ensuring that the operation strategies in different directions can be effectively executed.

In service frequency planning, we determine the number of services in both directions during the planned time horizon. The service in the upstream direction refers to the task of running rolling stock units from station 1 to station $S$ and halting them at each en-route station. In the downstream direction, the service refers to the tasks of running rolling stock units from station $S + 1$ to station $2S$. By considering the planned time horizon and the minimum headways in the upstream and downstream directions, we can obtain the maximum number of services during the planned time horizon (i.e., $\overline{N}_{\text{max}}$ and $\overline{\overline{N}}_{\text{max}}$). We define effective services as those services in which the arrival and departure times at each station are within a time set $\mathcal{T}$. $\mathcal{T}$ is generated by dividing the planned time horizon into a sequence of time intervals with the same duration $\delta$, namely $\mathcal{T} = \{1, \ldots, T\}$, while the timestamp $t \in \mathcal{T}$ is used to represent the time interval $[t\delta, t\delta + \delta)$. The arrival and departure operations of effective services at each station take these timestamp values. Only effective services can be executed by the available rolling stock unit. We use $\overline{N}$ and $\overline{\overline{N}}$ to represent the number of effective services during the planned time horizon in the two directions, where $\overline{N} \leq \overline{N}_{\text{max}}$ and $\overline{\overline{N}} \leq \overline{\overline{N}}_{\text{max}}$. We then define candidate service sets (i.e., $\overline{N}_{\text{max}} = \{1, \ldots, \overline{N}_{\text{max}}\}$ and $\overline{\overline{N}}_{\text{max}} = \{1, \ldots, \overline{\overline{N}}_{\text{max}}\}$) and effective service sets (i.e., $\overline{N} = \{1, \ldots, \overline{N}\}$ and $\overline{\overline{N}} = \{1, \ldots, \overline{\overline{N}}\}$) for the two directions.

In this study, we focus on peak/off-peak based timetables that are widely used in practice, in which the planned time horizon is split into different types of periods, that is, peak hours (p) and off-peak hours (o). The operators pre-determine the order of periods (e.g., an "o-p-o-p-o" order is usually followed during working days). Accordingly, we divide the effective services into peak hour services and off-peak hour services. The duration of each period is variable and is determined by the number of effective services in the corresponding period (e.g., duration is 0 if there are no effective services during this period.) The services offered during peak hours follow the headway $H^{(p)}$, while the services during off-peak hours follow the headway $H^{(o)}$. The $H^{(p)}$ and $H^{(o)}$ are pre-determined according to a network-level subway planning problem (Canca et al., 2016; 2018), which are usually changed either yearly or every half a year. Accordingly, the effective service set $\overline{N}$ in the upstream direction is divided into a service set during peak hours $\overline{N}^{(p)}$ and a service set in off-peak hours $\overline{N}^{(o)}$. Similarly, $\overline{\overline{N}}$ is divided into $\overline{\overline{N}}^{(p)}$ and $\overline{\overline{N}}^{(o)}$.

The selection of train speed profiles is integrated with the determination of train travel times in the timetable. As shown in Fig. 4, each segment (e.g., segment $l$) has a pre-determined set of train speed profiles (i.e., $\mathcal{L}_{l}$), one each containing multiple modes of train speed profiles (i.e., three modes in Fig. 4). Different modes of train speed profiles correspond to different train travel times, and the considered train speed profiles are all applicable in daily operations. We introduce a binary variable $x_{l,i}$ to describe the selection of train speed profiles over segment $i$, where $l \in \mathcal{L}_{l}$. If the train speed profile of mode $l$ is used by segment $i$, $x_{l,i} = 1$; otherwise, $x_{l,i} = 0$. When each $x_{l,i}$ is determined, the arrival and departure times of each service at each station are generated.

The timetables developed in the two directions must match with the rolling stock plans to ensure their applicability. The key decision is about connecting the effective services in different directions, that is, nearby depot 1 and depot 2 to circulate the rolling stock effectively. Therefore, we generate rolling stock plans and timetables simultaneously, such that they can be jointly optimized.

In Fig. 5, we present an example of an asymmetric operation strategy based on the asymmetry in the passenger demand. In terms of service frequency, the number of effective services in both directions must be determined, that is, 12 services are designed in the upstream direction (indicated as red lines) and 9 services are designed in the downstream direction.
(indicated as blue lines). More services are set in the upstream direction to meet a greater passenger demand in this direction. In terms of timetabling, both peak hour services (marked by shades) and off-peak hour services must be determined to generate the headways (we always refer to headway times) between effective services. Peak hours in the direction with more passenger demand (i.e., upstream) should be expanded. At the same time, the travel time in each segment should be selected to determine the optimal train speed profile. For rolling stock planning, the effective services must be matched with the rolling stock units. Specifically, different services must be connected (indicated by the turn-back lines in green), and the minimum fleet size of the required rolling stock units must be calculated based on the rolling stock movements to and from the depots (indicated by yellow dashed lines). In Fig. 5, seven rolling stock units are required to accomplish the timetable (marked by circled numbers).

2.2. Service quality and operating costs

We now examine service quality and operating costs. We consider passenger waiting and travel times to understand service quality. Passenger waiting time is the total time that passengers must wait at each station for the next adjacent service, that is, the time from when each passenger arrives at his/her origin station until the next service departs from that station, which is a common indicator to measure service quality in the passenger-oriented TP-SL (Niu and Zhou, 2013; Yin et al., 2017). In addition, passenger travel time is another key indicator of service quality in subway systems (Lam et al., 1999). Therefore, we combine passenger waiting and travel times as indicators of service quality in this study. For instance, passenger travel time refers to the amount of time each passenger spends in a service. The dwell time of each service at each station is not included in the passenger travel time because the dwell time at each station is not considered as a variable in our optimization problem.

Operating costs include rolling stock utilization costs, maintenance costs, and energy costs. Rolling stock utilization costs refer to fixed costs that are required to use rolling stock units during the planned time horizon, as a surrogate of the costs required for the necessary rolling stock units and related resources, including, e.g., crew costs. A rolling stock plan includes the fleet size of the rolling stock units required for executing the operation strategy. Thus, the utilization costs of the rolling stock can be calculated based on the rolling stock plan. Maintenance costs are the costs of rolling stock maintenance due to the operations, which are associated with the operating mileage of rolling stock and include the costs of executing services and the costs of deadhead operations (i.e., empty rolling stock movements). Therefore, maintenance costs are jointly determined by the service frequency and rolling stock plans. Energy costs are the inevitable costs associated with train operations and can be reduced by energy-saving methods. The considered energy-saving methods include optimizing the number of effective services, selecting train speed profiles to reduce traction energy consumption, adjusting timetables to increase energy reuse, and designing rolling stock plans to reduce empty rolling stock movements. Therefore, energy costs consider all the four procedures involved in planning the operation strategy.

2.3. Study assumptions

The following assumptions are made to formulate asymmetric passenger demand and operation strategies.

**Assumption 1.** The services in each direction run from an origin station to a terminal station, stopping at each intermediate station. Most subway lines use this type of services (also called full-length services).
Table 2
Notation for service frequency, timetable, and train speed profiles.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Detailed Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{S} = {1, \ldots, S} )</td>
<td>Set of involved stations in the upstream direction.</td>
</tr>
<tr>
<td>( \mathcal{S} = {S + 1, \ldots, 2S} )</td>
<td>Set of involved stations in the downstream direction.</td>
</tr>
<tr>
<td>( S )</td>
<td>Set of involved stations, ( S = \mathcal{S} \cup \bar{\mathcal{S}} ).</td>
</tr>
<tr>
<td>( \bar{\mathcal{N}}<em>{\text{max}} = {1, \ldots, \bar{N}</em>{\text{max}}} )</td>
<td>Candidate service set in the upstream direction.</td>
</tr>
<tr>
<td>( \bar{N}<em>{\text{max}} = {1, \ldots, \bar{N}</em>{\text{max}}} )</td>
<td>Candidate service set in the downstream direction.</td>
</tr>
<tr>
<td>( T = {1, \ldots, T} )</td>
<td>Set of discrete times with time unit ( \delta ). For discrete time ( t \in T ), the corresponding time interval is ( [t\delta, t\delta + \delta) ).</td>
</tr>
<tr>
<td>( \mathcal{L}_i = {1, \ldots, \mathcal{L}_i} )</td>
<td>Set of train speed profile modes over segment ( i ) (including ( \mathcal{L}_i ) modes), ( i \in \mathcal{S} \setminus {5, 25} ).</td>
</tr>
</tbody>
</table>

Indexes

| \( k, k', s, s' \) | Index of services, \( k, k' \in \bar{N}_{\text{max}} \cup \bar{N}_{\text{max}} \). |
| \( i, i', j \) | Index of stations and segments, \( i, i', j \in \mathcal{S} \). |
| \( t \) | Index of discrete times, \( t \in T \). |

Parameters

| \( H^{(p)}, H^{(o)} \) | Headways in peak hours and off-peak hours. |
| \( T_{i}^{(p)} \) | Train travel time over segment \( i \) of train speed profile \( l \), \( i \in \mathcal{S} \setminus \{5, 25\} \), \( l \in \mathcal{L}_i \). |
| \( T_{i}^{(o)} \) | Dwell time at station \( i \), \( i \in \mathcal{S} \). |
| \( N^{(p)} \) | Maximum difference between the numbers of effective services in the two directions. |

Decision Variables

| \( \bar{N} = \{1, \ldots, \bar{N}\} \) | Effective service set in the upstream direction. |
| \( \bar{N} = \{1, \ldots, \bar{N}\} \) | Effective service set in the downstream direction. |
| \( \bar{N}_{\text{max}}^{(p)}, \bar{N}_{\text{max}}^{(o)} \) | Set of services in peak hours and off-peak hours in the upstream direction. |
| \( \bar{N}_{\text{max}}^{(o)}, \bar{N}_{\text{max}}^{(o)} \) | Set of services in peak hours and off-peak hours in the downstream direction. |
| \( a_{ik} \) | Integer variable, arrival time of service \( k \) at station \( i \), \( i \in \mathcal{S} \), \( k \in \bar{N}_{\text{max}} \) or \( i \in \bar{S} \), \( k \in \bar{N}_{\text{max}} \). |
| \( d_{ik} \) | Integer variable, departure time of service \( k \) at station \( i \), \( i \in \mathcal{S} \), \( k \in \bar{N}_{\text{max}} \) or \( i \in \bar{S} \), \( k \in \bar{N}_{\text{max}} \). |
| \( x_{il} \) | 0–1 binary variable, \( =1 \) if train speed profile \( l \) is executed over segment \( i \); \( =0 \), otherwise, \( i \in \mathcal{S} \setminus \{5, 25\} \), \( l \in \mathcal{L}_i \). |

Assumption 2. The dwell time at each station is pre-determined. Since subway rolling stock units are usually not allowed to overtake while executing services, dwell time is determined based on the long-term demand to ensure a safe boarding/alighting process for passengers.

Assumption 3. The topological structure of the case considered in this study is a subway line with two depots at both ends. This topology is most suitable for asymmetric operation strategies. However, the model and solution methods proposed in this study can be extended to other topological structures, such as subway lines with one depot (Wang et al., 2018) or subway loop lines (Yue et al., 2017 and Huang et al., 2020). Moreover, we assume that both depots have enough capacity to accommodate the required rolling stock units during rail operations.

Assumption 4. Only the regenerative energy over each segment is considered to maximize its utilization. This assumption is common in real-world applications since the overall energy transmission losses in each segment are very small due to the limited relative distance between two rolling stock units.

Assumption 5. All passengers can be served, even when the subway system reaches its maximum capacity. Therefore, this study does not address passenger control measures (Shi et al., 2019; Liu et al., 2020b) and capacity expansion measures (Chen et al., 2019) to deal with over-saturated passenger demand. Furthermore, train capacity is not considered.

Assumption 6. Passenger demand includes transferring passengers at the transfer stations. A subway system usually estimates passenger paths in the subway network, according to a passenger flow assignment model (e.g., in Shang et al., 2019) in order to obtain the total daily passenger demand served by each subway line.

3. Mathematical model

The asymmetric operation strategy is formulated as an MINLP formulation, with the objective of improving service quality and reducing operating costs. First, we introduce the formulation for the asymmetric operation strategy in Sections 3.1 and 3.2. Second, we introduce the objective function components in Sections 3.3 and 3.4, the objective function in Section 3.5, and the integrated optimization in Section 3.6.

3.1. Formulation of service frequency, timetable, and train speed profiles

A detailed explanation of the notation used in this section is listed in Table 2.

In asymmetric operation strategies, the number of effective services must be determined in the upstream and downstream directions. In other words, the effective service sets \( \bar{N} \) and \( \bar{N} \) must be generated from the candidate sets \( \bar{N}_{\text{max}} \) and \( \bar{N}_{\text{max}} \).
\( \mathcal{N}_{\text{max}} \). Considering the relationship between arrival and departure times of all services at each station, the effective services can be expressed by the following formulas:

\[
k \in \mathcal{N} \iff d_{k,5} \leq T, \quad \forall k \in \mathcal{N}_{\text{max}}
\]

\[
k \in \mathcal{N} \iff d_{k,25} \leq T, \quad \forall k \in \mathcal{N}_{\text{max}}
\]

The maximum difference between the number of effective services in the upstream and downstream directions must be limited to \( \mathcal{N}^{(e)} \). Therefore, the following constraints should be satisfied:

\[
| \mathcal{N} - \mathcal{N}^{(e)} | \leq \mathcal{N}^{(e)}
\]

The effective services are classified into services during peak hours and off-peak hours, respectively. We thus have:

\[
\mathcal{N}^{(p)} \cup \mathcal{N}^{(o)} = \mathcal{N}
\]

\[
\mathcal{N}^{(p)} \cap \mathcal{N}^{(o)} = \emptyset
\]

\[
\mathcal{N}^{(p)} \cup \mathcal{N}^{(o)} = \mathcal{N}
\]

\[
\mathcal{N}^{(p)} \cap \mathcal{N}^{(o)} = \emptyset
\]

We use the form of set operations to formulate constraints (4)-(7), since this form can simplify the modeling of constraints involving effective services. The departure times of effective services from stations 1 and 2S can be determined as follows:

\[
d_{k,1} = \begin{cases} d_{k-1,1} + H^{(p)}, & k \in \mathcal{N}^{(p)}, \quad \forall k \in \mathcal{N} \setminus \{1\} \\ d_{k-1,1} + H^{(o)}, & k \in \mathcal{N}^{(o)}, \quad \forall k \in \mathcal{N} \setminus \{1\} \end{cases}
\]

\[
d_{k,25} = \begin{cases} d_{k-1,25} + H^{(p)}, & \forall k \in \mathcal{N}^{(p)}, \quad \forall k \in \mathcal{N} \setminus \{1\} \\ d_{k-1,25} + H^{(o)}, & \forall k \in \mathcal{N}^{(o)}, \quad \forall k \in \mathcal{N} \setminus \{1\} \end{cases}
\]

Given the dwell time at each station and the train travel time over each segment, both the arrival and departure times of all effective services at each station can be calculated by the following formulas:

\[
da_{k,i} = a_{k,i} + T^{(d)}_{i}, \quad \forall k \in \mathcal{N}, \quad i \in \mathcal{S} \quad \text{or} \quad \forall k \in \mathcal{N}, \quad i \in \mathcal{S},
\]

\[
a_{k,i} = d_{k,i-1} + \sum_{l \in L_{i}} x_{k,l} \cdot T^{(i)}_{l-1,i}, \quad \forall k \in \mathcal{N}, \quad i \in \mathcal{S} \setminus \{1\} \quad \text{or} \quad \forall k \in \mathcal{N}, \quad i \in \mathcal{S} \setminus \{S+1\}.
\]

Only a single mode of train speed profile must be used over each segment, which is formulated as follows:

\[
\sum_{l \in L_{i}} x_{k,l} = 1, \quad \forall i \in \mathcal{S} \setminus \{S, 2S\}.
\]

3.2. Formulation of rolling stock plan

Table 3 reports the notation used in this section.

In this study, rolling stock planning addresses the connection of effective services at depots 1 or 2. As shown in Fig. 6, there are two methods for connecting different services: i) turn-back on the line, and ii) connection through a depot. In the first method, after a rolling stock unit completes the current service, it runs into the turn-back segment to enter the other direction, without returning to the depot. However, this method has strict upper and lower time bounds to ensure safe operation of the rolling stock units. In the second method, after a rolling stock unit finishes the current service, it must move to the depot and run in the other direction. This method requires a longer time than the previous method and consumes more energy, but no upper time bound is needed to start the next service.

We next introduce our rolling stock plan formulation of nearby depot 1. With the binary variables \( \alpha_{k}^{(1)}, \beta_{k}^{(1)}, \gamma_{k,k'}^{(1)}, \) and \( \chi_{k,k'}^{(1)} \), the connection methods between the services in the two directions nearby depot 1 can be formulated uniquely. For service \( k' \) in the upstream direction, the rolling stock execution of this service can be described by binary variables \( \alpha_{k}^{(1)}, \beta_{k}^{(1)}, \gamma_{k,k'}^{(1)} \), and \( \chi_{k,k'}^{(1)} \), as shown in Table 4.

Based on the three cases mentioned in Table 4, we can formulate the required constraints to define all the possible values of binary variables \( \alpha_{k}^{(1)}, \beta_{k}^{(1)}, \gamma_{k,k'}^{(1)} \) and \( \chi_{k,k'}^{(1)} \). In case (1), when service \( k \) in the downstream direction and service
Table 3
Notation of the rolling stock plan.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Detailed Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T^{(1)}$, $T^{(2)}$</td>
<td>Lower and upper time bounds for the turn-back process on the line.</td>
</tr>
<tr>
<td>$T^{(1)}$</td>
<td>Lower time bound for each rolling stock unit to connect two services through depot 1 or depot 2.</td>
</tr>
<tr>
<td>Decision variables $\phi^{(1)}$, $\phi^{(2)}$</td>
<td>Integer variables, fleet size of the required rolling stock units in depot 1 or depot 2.</td>
</tr>
<tr>
<td>$\alpha_k^{(1)}$, $\alpha_k^{(2)}$</td>
<td>0–1 binary variables, =1 if service k is executed by a rolling stock unit departing from depot 1 or depot 2; =0, otherwise.</td>
</tr>
<tr>
<td>$\beta_k^{(1)}$, $\beta_k^{(2)}$</td>
<td>0–1 binary variables, =1 if the rolling stock unit executing service k returns to depot 1 or depot 2; =0, otherwise.</td>
</tr>
<tr>
<td>$\gamma_{k,k'}^{(1)}$, $\gamma_{k,k'}^{(2)}$</td>
<td>0–1 binary variables, =1 if the rolling stock unit executing service k turns back and is connected with service k’ on the line nearby depot 1 or depot 2; =0, otherwise.</td>
</tr>
<tr>
<td>$\chi_{k,k'}^{(1)}$, $\chi_{k,k'}^{(2)}$</td>
<td>0–1 binary variables, =1 if the rolling stock unit executing service k is connected with service k’ through depot 1 or depot 2; =0, otherwise.</td>
</tr>
</tbody>
</table>

Fig. 6. Connection methods of the effective services in the two directions nearby depot 1.

Table 4
The three cases for executing service k’ in the upstream direction.

<table>
<thead>
<tr>
<th>Case</th>
<th>Description of the three cases for executing service k’</th>
<th>$\beta_k^{(1)}$</th>
<th>$\alpha_k^{(1)}$</th>
<th>$\gamma_{k,k'}^{(1)}$</th>
<th>$\chi_{k,k'}^{(1)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Executes service k in the downstream direction, then turns back on the line</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(2)</td>
<td>Executes service k in the downstream direction, then returns to the depot</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(3)</td>
<td>Does not execute any service before executing service k’</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$k'$ in the upstream direction are executed by the same rolling stock unit continuously through turn-back on the line, the following constraints must be satisfied.

\[
\alpha_{k'}^{(1)} = 1 - \sum_{k\in N} \gamma_{k,k'}^{(1)}, \quad \forall k' \in N',
\]  
\[
\beta_k^{(1)} = 1 - \sum_{k'\in N} \gamma_{k,k'}^{(1)}, \quad \forall k \in N,
\]  
\[
a_{k,1} - d_{k,25} \geq T^{(i)} - T(1 - \gamma_{k,k'}^{(1)}), \quad \forall k \in N, k' \in N'
\]  
\[
a_{k,1} - d_{k,25} \leq T^{(i)} + T(1 - \gamma_{k,k'}^{(1)}), \quad \forall k \in N, k' \in N'.
\]

Constraints (13) and (14) ensure that when services k and k’ are connected by turn-back on the line, the rolling stock unit will neither return to nor depart from depot 1 (i.e., $\alpha_{k'}^{(1)} = 0$ and $\beta_k^{(1)} = 0$ when $\gamma_{k,k'}^{(1)} = 1$). Constraints (13) and (14) also ensure that each service can be connected to a maximum of one service before and after this service, respectively. In addition, case (3) can also be ensured by constraints (13) and (14) (i.e., $\gamma_{k,k'}^{(1)} = 0$ when $\alpha_{k'}^{(1)} = 1$). Constraints (15) and (16) guarantee the upper and lower time bounds of the turn-back on the line, where $T$ is the maximum value in the set $\mathcal{T}$ of discrete times.
Remark 3.1. Overtaking is allowed on the turn-back line if its capacity is more than one; otherwise, overtaking is not allowed. For example, two rolling stock units turn back on the line and then execute upstream services \( k' \) and \( s' \) after completing downstream services \( k \) and \( s \), that is, \( y_{k,k'}^{(1)} = y_{s,s'}^{(1)} = 1 \). The departure time of service \( k \) at station 25 is earlier than that of service \( s \) (i.e., \( d_{k,25} < d_{s,25} \)). If the capacity of the turn-back line is only one, it must be enforced that the arrival time of service \( k' \) at station 1 is earlier than that of service \( s' \) (i.e., \( a_{k',1} < a_{s',1} \)). In this study, constraints (16) can ensure this condition provided \( \bar{T}^{(i)} < H^{(p)} \).

In case (2), when service \( k \) in the downstream direction and service \( k' \) in the upstream direction are connected by a rolling stock unit through depot 1, the following constraints must hold:

\[
\sum_{k \in \bar{N}} \chi_{k,k'}^{(1)} \leq \alpha_{k'}^{(1)}, \quad \forall \ k' \in \bar{N},
\]

(17)

\[
\sum_{k' \in \bar{N}} \chi_{k,k'}^{(1)} \leq \beta_{k}^{(1)}, \quad \forall \ k \in \bar{N},
\]

(18)

\[
a_{k',1} - d_{k,25} \geq \bar{T}^{(1)} - T(1 - \chi_{k,k'}^{(1)}), \quad \forall \ k \in \bar{N}, \ k' \in \bar{N},
\]

(19)

Constraints (17) and (18) describe the necessary conditions for any two services to be connected by a rolling stock unit through depot 1, that is, the rolling stock unit executing service \( k \) returns to depot 1, and the rolling stock unit executing service \( k' \) departs from depot 1. In other words, \( \chi_{k,k'}^{(1)} = 0 \) when \( \alpha_{k'}^{(1)} = 0 \) or \( \beta_{k}^{(1)} = 0 \), while both \( \alpha_{k'}^{(1)} = 1 \) and \( \beta_{k}^{(1)} = 1 \) hold when \( \chi_{k,k'}^{(1)} = 1 \). Constraints (19) guarantee the minimum required time for the rolling stock unit to connect two services through depot 1. In short, constraints (13), (14), (17), and (18) guarantee the relationship between the involved binary variables for all three cases in Table 4, while constraints (15), (16), and (19) guarantee the time limits corresponding to cases 1 and 2.

When no rolling stock units can enter the upstream direction to execute service \( k' \) by the above-discussed connection method (i.e., case (3) in Table 4), a new rolling stock unit must be dispatched from depot 1. Furthermore, the number of newly dispatched rolling stock units is equal to the number of rolling stock units required in depot 1, which can be calculated by the following formula:

\[
\phi^{(1)} = \sum_{k \in \bar{N}} \left[ 1 - \sum_{k' \in \bar{N}} (\chi_{k,k'}^{(1)} + y_{k,k'}^{(1)}) \right].
\]

(20)

The rolling stock planning model related to depot 2 can be formulated in a similar way (see Appendix A). We note that a part of the rolling stock planning model (including, constraints (13)-(16) and (41)-(44)) is generalized from previous works (e.g., Wang et al., 2017; Wang et al., 2018; Mo et al., 2020). However, unlike the earlier studies, we model the connection method through the depot and adapt this to the topology of a bidirectional subway line with two depots. In addition, the minimum fleet size of the required rolling stock units is directly formulated as a linear formula.

Remark 3.2. The rolling stock planning model in this study can be directly extended to other topological lines. For example, the proposed model can be extended straight away to manage the subway line with only depot 1, by setting \( \bar{N}^{(e)} = 0 \) and \( \alpha_{k}^{(2)} = 0 \), \( \forall k \in \bar{N} \).

Remark 3.3. For the asymmetric operation strategy, the number of effective services between the two directions can be different (i.e., \( \bar{N} \neq \bar{N}^{(e)} \)), which leads to a change of fleet size in each depot before and after the train operations involved in the planned time horizon. For example, if there are two more effective services in the upstream direction (i.e., \( \bar{N} - \bar{N}^{(e)} = 2 \)), two rolling stock units dispatched from depot 1 can be moved to depot 2. If all the rolling stock units are required to return to their original depots after the current planned time horizon, we can set \( \bar{N}^{(e)} = 0 \) in constraint (3). Without loss of generality, we can replace constraint (3) by \( \bar{N} - \bar{N}^{(e)} = \bar{N}^{(d)} \) to meet the other specific requirements for the rolling stock plan.

3.3. Formulation of service quality

Table 5 presents the notation used to formulate passenger waiting and travel times.

Passenger waiting time refers to the time needed between the arrival of each passenger at his/her origin station and the next service departure from this station. Since two directions are involved in this problem, the total passenger waiting time can be formulated as follows:

\[
PW = \sum_{k \in \bar{N}} \sum_{i \in \bar{S}} \sum_{j \in S} \sum_{t=d_{k,1}}^{d_{k,i}-1} (d_{k,i} - t)P_{i,j,t} + \sum_{k \in \bar{N}} \sum_{i \in \bar{S}} \sum_{j \in S} \sum_{t=d_{k,1}}^{d_{k,i}-1} (d_{k,i} - t)P_{j,i,t}.
\]

(21)
Table 5
Notation for the service quality measurement.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Detailed Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td></td>
</tr>
<tr>
<td>$P_{i,j,t}$</td>
<td>Number of passengers arriving at station $i$ and traveling to station $j$ at time $t$ (i.e., time interval $[t-\delta, t+\delta]$), $i, j \in S, t \in T$.</td>
</tr>
<tr>
<td>$T^{(p)}_{i,t}$</td>
<td>The arrival time of the last passenger at station $i$, $i \in S$.</td>
</tr>
<tr>
<td>$n_i$</td>
<td>Number of passengers traversing segment $i$.</td>
</tr>
<tr>
<td>Objectives</td>
<td></td>
</tr>
<tr>
<td>$PW$</td>
<td>Passenger waiting time at stations.</td>
</tr>
<tr>
<td>$PT$</td>
<td>Passenger travel time during services.</td>
</tr>
</tbody>
</table>

This study assumes that all passengers can be served. Thus, we must ensure that there are no unserved passengers (at any station) after the last service in each direction has been executed. In other words, decision variables $d_{R,i} (i \in S)$ and $d_{N,i} (i \in S)$, the departure times of the last effective services from each in-route station in different directions, should be later than the arrival times of the last passengers at the same station (i.e., $T^{(p)}_{i,t}$). This condition can be enforced by the following equations:

\[ d_{R,i} \geq T^{(p)}_{i,t} \quad \forall i \in S \]  \hspace{1cm} (22)

\[ d_{N,i} \geq T^{(p)}_{i,t} \quad \forall i \in S. \]  \hspace{1cm} (23)

**Remark 3.4.** In order to keep the feasibility of the model, not all passengers in the planned time horizon can be considered as passenger demand. Here, we recommend using the practical timetable to extract the passenger demand. At each station, only those passengers, who arrive between the departure times (from this station) of the first and last (or penultimate) effective services in the practical timetable, can be considered as passenger demand.

Passenger travel time is obtained by multiplying train travel time over each segment by the number of passengers traversing this segment (namely, sectional passengers). This is formulated via the following equation:

\[ PT = \sum_{i \in S \setminus \{S_{25}\}} n_i \left( \sum_{i \in C_i} T^{(p)}_{i,t} x_{i,t} \right). \]  \hspace{1cm} (24)

where $n_i$ can be calculated by counting the number of passengers boarding before station $i$ (including station $i$) and alighting after station $i$. This can be computed as follows:

\[ n_i = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{J}} P_{i,j,t}, \quad \forall i \in S \setminus \{S\}, \]  \hspace{1cm}

\[ n_i = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{J}} P_{i,j,t}, \quad \forall i \in S \setminus \{2S\}. \]

3.4. Formulation of operating costs

Operating costs comprise rolling stock utilization costs, energy costs, and maintenance costs. While calculating energy costs, train speed profiles are considered at the detailed microscopic level. To balance the macroscopic and microscopic levels effectively, we introduce a new set $\mathcal{I}$ of (discrete) times, generated by discretizing the planned time horizon with a time unit $\sigma$. The latter time unit is used to formulate the train speed profile at a microscopic level. Compared with the time unit $\delta$ used above, the new time unit $\sigma$ is much smaller. Furthermore, we assume that $\delta$ is divisible by $\sigma$, namely $\sigma = 1s$ and $\delta = 30s$. With these definitions, $T \subset \mathcal{I}$. These additional notation is listed in Table 6.

For every rolling stock unit under the rolling stock plan, we consider utilization costs $C^{(u)}$. The overall rolling stock utilization costs $CU$ (to satisfy operational requirements) can be defined as follows:

\[ CU = C^{(u)} \cdot (\phi^{(1)} + \phi^{(2)}). \]  \hspace{1cm} (25)

The maximum number ($D$) of available rolling stock units must be limited, which is formulated as follows:

\[ \phi^{(1)} + \phi^{(2)} \leq D. \]  \hspace{1cm} (26)

Maintenance costs $CM$ are closely related to the mileage (or distance) traveled by the rolling stock units. Mileage is computed by considering the executed services and the movements to and from depots (i.e., shunting movements). $CM$ is thus calculated as follows:

\[ CM = C^{(s)} \cdot (\tilde{N} + \tilde{N}) + C_{1}^{(m)} \cdot \left( \sum_{k \in \mathcal{K}} \alpha_{k}^{(1)} + \sum_{k \in \mathcal{K}} \beta_{k}^{(1)} \right) + C_{2}^{(m)} \cdot \left( \sum_{k \in \mathcal{K}} \alpha_{k}^{(2)} + \sum_{k \in \mathcal{K}} \beta_{k}^{(2)} \right). \]  \hspace{1cm} (27)
Table 6  
Notation for the operating costs.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Detailed Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sets</strong></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{T}$</td>
<td>Set of discrete times with time unit $\sigma$, that is, ${1, 2, \ldots, \tilde{T}}$, where $\tilde{T} = T_T$ and $\tau \in \mathcal{T}$. For each discrete time $\tau \in \mathcal{T}$, the corresponding time interval is $[\tau \sigma, \tau \sigma + \sigma)$.</td>
</tr>
<tr>
<td><strong>Indexes</strong></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>Index of discrete time in set $\mathcal{T}$, $\tau \in \mathcal{T}$.</td>
</tr>
<tr>
<td><strong>Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>Maximum number of available $\mathcal{T}$, $\tau \in \mathcal{T}$.</td>
</tr>
<tr>
<td>$C_{\alpha}^{\text{mo}}$</td>
<td>Utilization costs of each rolling stock unit, that is, the utilization price of rolling stock units.</td>
</tr>
<tr>
<td>$E_1$</td>
<td>Energy consumption of each rolling stock unit running from depot 1 to station 1 or from station 2S to depot 1.</td>
</tr>
<tr>
<td>$E_2$</td>
<td>Energy consumption of each rolling stock unit running from depot 2 to station $S + 1$ or from station $S$ to depot 2.</td>
</tr>
<tr>
<td>$\eta$</td>
<td>The conversion efficiency from kinetic energy into regenerated electric energy during braking phases.</td>
</tr>
<tr>
<td>$C_{\text{cost}}$</td>
<td>Energy price, $$$/kWh.</td>
</tr>
<tr>
<td>$C_{\text{cost}}^{\text{mo}}$</td>
<td>Maintenance costs for each service executed by each rolling stock unit.</td>
</tr>
<tr>
<td>$C_{\text{cost}}^{\text{mo}}(k)$</td>
<td>Maintenance costs of each rolling stock unit from depot 1 to station 1 or from station 2S to depot 1.</td>
</tr>
<tr>
<td>$C_{\text{cost}}^{\text{mo}}(k)$</td>
<td>Maintenance costs of each rolling stock unit from depot 2 to station $S - 1$ or from station $S$ to depot 2.</td>
</tr>
<tr>
<td><strong>Variables</strong></td>
<td></td>
</tr>
<tr>
<td>$r_i^{(s)}(\tau)$</td>
<td>Traction electric power over segment $i$ at time $\tau$, $i \in \mathcal{S} \setminus {S, 2S}$, $\tau \in \mathcal{T}$.</td>
</tr>
<tr>
<td>$r_i^{(b)}(\tau)$</td>
<td>Braking electric power over segment $i$ at time $\tau$, $i \in \mathcal{S} \setminus {S, 2S}$, $\tau \in \mathcal{T}$.</td>
</tr>
<tr>
<td>$e_i(\tau)$</td>
<td>Total energy consumption over segments $i$ and $2S - i$ at time $\tau$, $i \in \mathcal{S} \setminus {S}$, $\tau \in \mathcal{T}$.</td>
</tr>
<tr>
<td><strong>Objectives</strong></td>
<td></td>
</tr>
<tr>
<td>$CU$</td>
<td>Rolling stock utilization costs.</td>
</tr>
<tr>
<td>$CE$</td>
<td>Energy costs.</td>
</tr>
<tr>
<td>$CM$</td>
<td>Maintenance costs.</td>
</tr>
</tbody>
</table>

The specific modeling of energy costs $CE$ is as follows:

$$CE = C_{\alpha}^{\text{mo}} \cdot \left[ E_1 \cdot \left( \sum_{k \in \mathcal{N}} \alpha_k^{(1)} + \sum_{k \in \mathcal{N}} \beta_k^{(1)} \right) + E_2 \cdot \left( \sum_{k \in \mathcal{N}} \alpha_k^{(2)} + \sum_{k \in \mathcal{N}} \beta_k^{(2)} \right) + \sum_{i \in \mathcal{S} \setminus \{S\}} \sum_{\tau \in \mathcal{T}} e_i(\tau) \right] .$$  \hspace{1cm} (28)

In equation (28), energy consumption related to the shunting movements in the rolling stock plan is computed as the energy consumed by the rolling stock units running between the depots and their first stations of the line. Since the subway line in this study contains two depots with different geographical locations, we use $E_1$ to represent the energy consumption of each rolling stock unit running between depot 1 and station 1 (or station $2S$) and $E_2$ to represent the energy consumption of each rolling stock unit running between depot 2 and station $S$ (or station $S + 1$). Therefore, energy consumption during the shunting movements is formulated via $E_1$, $E_2$, $\alpha_k^{(1)}$, $\beta_k^{(1)}$, $\alpha_k^{(2)}$, and $\beta_k^{(2)}$. Energy consumption related to the execution of effective services is also formulated by constraints (28), in which $e_i(\tau)$ is the energy consumed over segments $i$ and $2S - i$ at time $\tau$. The formula for $e_i(\tau)$ is as follows:

$$e_i(\tau) = \max \left\{ \left[ r_i^{(s)}(\tau) + r_i^{(b)}(2S - 1, \tau) - r_i^{(b)}(\tau) + r_i^{(s)}(2S - 1, \tau) \right] \cdot \eta \right\} . \hspace{1cm} \forall i \in \mathcal{S} \setminus \{S\}, \tau \in \mathcal{T} \hspace{1cm} (29)$$

Formula (29) denotes the total amount of energy consumed by all rolling stock units over each segment and the opposite segment (e.g., segments 1 and $2S-1$). Here, regenerative energy can be used for traction by other rolling stock units. However, the regenerative energy, which is greater than the amount of energy required for consumption during traction, will not be used and stored in our study. Furthermore, $e_i(\tau) = 0$ when $i = S$, as segments $S$ and $2S$ are turn-back segments. We refer to the previous literature works (e.g., Huang et al., 2017 and Mo et al., 2019a) for calculating the electric power for traction $r_i^{(s)}(\tau)$ and braking $r_i^{(b)}(\tau)$. For clarity, the necessary details of this modeling process are presented in Appendix B.1. The core relationships between the operation strategy and the electric powers can be briefly described by using the following formulas:

$$\{a_{k,i}, d_{k,i}, x_{s,i}, \text{input parameters} | \forall k \in \mathcal{N}, l \in \mathcal{L}_i \} \Rightarrow \{r_i^{(s)}(\tau), r_i^{(b)}(\tau) \mid \forall \tau \in \mathcal{T} \} . \hspace{1cm} \forall i \in \mathcal{S} \setminus \{S\} . \hspace{1cm} (30)$$

$$\{a_{k,i}, d_{k,i}, x_{s,i}, \text{input parameters} | \forall k \in \mathcal{N}, l \in \mathcal{L}_i \} \Rightarrow \{r_i^{(s)}(\tau), r_i^{(b)}(\tau) \mid \forall \tau \in \mathcal{T} \} . \hspace{1cm} \forall i \in \mathcal{S} \setminus \{2S\} . \hspace{1cm} (31)$$
In formulas (30) and (31), the left parts represent the variables and parameters required to calculate the amount of electric power, where \( a_{k,j} \) and \( d_{k,i} \) are the timetable variables, while \( x_{j,i} \) represents the train speed profile selection variable. The input parameters include key information on speed profile (i.e., speed, acceleration, and resistance), rolling stock (i.e., mass of rolling stock unit and energy conversion coefficients), and passenger demand (i.e., the ODT matrix).

3.5. Objective function

As discussed in Section 2.2, we consider performance indicators related to service quality (including passenger waiting time \( PW \) and passenger travel time \( PT \)) and operating costs (including rolling stock utilization costs \( CU \), maintenance costs \( CM \), and energy costs \( CE \)). We then use the following combination of these performance indicators as the objective in our mathematical formulation:

\[
\text{obj} = \min w_p \cdot \frac{PW + PT}{ob_{j,p,\text{nom}}} + w_c \cdot \frac{CU + CM + CE}{ob_{j,c,\text{nom}}}. \tag{32}
\]

In the objective function (32), the sum of three costs (i.e., \( CU, CM, \) and \( CE \)) is used as the operating costs indicator \( \text{obj}_c \) (i.e., \( \text{obj}_c = CU + CM + CE \)), while a weighted sum of \( PW \) and \( PT \) is used as the indicator of service quality \( \text{obj}_p \) (i.e., \( \text{obj}_p = w_p \cdot PW + PT \)), where a parameter \( w_p > 1 \) is used to increase the importance of \( PW \), since passengers usually have a more accurate perception of their waiting time compared to their travel time (as described, e.g., in Fan et al. (2016)). \( w_p \) and \( w_c \) are used to adjust the importance of \( \text{obj}_p \) and \( \text{obj}_c \) according to operators’ preferences. The nominal values \( ob_{j,p,\text{nom}} \) and \( ob_{j,c,\text{nom}} \) are used to normalize \( \text{obj}_p \) and \( \text{obj}_c \). The following procedures can be used to obtain the nominal values. In detail, we first set \( ob_{j,p,\text{nom}} = ob_{j,c,\text{nom}} = 1 \) in the objective function (32). We then set \( w_p >> w_c \) and optimize (32) by using the solution method proposed in Section 4. Thus, a solution with the smallest \( \text{obj}_p \) is generated, where the corresponding \( \text{obj}_p \) is denoted by \( \text{obj}_{p,1} \), and the corresponding \( \text{obj}_c \) is denoted by \( \text{obj}_{c,2} \). Similarly, when \( w_c >> w_p \), a solution with the smallest \( \text{obj}_c \) is generated, where the corresponding \( \text{obj}_c \) (\( \text{obj}_p \)) is denoted by \( \text{obj}_{c,1} \) (\( \text{obj}_{p,2} \)). \( ob_{j,p,\text{nom}} \) and \( ob_{j,c,\text{nom}} \) can also be defined by other methods (e.g., used in Ghoseiri et al. (2004)), according to the preference of operators in the modelling process.

3.6. Integrated optimization model

With the aim of optimizing service quality and operating costs, the integrated optimization problem (IOP) examined in this study can be formulated as the following MINLP:

**Integrated Optimization Problem (IOP):**

\[
\begin{align*}
\text{(32)} & & \text{Weighted sum of } PW, PT, CU, CM, \text{ and } CE \\
\text{subject to} & & \\
\text{Constraints (1) – (12)} & & \text{Service frequency, timetable, and train speed profiles} \\
\text{Constraints (13) – (20), (41) – (48)} & & \text{Rolling stock plan} \\
\text{Constraints (21) – (23)} & & \text{Passenger waiting time } (PW) \\
\text{Constraints (24)} & & \text{Passenger travel time } (PT) \\
\text{Constraints (25) – (26)} & & \text{Rolling stock utilization costs } (CU) \\
\text{Constraints (27)} & & \text{Maintenance costs } (CM) \\
\text{Constraints (28) – (31)} & & \text{Energy costs } (CE)
\end{align*}
\]

4. Solution method

In this section, we first analyze the complexity of the IOP model in terms of the type of variables and constraints and introduce our exact method to solve the IOP.

4.1. Model analysis

The decision variables in the IOP model are related to: (1) service frequency planning (i.e., effective service set \( \mathcal{N} \) in the upstream direction and effective service set \( \mathcal{N} \) in the downstream direction); (2) train timetabling (i.e., arrival and departure times \( a_{k,i} \) and \( d_{k,i} \) of effective service \( k \) at station \( i \)); (3) speed profile selection (i.e., \( x_{j,i} \) over segment \( i \) with mode \( l \)); and (4) rolling stock planning (i.e., binary variables \( \alpha_k^{(1)} \), \( \beta_k^{(1)} \), \( \gamma_{k,k'}^{(1)} \), and \( \chi_{k,k'}^{(1)} \) nearby depot 1 and binary variables \( \alpha_k^{(2)} \), \( \beta_k^{(2)} \), \( \gamma_{k,k'}^{(2)} \), and \( \chi_{k,k'}^{(2)} \) nearby depot 2).

In Appendix C, we provide a detailed complexity analysis of the proposed model. Most of the constraints, except (3), (12), are non-linear and involve more than two categories of decision variables. Specifically, the decision variables related to the service frequency are involved in most of the constraints, except for constraints (12).

Therefore, efficiently handling the coupling relationship between the service frequency variables and other variables is a key factor in the solution process. Furthermore, the constraints between train speed profiles and timetables and those between rolling stock plans and timetables must be carefully addressed when solving the IOP.
4.2. An exact method for the integrated optimization problem

Based on the above model analysis, we next propose an exact method to solve the IOP. To this end, we first handle the relationship between the service frequency variables and the other variables by enumerating all possible schemes, where each scheme represents the number of effective services and their respective departure times from stations 1 and 2S. Given each scheme in the IOP, a corresponding sub-problem, termed scheme optimization problems (SOP), is generated. Thus, the IOP is decomposed into a series of SOPs. To compute the optimal solution of each SOP, we propose a forward dynamic programming (FDP) algorithm. Finally, an optimal solution is obtained for the IOP by analyzing the optimal solutions of the SOPs.

To improve the computational efficiency of the exact method, two types of speed-up techniques are proposed. First, a pre-processing process is implemented to efficiently evaluate and rank all the schemes. Then, two effective stopping criteria are given to terminate the algorithm (with an optimal solution), without the need to solve all the SOPs.

4.2.1. Generation of SOPs

We first explain how the various schemes used to generate SOPs are obtained. Note that each scheme involves both the upstream and downstream directions (i.e., stations 1 and 2S). To generate a scheme, we firstly produce a half-scheme for each direction, and later combine these two half-schemes into a complete one, in which each half-scheme is generated through an enumeration method. For instance, in the upstream direction, we enumerate those services that switch from off-peak periods to peak periods (based on the order of these periods) in the candidate service set and vice versa (i.e., we enumerate services that switch from peak periods to off-peak periods), in order to divide the candidate services into services during peak and off-peak hours, where the departure times of candidate services at station 1 are obtained by constraints (8). Next, we consider the maximum and minimum travel times over segments 1, 2, ···, S − 1 to obtain all possible effective services by constraints (1). Thus, all possible half-schemes in the upstream direction are obtained. Likewise, we combine constraints (2) and (9) to obtain possible half-schemes in the downstream direction (including the effective services in the downstream direction and their departure times at station 2S). By combining the half-schemes in the two directions without violating the constraints (3), we finally obtain the complete set of schemes.

Remark 4.1. The number of generated SOPs can be reduced effectively by pre-determining a part of the timetable information. For example, operators usually pre-determine the departure times of the first and last services from a station or multiple stations (Kang et al., 2015; Guo et al., 2016, and Zhou et al., 2019), especially from the transfer stations and the stations nearby depots, to improve the reachability of a subway network or to ensure specific rolling stock requirements.

Each scheme corresponds to a SOP by formulas (10) - (32). Constraints (1) - (9) (also the decision variables \( \overline{N}, \overline{N}, d_{k,1} \) (\( \forall k \in \overline{N} \)), and \( d_{k,2S} \) (\( \forall k \in \overline{N} \))) in the IOP formulation are not involved in the SOP formulation. Table 7 reports the detailed information about the linear formulas used in each SOP. We note that most of the non-linear constraints in the IOP become linear constraints in the SOP, while the remaining non-linear constraints are mainly used to calculate the passenger waiting time \( PW \) and energy consumption \( CE \).

4.2.2. Forward dynamic programming for solving SOPs

For each SOP, the speed profile of trains in each segment and the rolling stock plan nearby the two depots must be determined. We treat this process as a multi-stage optimization problem. As mentioned in Section 4.2.1, each scheme includes the departure times of effective services from stations 1 and 2S. After determining the train speed profiles over segments 1 and 2S − 1, we can calculate the departure times of effective services at stations 2 and 2S − 1. Similarly, after determining the train speed profiles over segments 2 and 2S − 2, we can calculate the departure times at stations 3 and 2S − 2. By analogy, the departure times at stations S and S + 1 can be calculated after determining the train speed profiles over segments S − 1 and S + 1. Based on the departure times of effective services at stations 1, S, S + 1, and 2S, the rolling stock plan can be optimized, and all the decision variables of the SOP can be determined. In addition, as parameter \( S \) increases, the feasible region of the SOP grows exponentially (see the complexity analysis of the train speed profile variables in Table 15 of Appendix C).

Based on the above reasons, the FDP algorithm, originally proposed by Bellman (1966), is suitable for solving each SOP. The SOP can be divided into \( S \) stages, according to the aforementioned station pair; that is, station pair \( i \) consists of stations

<table>
<thead>
<tr>
<th>Part</th>
<th>Linear formulas</th>
<th>Non-linear formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(32)</td>
<td>(28)</td>
</tr>
<tr>
<td>II</td>
<td>(10)</td>
<td>(21)-(23)</td>
</tr>
<tr>
<td>III</td>
<td>(12)</td>
<td>-</td>
</tr>
<tr>
<td>IV</td>
<td>(13)-(20), (25)-(27), (41)-(42), (45)-(46), (48)</td>
<td>-</td>
</tr>
<tr>
<td>V</td>
<td>(11), (24)</td>
<td>(29)-(31)</td>
</tr>
<tr>
<td>VI</td>
<td>(43)-(44), (47)</td>
<td>-</td>
</tr>
</tbody>
</table>
and $2S - i + 1$ ($i \in \overline{S}$). To obtain the optimal solution of each SOP, we introduce a virtual endpoint after stage $S$, and there are $S + 1$ stages in the FDP. Each stage is only related to its previous stage (if any) and contains the states that are used to record the variables. For stage $i \in \overline{S}$, we use states $s_i$ to record the departure times of the first effective services for the station pair $i$ (e.g., states in stage 1 correspond to $d_{1,1}$ and $d_{1,2S}$). In other words, these two departure times constitute the label for each state. With this labeling method, the departure times of all subsequent effective services at the same stations can be obtained, since the headways between adjacent effective services are known from the scheme information, and all effective services have the same train travel time over the same segment. As shown in Fig. 7, the states of each stage can be represented by the points on a bidimensional plane. The state in stage 1 is fixed by the scheme information, which can be transferred to the next stage through state transition arcs (the dotted lines between two states in Fig. 7). State transition arcs represent the state transition equations that show how the states in the next stage are generated through the states and decisions in the current stage. When the states transit from stage $i$ to $i + 1$ ($i, i + 1 \in \overline{S}$), the corresponding decisions represent the train speed profiles over segments $i$ and $2S - i$ (e.g., state transition arcs from stage 1 to stage 2 correspond to the train speed profiles over segments 1 and $2S - 1$). When the states transit from stage $S$ to stage $S + 1$, the decisions represent the rolling stock plans that are close to the two depots. For each state transition, we introduce a contribution function to calculate its cost, and we use a value function to calculate the optimal cost from the initial state (i.e., the state in stage 1) to the newly generated state according to the state transitions. At the end of the problem solution process, the state transition path with the best cost from the start to the end state (i.e., the virtual endpoint of stage $S + 1$) is the optimal solution of the SOP (the red path in Fig. 7). This solution represents the optimal rolling stock plan for nearby the two depots and the optimal train speed profile in each segment. To summarize, we have discussed the basic items in our FDP algorithm. Next, we introduce the mathematical formulas for these basic items.

- **Stage**

  The process of solving each SOP by our FDP algorithm includes $S+1$ stages. Each stage $i \in \overline{S}$ is associated with a pair of stations (i.e., stations $i$ and $2S - i + 1$), while stage $i = S+1$ corresponds to a virtual endpoint.

- **State**

  We use two elements (i.e., $d_{1,i}$ and $d_{1,2S-i+1}$) to label states $s_i$ for stage $i \in \overline{S}$, where $d_{1,i}$ represents the departure time of the first service at station $i$, while $d_{1,2S-i+1}$ represents the departure time of the first service from station $2S - i + 1$. That is,

  $$s_i = \{d_{1,i}, d_{1,i}^\prime\},$$

  where $i^\prime = 2S - i + 1$. The end state $s_{S+1}$ corresponds to the virtual endpoint (as shown in the illustrative example of Fig. 7), such that all the states in stage $S$ can reach the same end state $s_{S+1}$.

- **Decision**

  We use $u_i$ to represent the decisions for the states in stage $i \in \overline{S}$. The decisions in stage $i \in \overline{S} \setminus \{S\}$ refer to the train speed profile selection over segments $i$ and $2S - i$, while decisions in stage $S$ refer to the rolling stock plan nearby the two depots. Therefore, $u_i$ in stage $i \in \overline{S}$ can be represented as follows:

  $$u_i = \begin{cases} 
  \{x_{1,i}, x_{2S-i+1}^\prime \mid \forall l \in \mathcal{L}_i, l^\prime \in \mathcal{L}_{2S-i}\}, & \text{if } i \in \overline{S} \setminus \{S\} \\
  \{(\alpha_k^{(1)}, \beta_k^{(1)}, \gamma_k^{(1)}, x_{k,k}^{(1)}, \alpha_k^{(2)}, \beta_k^{(2)}, \gamma_k^{(2)}, x_{k,k}^{(2)} \mid \forall k \in \overline{N}, k^\prime \in \overline{N}\}, & \text{if } i = S 
  \end{cases}$$

- **State transition equation**

  The states in each stage are only related to the states and decisions in their previous stage (if any). The state transition equation from stage $i$ to stage $i + 1$ can be denoted as $s_{i+1} = T_i(s_i, u_i)$, where $i, i + 1 \in \overline{S}$. All the states in stage $S$ are reduced.
to the end state in stage $S + 1$. The state transition process from stage 1 to stage $S$ includes the transitions of $d_{1,i}$ and $d_{1,2S-i+1}$, which are established respectively as follows:

$$
\begin{align*}
d_{1,i} &= \begin{cases} 
d_{1,1,} & i = 1 
d_{1,i-1} + T^{(d)}_i + \sum_{l \in \ell_i} X_{i-1,l} \cdot T^{(l)}_{i-1,l}, & i \in \mathcal{S} \setminus \{1\} \end{cases} 
d_{1,2S-i+1} &= \begin{cases} 
d_{1,2S,} & i = 1 
d_{1,i+1} - T^{(d)}_i - \sum_{l \in \ell_i} X_{f,l} \cdot T^{(l)}_{f,l}, & i \in \mathcal{S} \setminus \{1\}, i' = 2S - i + 1 \end{cases}
\end{align*}
$$

The state transition equation from stage 1 to stage $S$ is mainly formulated via constraints (10) and (11). In addition, we also perform the following pruning method during the state transition process. Each newly generated state is recorded and used in the subsequent transitions only when all the constraints (22), (23), and (26) are satisfied. Otherwise, the newly generated state is infeasible and pruned in advance. Moreover, if all the current states cannot transit to the next stage, our FDP algorithm will terminate, since there will be no feasible solution in the corresponding SOP.

**Contribution function**

We use $F_i(s_i, u_i, s_{i+1})$ to represent the contribution function for modelling the transition from state $s_i$ to state $s_{i+1}$ by taking decision $u_i$. The contribution function is formulated by treating the objective function (32) in a multi-stage fashion, as given by the following formula:

$$
F_i(s_i, u_i, s_{i+1}) = \begin{cases} 
w_p \cdot \frac{PW_i}{\text{obj}_{p,nom}} + w_c \cdot \frac{CE_i}{\text{obj}_{c,nom}}, & i \in \mathcal{S} \setminus \{S\} 
w_c \cdot \frac{CU + CM + CE_i}{\text{obj}_{c,nom}}, & i = S
\end{cases}
$$

where $CU$ and $CM$ still refer to formulas (25) and (27), while $PW_i$, $PT_i$, and $CE_i$ are obtained according to linearly separable formulas (21), (24), and (28), $PW_i$ represents passenger waiting time at stations $i$ and $2S - i + 1$, which is based on constraints (21). For $i \in \mathcal{S} \setminus \{S\}$, $PW_i$ is formulated as follows:

$$
PW_i = \sum_{k \in \mathcal{L}_i} \sum_{j=1}^{d_{k,i} - 1} (d_{k,j} - t) \cdot R_{i,j} + \sum_{k \in \mathcal{L}_i} \sum_{j=1}^{d_{k,i} - 1} \sum_{l=1}^{\sum_{e \in \mathcal{L}_i} T^{(l)}_{e,l}} (d_{k,j} - t) \cdot P_{i,j}^{\tau},
$$

where $i' = 2S - i + 1$. For $i \in \mathcal{S} \setminus \{S\}$, $PT_i$ is passenger travel time over segments $i$ and $2S - i$, which is based on constraints (24), and is formulated as follows:

$$
PT_i = \left( \sum_{l \in \ell_i} T^{(l)}_{l,i} \cdot X_{l,i} \right) \cdot n_i + \left( \sum_{l \in \ell_i} T^{(l)}_{f,l} \cdot X_{f,l} \right) \cdot n_f,
$$

where $i' = 2S - i$. For $i \in \mathcal{S}$, $CE_i$ is defined as follows:

$$
CE_i = \begin{cases} 
\sum_{\tau \in \mathcal{X}} e_i(\tau), & i \in \mathcal{S} \setminus \{S\} 
C^{(1)} \left( \sum_{k \in \mathcal{N}} \alpha_k^{(1)} + \sum_{k \in \mathcal{N}} \beta_k^{(1)} \right) + C^{(2)} \left( \sum_{k \in \mathcal{N}} \alpha_k^{(2)} + \sum_{k \in \mathcal{N}} \beta_k^{(2)} \right), & i = S
\end{cases}
$$

(33)

It is worth noting that the notation $e_i(\tau)$, introduced in formula (29) and used in formula (33), motivates us to consider pairs of stations (i.e., stations $i$ and $2S - i + 1$) at each stage.

**Value function**

We define $\Psi^*_i(s_i)$ as the following value function to represent the minimum cost from the start state (i.e., state $s_1$) to state $s_i$:

$$
\Psi^*_{i+1}(s_{i+1}) = \min_{s_i} \{ \Psi^*_i(s_i) + \min_{u_i} F_i(s_i, u_i, s_{i+1}) \}, \forall i \in \mathcal{S},
$$

(34)

and

$$
\Psi^*_1(s_1) = w_p \cdot \frac{PW_1}{\text{obj}_{p,nom}}.
$$

(35)

Considering that each scheme contains the departure times of the effective services from stations 1 and $2S$, $PW_1$ can be obtained only when the corresponding scheme is determined. The process of calculating $\Psi^*_i(s_{i+1})$ in equations (34) corresponds to the process of solving an equivalent shortest path problem. Specifically, the state transition path corresponding to $\Psi^*_i(s_{i+1})$ in equations (34) is the shortest path from state $s_1$ to $s_{i+1}$, which can be obtained by evaluating all the states in stage $i$ that can transfer to state $s_{i+1}$, namely possible states. For any possible state $s_i$, the cost of the shortest path from $s_1$ to $s_{i+1}$ (via $s_i$) is obtained by summing the cost of the shortest path from $s_1$ to $s_i$ (i.e., $\Psi^*(s_i)$) and the best cost of the state transition from $s_i$ to $s_{i+1}$ (i.e., $\min_{u_i} F_i(s_i, u_i, s_{i+1})$). The path with the minimum cost is thus the shortest path from $s_1$ to $s_{i+1}$, and the corresponding cost is $\Psi^*_i(s_{i+1})$. 

299
For \( i \in \mathcal{S} \setminus \{ s \} \), the decision \( u_i \) used to transit from state \( s_i \) to state \( s_{i+1} \) is unique, i.e., \( \min_{u_i} F_i(s_i, u_i, s_{i+1}) = F_i(s_i, u_i, s_{i+1}) \). However, for each state transition from stage \( S \) to stage \( S + 1 \), the possible decisions correspond to \( 2^{\overline{N} + 2\overline{N} + 4\overline{N} \cdot \overline{N}} \) binary variables about the rolling stock planning (as it can be derived from Table 15). We note that linear formulas are used for rolling stock planning (see parts IV and VI in Table 7), and these can be efficiently handled by a state-of-the-art commercial solver, such as CPLEX or GUROBI. Thus, we solve the following sub-problem directly with a solver to obtain the \( \min_{u_i} F_i(s_i, u_i, s_{i+1}) \) for \( i = S \) and the corresponding optimal decision \( u_S \):

\[
\begin{align*}
\min & \quad w_c \cdot \frac{CU + CM + CE_S}{obf_{c,nom}} \\
\text{s.t.} & \quad (13)-(20) \quad \text{Constraints about depot 1} \\
& \quad (41)-(48) \quad \text{Constraints about depot 2} \\
& \quad (25)-(27) \quad \text{Constraints about CU and CM} \\
& \quad \text{Constraints (33)} \quad \text{Energy costs in rolling stock plan CE}_S
\end{align*}
\]

(36)

We observe that \( \Psi_1^*(s_{S+1}) \) is the minimum cost from the start state \( (s_1) \) to the end state \( (s_{S+1}) \), which is also the optimal objective value of the SOP. The optimal SOP solution is deduced by tracing back the shortest path from the end state to the start state.

When optimal solutions are obtained for the SOPs under all the schemes, the optimal IOP solution is computed by comparing the SOP solutions. However, if the number of schemes is large, the process of solving all the SOPs can be intensive. With this concern, we next introduce the speed-up techniques to improve the efficiency of our proposed method.

### 4.2.3. Speed-up techniques: Pre-processing procedure based on scheme ranking

This section describes the pre-processing procedure that is used to accelerate the solution process. The key idea underlying this pre-processing is to set a lower bound value, named scheme potential, for each SOP. We first evaluate all the scheme potentials, and rank all the schemes in ascending order according to their scheme potentials to obtain a scheme list. Following this pre-processing, we use the FDP algorithm introduced in Section 4.2.2 to solve the SOPs, according to the pre-computed scheme list.

Next, we introduce the method adopted in this study to calculate the scheme potential. We first divide the objective function (32) into three parts, namely PPW, PTE, and PME, as defined in equation (37).

\[
\text{obj} = \min w_p \cdot \frac{w_{w_p} \cdot PW}{obj_{p,nom}} + w_p \cdot \frac{PT}{obj_{p,nom}} + w_c \cdot \frac{\sum_{i=1}^{S-1} CE_i}{obf_{c,nom}} + w_c \cdot \frac{CU + CM + CE_S}{obf_{c,nom}}.
\]

(37)

To evaluate the three parts, three corresponding sub-problems are formulated as SP-PPW, SP-PTE, and SP-PME. We use \( \text{PPW}^* \), \( \text{PTE}^* \), and \( \text{PME}^* \) to represent the optimal objective values of these three sub-problems, and introduce PO to represent the scheme potential, which is defined as follows:

\[
\text{PO}=\text{PPW}^*+\text{PTE}^*+\text{PME}^*.
\]

The detailed formulations of SP-PPW, SP-PTE, and SP-PME are introduced below.

**SP-PPW**

\[
\begin{align*}
\min & \quad w_p \cdot \frac{w_{w_p} \cdot PW}{obj_{p,nom}} \\
\text{s.t.} & \quad (10), (21)-(23) \\
& \quad T_{i}^{(x)} \leq d_{1,i} \leq T_{i}^{(x)}, \quad \forall i \in \mathcal{S} \quad \text{The relaxed form}
\end{align*}
\]

(38)

First, we discuss the sub-problem to minimize the PPW. The SP-PPW mainly relates to train timetabling in the SOP (i.e., parts II and V in Table 7), in which we slacken the constraints (11) to release the connection between train arrival and departure times at adjacent stations. Specifically, the departure time of the first service at each station \( i \) is delimited between an upper bound \( T_{i}^{(x)} \) and a lower bound \( T_{i}^{(x)} \) rather than fixing this value based on the departure time at the previous station. \( T_{i}^{(x)} \) and \( T_{i}^{(x)} \) are computed based on the maximum and minimum train travel times to reach station \( i \) and their departure times at station 1 (upstream) or 2S (downstream). With this definition, \( \text{PPW}^* \) is equal to the sum of the minimum passenger waiting time at all stations. An enumeration algorithm can quickly solve the SP-PPW since the feasible region of each station \( i \) is limited by \( T_{i}^{(x)} \) and \( T_{i}^{(x)} \). In addition, it can be observed that the \( \text{PPW}^* \) cannot be greater than the optimal value obtained when minimizing the PPW in the SOP.

We next introduce the sub-problem to minimize the PTE. The formulation for SP-PTE is as follows (after removing redundant constraints, that is, formulations (10), (13)-(23), and (25)-(27), in the SOP).
**SP-PTE**

\[
\begin{align*}
\min & \quad w_p \cdot \frac{PT}{obf_{p,nom}} + w_c \cdot \frac{\sum_{l \in I^c} CE_l}{obf_{c,nom}} \\
\text{s.t.} & \quad (12), (24) \\
& \quad CE_l = \sum_{i \in I^c} x_{i,l} \tilde{r}_{i,l} + \sum_{i \in I^c} x_{i,l} \tilde{r}_{i',l}, \forall i \in S \setminus \{S\}, i' = 2S - i. \\
& \quad \text{Constraints in Part III} \\
& \quad \text{Constraints in Part V} \\
& \quad \text{The relaxed form}
\end{align*}
\]

The SP-PTE is mainly related to the train speed profile selection process in IOP formulation, in which the constraints correspond to parts I and V (see Table 7 for more information). In the SP-PTE, we use a scaling method to replace constraints (28)-(31) with their relaxed form. \(\tilde{r}_{i,l}\) is an underestimated energy consumption of the executing train speed profile \(l\) over segment \(i\), which is lower than the energy consumption calculated according to constraints (28)-(31). Refer to Appendix B.2 for a detailed description of \(\tilde{r}_{i,l}\). Under these conditions, the SP-PTE is a linear model with binary variables \(x_{i,l}\). The SP-PTE can be effectively solved using a commercial solver, such as CPLEX or GUROBI. We observe that PTE* is not greater than the optimal value obtained by minimizing the PTE in the SOP model.

The third sub-problem, that is, the SP-PME, aims to minimize PME and generate a SOP solution. The SP-PME formulation is based on the SOP model and only removes the redundant formulas to calculate the passenger waiting time, passenger travel time, and energy consumption to execute effective services in the SOP (i.e., formulas (21), (24), and (28)-(32)). Thus, optimizing the SP-PME is equivalent to minimizing the PME in the SOP, and a feasible solution of the SP-PME is also feasible for the SOP. The SP-PME is a linear model and can be effectively solved by commercial solvers.

**SP-PME**

\[
\begin{align*}
\min & \quad w_c \cdot \frac{CU + CM + CE_l}{obf_{c,nom}} \\
\text{s.t.} & \quad \text{Constraints (10)-(12), (22)-(23)} \quad \text{Timetable and train speed profiles} \\
& \quad \text{Constraints (13)-(20)} \quad \text{Rolling stock plan nearby depot 1} \\
& \quad \text{Constraints (41)-(48)} \quad \text{Rolling stock plan nearby depot 2} \\
& \quad \text{Constraints (25)-(27)} \quad \text{Rolling stock utilization costs \(CU\) and maintenance costs \(CM\)} \\
& \quad \text{Constraints (33)} \quad \text{Energy costs in rolling stock plan \(CE_l\)}
\end{align*}
\]

**Remark 4.2.** The feasibility of the three sub-problems (i.e., of the SP-PPW, SP-PTE, and SP-PME) can be checked by their corresponding solving methods directly. If not all three sub-problems are feasible, there is no feasible solution in the corresponding SOP. Then, this invalid SOP will no longer be considered during the problem solution process.

### 4.2.4. Speed-up techniques: Stopping criteria

Another aspect to reduce computation time comprises two stopping criteria proposed in this section. The first stopping criteria aim to terminate the solution process of the IOP, while the second stopping criterion aims to terminate the solution processes of the SOPs. Specifically, the core idea of the first stopping criterion is as follows: an optimal IOP solution can be determined using the scheme potentials before solving all the SOPs. We next propose the following theorem to validate the first stopping criterion.

**Theorem 4.1.** Assume that \(q\) schemes are generated and ranked (based on a list of increasing potentials), and the corresponding scheme list is \(Q = \{1, 2, \ldots, q\}\). Let the potential of each scheme \(n \in Q\) be \(PO(n)\), and \(\forall n_1, n_2 \in Q\), if \(n_1 < n_2\), \(PO(n_1) < PO(n_2)\) holds. Let \(UB(n - 1)\) be the minimum objective value in the first \(n - 1\) SOPs. Then, if \(PO(n) > UB(n - 1)\), the optimal objective value of the IOP is \(UB(n - 1)\).

**Proof.** Recall that the PPW, PTE, and PME are defined by the formula (37). Let \(PO^*, PTE^*, \text{ and } PME^*\) be the optimal objective values of the SOP model to minimize PPW, PTE, and PME, respectively. Let \(obj^j\) be the optimal objective value of the SOP. It is easy to show that \(PPW^* + PTE^* + PME^* < obj^j\). As mentioned in Section 4.2.3, we also have \(PPW^* \leq PPW^j\), \(PTE^* \leq PTE^j\), and \(PME^* = PME^j\), where \(PPW^*, PTE^*, \text{ and } PME^*\) are the optimal objective values of the SP-PPW, SP-PTE, and SP-PME, respectively. Furthermore, we have \(PPW^* + PTE^* + PME^* < obj^j\), that is, \(PO \leq \text{obj}^j\). Therefore, \(\forall n \in Q\) and \(n > 1\), if \(PO(n) > UB(n - 1)\), then \(obj^j(n) > UB(n - 1)\). According to the definition of \(UB(n - 1)\), the optimal objective value of the IOP is thus \(UB(n - 1)\). ■

Based on Theorem 4.1, we define the first stopping criterion as follows: we prevent the solution process of the IOP (i.e., solving the SOPs from the first scheme according to the scheme list) if the next scheme potential is higher than the current best-known SOP objective value.

We next introduce the second stopping criterion, which is used during the computation of an optimal SOP solution, that is, during the solution process of our FDP algorithm. We note that optimizing the decisions on state transitions from stage \(S\) to stage \(S + 1\) is time consuming, since each state transition corresponds to a linear integer programming formulation (i.e.,
model (36)). Therefore, we propose the second stopping criterion to conditionally skip these state transitions. The detailed skipping conditions are obtained by the following theorem.

**Theorem 4.2.** $\text{obj}^f$ (the optimal solution of the SOP under each scheme $n \in Q$) is not smaller than $UB(n-1)$ if $\min_{S_5} \Psi^*_S(S_5) + PME^* > UB(n-1)$, where $\min_{S_5} \Psi^*_S(S_5)$ is the minimum cost from the start state to a state in stage $S$ according to the state transition law in the proposed FDP algorithm.

**Proof.** $\min_{S_5} \Psi^*_S(S_5)$ is equivalent to the optimal objective values when minimizing the PPW + PTE in the SOP model. PME* is equal to the optimal objective value when minimizing the PME in the SOP model. The subsequent proof is thus similar to the one of Theorem 4.1. 

Based on Theorem 4.2, we perform the second stopping criterion, before transiting states from stage $S$ to stage $S + 1$ when solving the SOP, as follows: if $\min_{S_5} \Psi^*_S(S_5) + PME^*$ is lower than the current best-known objective value of the SOP, transit the states from stage $S$ to stage $S + 1$ (i.e., continue the solution process of the current SOP); otherwise, stop the solution process of the current SOP and move to the next SOP in the scheme list (if any), since the optimal solution value of the current SOP cannot be lower than the best-known objective value of the SOP.

### 4.2.5. Flowchart of proposed exact algorithm

The flowchart of our proposed exact algorithm is illustrated in Fig. 8. The proposed algorithm is mainly divided into two phases: pre-processing and solving. In the pre-processing phase, we first enumerate all the schemes, with the total number of $q$ schemes (see Section 4.2.1). We then calculate the scheme potential PO for each SOP by solving the three sub-problems (i.e., the SP-PPW, SP-PTE, and SP-PME). When all the schemes are generated, we rank them according to their scheme potentials and insert them in the scheme list in an increasing order of PO.

Similarly, an optimal solution must be generated for each SOP in the solving phase, according to the generated scheme list. For each SOP, if the first stopping criterion of Section 4.2.4 is not satisfied, the FDP algorithm will continue to compute the optimal solution of this SOP (see Section 4.2.2). When the FDP algorithm is executed at stage $S$, the second stopping
criterion is judged. If the second stopping criterion is satisfied, the FDP algorithm skips the remaining calculation process for the current SOP and starts to solve the SOP corresponding to the next scheme in the ordered scheme list; otherwise, the FDP algorithm computes the optimal objective value \( obj^j \) for the current SOP. The FDP algorithm terminates when the first stopping criterion is satisfied or when the SOPs for all the schemes are processed. The FDP algorithm outputs the best known SOP solution, that is, an optimal IOP solution.

5. Numerical experiments

This section presents computational experiments based on the Yizhuang line to illustrate the performance of our modeling approach and solution method. The proposed algorithm has been coded in Visual Studio C++ 2015 on a Windows 10 personal computer with an Intel Core i7-8750H CPU and 16 GB RAM. We use the commercial solver IBM ILOG CPLEX 12.3 to solve the linear programming models of Section 4. For the input data and parameters used in the experiments, please refer to Appendix D.

To evaluate the effectiveness and efficiency of the proposed methodology, we investigated the following four aspects:

1. Nondominated solutions. In this group of experiments, we generate the Pareto frontiers by adjusting the weights of the bi-objective optimization function, by which we can analyze the performance of the asymmetric operation strategy under different objective preferences.

2. Asymmetric versus symmetric operation strategy. This group of experiments compares the performance of asymmetric and symmetric operation strategies from the perspective of the objective preference, primarily to assess the benefits of the optimized asymmetric operation strategy.

3. Practical operation strategy versus optimal operation strategy. We compare our results with the practical operation strategy (that is a symmetric operation strategy). These comparative experiments are useful for extrapolating managerial insights by analyzing where and how the optimal asymmetric operation strategy improves the practical operation strategy.

4. Analysis of algorithm efficiency. Based on a further investigation of the above three groups of experimental results, we quantify the algorithm efficiency, especially regarding the performance of the proposed speed-up techniques.

5.1. Nondominated solutions

To generate the Pareto frontier of the IOP, we first set \( w_p \) as 0.999 and \( w_c \) as 0.001 to generate the nondominated point with \( obj_{p,1} = 903169 \) and \( obj_{c,2} = 7411 \) (i.e., the upper-left point). We then set \( w_p \) as 0.001 and \( w_c \) as 0.999 to generate the nondominated point with \( obj_{p,2} = 1367260 \) and \( obj_{c,1} = 4477 \) (i.e., the lower-right point). Then, \( obj_{p,norm} = 464091 \) and \( obj_{c,norm} = 2934 \). After that, we change \( w_p \) from 0.9 to 0.1 (and correspondingly change \( w_c \) from 0.1 to 0.9) to obtain more nondominated points. The Pareto frontier of the IOP is shown in Fig. 9. The weight ratio \( (w_p : w_c) \) corresponding to each point is marked in Fig. 9. Moreover, without loss of generality, we propose a new problem, named IOP-noCU, in which CU is not optimized in the IOP (this is obtained by multiplying the CU in objective function (32) by a decimal very close to 0), since some subway operators may minimize rolling stock utilization costs in other procedures related to the process of rail traffic planning, such as crew scheduling. The values of \( obj_{p,norm} \) and \( obj_{c,norm} \) related to IOP-noCU are calculated as 462,237 and 25666, respectively. The Pareto frontier of the IOP-noCU is also shown in Fig. 9.

From the Pareto frontier of the IOP in Fig. 9, we find that, in some cases, the optimal values corresponding to different weight ratios may overlap (such as the points corresponding to weight ratios 0.4:0.6 and 0.5:0.5), which means that a set of weight ratios may produce a single nondominated solution. Other informative points in Fig. 9 show the existence of two unexpected jumps. The first jump is from 0.1:0.9 to 0.2:0.8, while the second jump is from 0.5:0.5 to 0.6:0.4. The positions of the two points (before and after each jump) in the figure have a significant distance. The IOP is thus sensitive to these weight ratio intervals, which means that there may be more nondominated points in these intervals. To investigate this, we perform additional experiments to examine the most significant points around the two jumps. These experimental results are shown in Table 8.

In Table 8, columns 2 and 3 report the values of weights \( w_p \) and \( w_c \). The columns from 4 to 8 show the five parts of the objective function. Columns 9 and 10 indicate the number of effective services (i.e., service frequencies) in both the upstream and downstream directions. Aggregate indicators on the train speed profiles are reported in columns 11 and 12. \( T \) in column 11 represents the sum of the selected train speed profile modes over all the other segments in the upstream direction, that is, \( \bar{T} = \sum_{i \in \mathbb{N}_s \backslash \{S\}} \sum_{l \in L_i} x_{i,l} \cdot l \). Similarly, \( \bar{T} \) in column 12 represents the sum of train speed profile modes for the downstream direction.

From the results in Table 8, the first jump “0.1:0.9→0.2:0.8” is reduced to “0.18:0.82→0.2:0.8”. We marked in bold the indicators that change significantly in cases 6 and 7. Compared with case 6, the values of \( \bar{T} \) and \( \bar{T} \) in case 7 are obviously reduced. More train speed profiles with shorter travel times are used in case 7, leading to a notable reduction (increase) of PT (CE). Notably, CU is also reduced, that is, the number of required rolling stock units to match all the effective services is reduced, since the shorter train travel times accelerate the circulation of the available rolling stock unit. A similar conclusion is also supported by the experimental results in Canca and Zarzo (2017). The second jump “0.5:0.5→0.6:0.4” is reduced to “0.58:0.42→0.6:0.4”. The most relevant indicators about the second jump are marked in bold for cases 14 and 15, where the most significant difference is related to the number of effective services in both directions. When comparing the latter two,
PW is reduced considerably, while the corresponding CM, CU, and CE are evidently increased. In Table 8, the most significant difference between the two mentioned jumps is the change in CU, implying that a small change in the fleet size can lead to an obvious change in the operating costs.

Similar to the Pareto frontier of the IOP, there are two jumps (one is from 0.4:0.6 to 0.5:0.5, and the other is from 0.5:0.5 to 0.6:0.4) on the Pareto frontier of the IOP-noCU, as shown in Fig. 9. Furthermore, we find eight new nondominated solutions (indicated via the “Δ” in Fig. 9) without weight ratios from the additional cases tested between these two jumps. The Pareto frontier of the IOP-noCU is thus much smoother than the Pareto frontier of IOP. We report the complete experimental results of the IOP-noCU in Table 9, where we replace the CU column of Table 8 with a new column about the fleet size, named “Fleet.”

In Table 9, as \( w_c \) increases, the number of effective services (\( \hat{N} \) and \( \hat{N} \)) gradually decrease and the sum of train speed profile modes (\( \hat{T} \) and \( \hat{T} \)) gradually increases, thus reducing the operating costs (CM and CE). However, the effective services

and train speed profiles may not always change simultaneously, since increasing train travel times can influence the cost of reducing an effective service, while reducing the effective services can influence the unit cost of increasing the train travel time.

In this section, we present two sets of experiments to generate nondominated solutions for the IOP and IOP-noCU. For the IOP, we observe that the nondominated solutions (locally) aggregate around several centers, where each center corresponds to a fleet size case. When comparing the results obtained for the IOP and IOP-noCU, the Pareto frontier of the IOP-noCU is smoother. Furthermore, we find that the service frequency and train speed profiles show a close relationship with each other in the experimental results of the IOP-noCP. Moreover, we notice asymmetric features (i.e., $\bar{N} \neq \check{N}$ and $\bar{T} \neq \check{T}$) in nondominated solutions of the IOP-noCU, which are examined in Section 5.2.

### 5.2. Asymmetric strategy versus symmetric strategy

In Table 9, the experimental cases 7–15 (corresponding to weight ratios from 0.44:0.56 to 0.6:0.4) represent asymmetric strategies. Specifically, these asymmetric strategies have similarities, namely, $\bar{N} \geq \check{N}$ and $\bar{T} \leq \check{T}$. In other words, the number of effective services in the upstream direction surpass those in the downstream direction, and train travel time in the upstream direction is shorter than that in the downstream direction, which complies the asymmetric passenger demand in the morning peak hours on the Yizhuang line. As passenger demand in the upstream direction is much higher, more effective services are assigned asymmetrically to reduce passenger waiting time, while shorter train travel times are used to reduce passenger travel time.

To further investigate if asymmetric strategies perform better than symmetric strategies, we execute an additional set of experiments for the IOP-noCU, where the difference between the number of effective services in the two directions (i.e., $\bar{N}^{(e)}$ in constraints (3)) is equal to 0. For clarity, we use IOP-noCU$^5$ to show the experimental results with $\bar{N}^{(e)} = 5$, while we use IOP-noCU$^0$ to represent the additional experimental results with $\bar{N}^{(e)} = 0$. The comparison results are shown in Table 10, where we mark the better objective values, service quality indicators, and operating costs indicators in bold, showing that asymmetric strategies (IOP-noCU$^5$) are better than symmetric strategies (IOP-noCU$^0$) in all cases. Furthermore, it is very in-

### Table 9

<table>
<thead>
<tr>
<th>Case</th>
<th>$w_p$</th>
<th>$w_c$</th>
<th>PW($\delta$)</th>
<th>PT($\delta$)</th>
<th>CM($\psi$)</th>
<th>Fleet</th>
<th>CE($\psi$)</th>
<th>$\bar{N}$</th>
<th>$\check{N}$</th>
<th>$\bar{T}$</th>
<th>$\check{T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.999</td>
<td>414,567</td>
<td>743,555</td>
<td>1085.76</td>
<td>6</td>
<td>1183.22</td>
<td>25</td>
<td>25</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.9</td>
<td>410,402</td>
<td>734,713</td>
<td>1085.76</td>
<td>6</td>
<td>1183.45</td>
<td>25</td>
<td>25</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.8</td>
<td>410,457</td>
<td>711,617</td>
<td>1085.76</td>
<td>6</td>
<td>1210.71</td>
<td>25</td>
<td>25</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.7</td>
<td>410,131</td>
<td>619,349</td>
<td>1085.76</td>
<td>6</td>
<td>1393.90</td>
<td>25</td>
<td>25</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>0.4</td>
<td>0.6</td>
<td>410,110</td>
<td>582,508</td>
<td>1085.76</td>
<td>6</td>
<td>1497.73</td>
<td>25</td>
<td>25</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>0.42</td>
<td>0.58</td>
<td>410,611</td>
<td>541,891</td>
<td>1085.76</td>
<td>6</td>
<td>1648.24</td>
<td>25</td>
<td>25</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>0.56</td>
<td>382,327</td>
<td>541,891</td>
<td>1163.34</td>
<td>9</td>
<td>1758.04</td>
<td>28</td>
<td>25</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>0.46</td>
<td>0.54</td>
<td>368,563</td>
<td>541,891</td>
<td>1205.64</td>
<td>9</td>
<td>1814.66</td>
<td>29</td>
<td>26</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>0.48</td>
<td>0.52</td>
<td>350,959</td>
<td>541,891</td>
<td>1264.38</td>
<td>8</td>
<td>1891.27</td>
<td>30</td>
<td>28</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
<td>0.5</td>
<td>326,095</td>
<td>541,891</td>
<td>1348.98</td>
<td>8</td>
<td>2004.52</td>
<td>32</td>
<td>30</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>0.52</td>
<td>0.48</td>
<td>314,767</td>
<td>541,891</td>
<td>1391.28</td>
<td>8</td>
<td>2061.14</td>
<td>33</td>
<td>31</td>
<td>21</td>
<td>23</td>
</tr>
<tr>
<td>12</td>
<td>0.54</td>
<td>0.46</td>
<td>304,513</td>
<td>530,423</td>
<td>1433.58</td>
<td>8</td>
<td>2191.13</td>
<td>34</td>
<td>32</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>13</td>
<td>0.56</td>
<td>0.44</td>
<td>295,080</td>
<td>522,316</td>
<td>1475.88</td>
<td>8</td>
<td>2301.63</td>
<td>35</td>
<td>33</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>14</td>
<td>0.58</td>
<td>0.42</td>
<td>291,600</td>
<td>522,316</td>
<td>1500.18</td>
<td>8</td>
<td>2329.05</td>
<td>35</td>
<td>34</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>15</td>
<td>0.6</td>
<td>0.4</td>
<td>290,472</td>
<td>522,316</td>
<td>1500.18</td>
<td>8</td>
<td>2329.06</td>
<td>35</td>
<td>34</td>
<td>19</td>
<td>23</td>
</tr>
<tr>
<td>16</td>
<td>0.7</td>
<td>0.3</td>
<td>289,616</td>
<td>498,501</td>
<td>1518.18</td>
<td>8</td>
<td>2567.83</td>
<td>35</td>
<td>35</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>17</td>
<td>0.8</td>
<td>0.2</td>
<td>282,378</td>
<td>490,685</td>
<td>1560.48</td>
<td>8</td>
<td>2799.69</td>
<td>36</td>
<td>36</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>18</td>
<td>0.9</td>
<td>0.1</td>
<td>282,645</td>
<td>482,829</td>
<td>1560.48</td>
<td>8</td>
<td>3012.06</td>
<td>36</td>
<td>36</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>19</td>
<td>0.999</td>
<td>0.001</td>
<td>283,228</td>
<td>478,237</td>
<td>1576.92</td>
<td>7</td>
<td>3257.88</td>
<td>37</td>
<td>36</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

### Table 10

<table>
<thead>
<tr>
<th>Case</th>
<th>IOP-noCU$^0$</th>
<th>IOP-noCU$^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$ab_j(\delta)$</td>
<td>$ab_j(\psi)$</td>
</tr>
<tr>
<td>7</td>
<td>1,157,952</td>
<td>2740.56</td>
</tr>
<tr>
<td>8</td>
<td>1,072,236</td>
<td>3151.46</td>
</tr>
<tr>
<td>9</td>
<td>1,051,122</td>
<td>3259.31</td>
</tr>
<tr>
<td>10</td>
<td>1,034,022</td>
<td>3349.42</td>
</tr>
<tr>
<td>11</td>
<td>1,015,734</td>
<td>3457.09</td>
</tr>
<tr>
<td>12</td>
<td>988,165</td>
<td>3627.79</td>
</tr>
<tr>
<td>13</td>
<td>965,944</td>
<td>3779.01</td>
</tr>
<tr>
<td>14</td>
<td>955,576</td>
<td>3872.29</td>
</tr>
<tr>
<td>15</td>
<td>955,576</td>
<td>3872.29</td>
</tr>
</tbody>
</table>
interesting to note that the train speed profiles corresponding to the optimal solutions in both IOP-noCU⁰ and IOP-noCU⁵ are the same (not just $T$ and $\bar{T}$). Therefore, a comparison between asymmetric and symmetric strategies focuses on the difference in effective services. In case 7, the asymmetric strategy increases effective services in the upstream direction, thereby reducing the service quality costs (i.e., passenger waiting time). In cases 8, 9, 14, and 15, asymmetric strategies reduce effective services in the downstream direction, which can reduce the operating costs (i.e., maintenance costs). In cases 10–13, asymmetric strategies simultaneously reduce effective services in the downstream direction and increase effective services in the upstream direction, which reduces service quality costs (i.e., passenger waiting time) in all four cases. The inherent reason for these cost reductions is because asymmetric strategies can better address the asymmetry in passenger demand by flexibly adjusting the effective services in the upstream and downstream directions. Compared to symmetric strategies, asymmetric strategies have a better performance overall in balancing both service quality and operating costs.

### 5.3. Practical operation strategy versus optimal operation strategy

We next compare the optimal solutions obtained applying our solution method with the practical operation strategy of the Yizhuang line in 2015. This practical operation strategy was manually designed by the personnel of operations management based on passenger data and their experiences. First, service frequency, timetable, and rolling stock planning are designed sequentially, and then fine-tuned without careful consideration of how to select optimal train speed profiles. In addition, the planned time horizon studied in the numerical experiments is from 6:00 am to 10:00 am, which comes under the morning peak hours of the Yizhuang line. During this period, all services maintain a fixed headway of six minutes.

We show the optimization results on the IOP and IOP-noCU and compare them with the practical operation strategy. Both service quality costs and operating costs of the practical operation strategy are represented by the red dots in Fig. 9. The two red dots in this figure represent operating costs, which have two different forms depending on whether the rolling stock utilization costs are included.

To examine how optimal solutions can save operating costs, we select the optimal solution with service quality costs similar to (but not higher than) the practical operation strategy. The selected cases for the IOP and IOP-noCU are shown in Fig. 9, marked by IOP⁰ and IOP-noCU⁰. Similarly, to analyze how service quality can be improved we choose optimal solutions with operating costs similar to (but not higher than) the practical operation strategy. The selected cases are marked with IOP⁰ and IOP-noCU⁰.

Table 11 compares the performance of the practical operation strategy and the above-mentioned optimal operation strategies. In the CU column, parentheses indicate the corresponding fleet size, $obj_c$ indicates operating costs (including CU), and $obj_{j,nocu}$ indicates the operating costs without considering CU. We first compare the practical operation strategy with IOP⁰ and IOP-noCU⁰. These computational results show that both IOP⁰ and IOP-noCU⁰ save operating costs effectively, while ensuring the same level of service quality. We use boldface to indicate the best indicators in the IOP⁰ and IOP-noCU⁰ solutions. It should be noted that $PW$ is not improved in optimization-based solutions (compared to the practical operation strategy), as the practical operation strategy sets the number of effective services to their (possible) maximum value and addresses the $PW$ minimization. Overall, both IOP⁰ and IOP-noCU⁰ solutions increase the passenger waiting time, while reducing the passenger travel time (to ensure service quality), thus triggering the cancellation of some effective services in the practical operation strategy. This approach also reduces the CM. Moreover, the IOP⁰ solutions reduce operating costs from the perspective of reducing CU, while the IOP-noCU⁰ solutions reduce operating costs from the perspective of reducing CE.

In Appendix E, we present and discuss the obtained service frequencies, timetables, rolling stock plans, and train speed profile selections for the five operation strategies of Table 11.

### 5.4. Analysis of algorithm efficiency

In this section, we analyze the computational efficiency of our methodology. For the IOP and IOP-noCU cases discussed in Section 5.1, a total of 117,708 schemes (i.e., the SOPs) are generated during the pre-processing procedure. Applying our FDP algorithm, approximately 25 seconds are required to generate the optimal solution for each SOP. Accordingly, 776 hours are required to generate an optimal solution for the IOP or IOP-noCU without considering the speed-up techniques. We report computation times for all the cases of Tables 8 and 9 in column 4 of Tables 12 and 13, respectively. After embedding the proposed speed-up techniques, the computation time (CPU) for each case in the IOP is below three hours, while the

---

**Table 11**

Performance comparison of the practical operation strategy and the four optimal strategies.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>$PW(\delta)$</th>
<th>$PT(\delta)$</th>
<th>CM($\forall$)</th>
<th>CU($\forall$)</th>
<th>CE($\forall$)</th>
<th>$obj_c(\delta)$</th>
<th>$obj_c(\forall)$</th>
<th>$obj_{j,nocu}(\forall)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practical</td>
<td>282745</td>
<td>617553</td>
<td>1602.78</td>
<td>2944.32 (8)</td>
<td>2339.68</td>
<td>1041670.5</td>
<td>6886.78</td>
<td>3942.46</td>
</tr>
<tr>
<td>IOP</td>
<td>292100</td>
<td>494298</td>
<td>1498.62</td>
<td>2576.28 (7)</td>
<td>2606.89</td>
<td>932448</td>
<td>6681.79</td>
<td>-</td>
</tr>
<tr>
<td>IOP-noCU⁰</td>
<td>328085</td>
<td>541891</td>
<td>1348.98</td>
<td>2944.32 (8)</td>
<td>2004.52</td>
<td>1031033.5</td>
<td>-</td>
<td>3353.50</td>
</tr>
<tr>
<td>IOP</td>
<td>286964</td>
<td>492128</td>
<td>1534.62</td>
<td>2576.28 (7)</td>
<td>2698.06</td>
<td>922574</td>
<td>6808.96</td>
<td>-</td>
</tr>
<tr>
<td>IOP-noCU⁰</td>
<td>290472</td>
<td>522316</td>
<td>1500.18</td>
<td>2944.32 (8)</td>
<td>2329.06</td>
<td>958024</td>
<td>-</td>
<td>3829.24</td>
</tr>
</tbody>
</table>
computation time for each case in the IOP-noCU is below six hours, which are both acceptable for tactical operational planning.

To evaluate the effectiveness of our speed-up techniques further, we provide additional performance indicators in Tables 12 and 13. CPU in the fifth column represents the computation time for the pre-processing phase (i.e., scheme ranking). In Tables 12 and 13, the CPU is approximately two hours for each case, which is equivalent to an average time of 0.06 seconds for pre-processing each scheme. In addition, since the pre-processing of each scheme is independent of the other, we observe that parallel computing techniques can potentially reduce this computation time.

During the scheme ranking process, all the schemes are ranked based on an increasing order of POs. The resulting SOPs are solved from the lowest PO value. We use the optimal solution of the first SOP as the initial upper bound (IUB). Let OB be the value of the optimal IOP solution. We use the formula (IUB-OB)/OB to generate the initial gap (IGAP). The IGAP is reported in column 6 of Tables 12 and 13. In most cases, IGAP is controlled below 1%. Even in cases 7–14 of the IOP, the optimal solutions are obtained after solving the first SOP, which confirms that the scheme ranking can successfully find a very tight initial upper bound to accelerate the process of finding the optimal IOP solution.

The CPU in column 7 of Tables 12 and 13 represents the computation time of the solving phase, that is, the time taken from the completion of the pre-processing phase to the completion of the overall IOP solution process. Although we find that the CPU in Tables 12 and 13 satisfy the computational requirements of the tactical decision level, the CPU can be very different for various cases. For example, in Table 12, case 8 takes only 25 seconds, while case 1 takes more than 3 hours. We recall that it takes 25 seconds to complete each solution process of SOP. 3 seconds to terminate the solution process in advance (i.e., the second stopping criterion is met), and only 0.06 seconds to calculate the potential of each

| Table 12 | Indicators of computational efficiency for the IOP. |
|---|---|---|---|---|---|---|
| Case | w_p | w_i | CPU(s) | CPU(s) | IGAP(%) | CPU(s) | Iter.E | Iter.EX |
| 1 | 0.001 | 0.999 | 20,488 | 7450 | 3.76077 | 13,038 | 626 | 550 |
| 2 | 0.1 | 0.9 | 8503 | 6924 | 0.07311 | 1579 | 626 | 2 |
| 3 | 0.12 | 0.88 | 8800 | 7239 | 0.08544 | 1561 | 626 | 2 |
| 4 | 0.14 | 0.86 | 8929 | 7382 | 0.09851 | 1547 | 626 | 2 |
| 5 | 0.16 | 0.84 | 8843 | 7257 | 0.11062 | 1586 | 626 | 2 |
| 6 | 0.18 | 0.82 | 8923 | 7355 | 0.12295 | 1568 | 626 | 2 |
| 7 | 0.2 | 0.8 | 6824 | 6778 | 0 | 46 | 3 | 2 |
| 8 | 0.3 | 0.7 | 7924 | 7899 | 0 | 25 | 2 | 1 |
| 9 | 0.4 | 0.6 | 8192 | 8168 | 0 | 24 | 2 | 1 |
| 10 | 0.5 | 0.5 | 7805 | 7781 | 0 | 24 | 2 | 1 |
| 11 | 0.52 | 0.45 | 7240 | 7215 | 0 | 25 | 2 | 1 |
| 12 | 0.55 | 0.45 | 7239 | 7214 | 0 | 25 | 2 | 1 |
| 13 | 0.56 | 0.44 | 7258 | 7233 | 0 | 25 | 2 | 1 |
| 14 | 0.58 | 0.42 | 8327 | 7187 | 0 | 1140 | 482 | 1 |
| 15 | 0.6 | 0.4 | 9809 | 7978 | 0.22308 | 1831 | 666 | 16 |
| 16 | 0.7 | 0.3 | 8328 | 7494 | 0.30460 | 834 | 302 | 6 |
| 17 | 0.8 | 0.2 | 7715 | 7319 | 0.10085 | 396 | 153 | 2 |
| 18 | 0.9 | 0.1 | 8433 | 8041 | 0.00839 | 392 | 149 | 1 |
| 19 | 0.999 | 0.001 | 7441 | 7056 | 0 | 385 | 149 | 2 |

| Table 13 | Indicators of computational efficiency for the IOP-noCU. |
|---|---|---|---|---|---|---|
| Case | w_p | w_i | CPU(s) | CPU(s) | IGAP(%) | CPU(s) | Iter.E | Iter.EX |
| 1 | 0.001 | 0.999 | 21,478 | 7835 | 5.70879 | 13,643 | 626 | 550 |
| 2 | 0.1 | 0.9 | 8706 | 7139 | 3.92652 | 1567 | 626 | 3 |
| 3 | 0.2 | 0.8 | 8825 | 7275 | 0.18659 | 1550 | 626 | 2 |
| 4 | 0.3 | 0.7 | 7729 | 7324 | 0.20930 | 405 | 151 | 2 |
| 5 | 0.4 | 0.6 | 8512 | 7895 | 1.46575 | 617 | 68 | 18 |
| 6 | 0.42 | 0.58 | 8973 | 7493 | 0.28893 | 1480 | 333 | 44 |
| 7 | 0.44 | 0.56 | 10,293 | 7530 | 0.19067 | 2763 | 850 | 56 |
| 8 | 0.45 | 0.54 | 12,556 | 7750 | 0.24855 | 4806 | 1780 | 48 |
| 9 | 0.48 | 0.52 | 17,716 | 7705 | 0.34958 | 10,011 | 4012 | 34 |
| 10 | 0.5 | 0.5 | 20,635 | 7981 | 0.37745 | 12,654 | 4404 | 21 |
| 11 | 0.52 | 0.48 | 19,653 | 7626 | 0.16767 | 12,027 | 4870 | 19 |
| 12 | 0.54 | 0.46 | 21,582 | 8090 | 0.34320 | 13,492 | 5511 | 13 |
| 13 | 0.56 | 0.44 | 21,000 | 8107 | 0.30768 | 12,893 | 5217 | 12 |
| 14 | 0.58 | 0.42 | 19,960 | 8418 | 0.2111 | 11,542 | 4912 | 4 |
| 15 | 0.6 | 0.4 | 18,369 | 7382 | 0.13910 | 11,334 | 4727 | 2 |
| 16 | 0.7 | 0.3 | 12,315 | 7516 | 0.06148 | 4799 | 2012 | 5 |
| 17 | 0.8 | 0.2 | 8763 | 7326 | 0.22735 | 1437 | 594 | 2 |
| 18 | 0.9 | 0.1 | 8229 | 7402 | 0.03046 | 827 | 326 | 3 |
| 19 | 0.999 | 0.001 | 7992 | 7575 | 0.00014 | 417 | 149 | 2 |
scheme (used in the first stopping criterion). The CPU\(^5\) difference clearly relates to the performance of the first and second stopping criteria.

Iter.E and Iter.EX in the last two columns of Tables 12 and 13 indicate the performance of the first and second stopping criteria, respectively. Specifically, Iter.E indicates the number of SOPs that must be solved (until the first stopping criterion is met), while Iter.EX indicates the number of SOP solution processes involving state transitions from stage \(S\) to stage \(S+1\) (i.e., the second stopping criterion is not met). From the Iter.E column, we observe that the number of SOPs that must be solved is greatly reduced through the first stopping criterion. Only a maximum of 5217 schemes (case 13 in Table 13) of the 111,708 generated schemes must be considered during the IOP solution process. From the Iter.EX column, most of the considered schemes meet the second stopping criterion. Thus, these schemes are terminated in advance. Overall, the proposed stopping criteria effectively reduce the computation time of the solving phase.

6. Conclusions and further research

In this study, we propose an integrated optimization approach for managing the asymmetric operation strategies based on the asymmetry in passenger demand on different directions on a subway line. Our approach includes the following four aspects: service frequency planning (determining the number of effective services), timetabling (generating effective services during peak and off-peak hours), rolling stock planning (matching the rolling stock units with the effective services), and optimization of train speed profiles (selecting the train speed profile mode over each segment). These four aspects are integrated via a mixed-integer non-linear programming model to simultaneously optimize the service quality (i.e., passenger waiting and travel times) and operating costs (i.e., rolling stock utilization costs, maintenance costs, and energy costs, including reused energy considerations). Based on this modeling approach, we propose an exact algorithm, which first decomposes the original IOP into multiple sub-problems (i.e., SOPs) and uses a dedicated forward dynamic programming algorithm to solve each SOP. Speed-up techniques, including scheme ranking and stopping criteria, are embedded into the proposed algorithm to improve its computational efficiency. We design numerical experiments based on the Yizhuan line, to analyze the performance of the proposed methodology. As the objective preference changes, the nondominated solutions for the IOP aggregate around several fleet size cases, while the nondominated solutions for the IOP-noC\(P\) clearly outperform symmetric strategies, especially when there is a significant trade-off between the two objectives (e.g., when the weight ratio is 0.5:0.5). By comparing the optimization-based solutions with the practical operation strategy, the former strategies present a much better performance in terms of both improving the service quality and reducing operating costs. Furthermore, the proposed speed-up techniques significantly reduce the computation time required to find an optimal IOP solution, thus satisfying the computational requirements at a tactical decision level.

Future research could address the following aspects: (1) extend our methodology to robust optimization approaches (as indicated, e.g., in Liu et al. (2020a), Gong et al. (2021)); (2) investigate further traffic management measures, such as passenger control strategies and train capacity optimization, to handle extreme passenger demand during peak hours; (3) improve selection of train speed profiles, for instance, by optimizing train speed profiles according to the specific characteristics of the services scheduled over each track (as indicated, e.g., in Wang et al. (2020), Ghasempour and Heydecker (2020)); and (4) develop optimization-based energy reuse methods, involving the consideration of the dynamic onboard passengers and the adjustment of train dwell and travel times.

CRediT Authorship contribution statement

Pengli Mo: Conceptualization, Methodology, Writing - original draft. Andrea D’Ariano: Methodology, Writing - review & editing. Lixing Yang: Methodology, Writing - review & editing, Funding acquisition. Lucas P. Veelenturf: Validation, Writing - review & editing. Ziyou Gao: Supervision, Project administration.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 71825004, 71771016, 71621001, 72001020) and the Beijing Natural Science Foundation (L191015, L201016). The authors would also like to thank the three anonymous referees for their valuable comments and suggestions.

Appendix A. Formulation of rolling stock plan related to depot 2

\[
\alpha_k^{(2)} = 1 - \sum_{k \in \mathcal{N}} \gamma_{k,k'}^{(2)}, \quad \forall \ k' \in \mathcal{N}, \tag{41}
\]

\[
\beta_k^{(2)} = 1 - \sum_{k \in \mathcal{N}} \gamma_{k,k'}^{(2)}, \quad \forall \ k \in \mathcal{N}, \tag{42}
\]

\[
a_{k,S} - d_{k,S} \geq T(1) - T(1 - \gamma_{k,k}^{(2)}), \quad \forall \ k \in \mathcal{N}, k' \in \mathcal{N}, \tag{43}
\]
\[ a_{k',s+1} - d_{k,5} \leq T^{(l)} + T(1 - y_{k,k}^{(2)}), \forall k' \in \mathcal{N}, k' \in \mathcal{N} \]  
\[ \sum_{k \in \mathcal{N}} \chi_{k,k}^{(2)} \leq \alpha_{k'}^{(2)}, \forall k' \in \mathcal{N}, \]  
\[ \sum_{k' \in \mathcal{N}} \chi_{k,k}^{(2)} \leq \beta_{k}^{(2)}, \forall k \in \mathcal{N}, \]  
\[ a_{k',s+1} - d_{k,5} \geq T^{(2)} - T(1 - x_{k,k}^{(2)}), \forall k \in \mathcal{N}, k' \in \mathcal{N}, \]  
\[ \phi^{(2)} = \sum_{k' \in \mathcal{N}} \left[ 1 - \sum_{k \in \mathcal{N}} \left( \chi_{k,k}^{(2)} + y_{k,k}^{(2)} \right) \right] \]  

Appendix B. Formulation of electric power

B1. Formulation of \( r_i^{(a)}(\tau) \) and \( r_i^{(b)}(\tau) \)

As shown in Fig. 10, the energy supplied by the power system is used to accelerate rolling stock units. The resistance must be overcome during train operations, and the train engine has a mechanical loss. When a train brakes, its engine reverses and converts a part of the kinetic energy into electric energy. The regenerated energy can be used to accelerate other rolling stock units. However, the excess energy needs to be converted into heat to dissipate (air braking or resistance consumption). During train operations, the energy provided by the power system is converted into other forms of energy, some of which can be measured by power conversion coefficients, such as traction efficiency \( \eta_t \) and braking efficiency \( \eta_b \).

This study calculates the kinetic energy (green part in Fig. 10) during train operations and uses a power conversion coefficient to obtain the electric power during traction and braking, that is, \( r_i^{(a)}(\tau) \) and \( r_i^{(b)}(\tau) \). In this study, without loss of generality, we consider the train speed profiles during real-world operations. As shown in Fig. 11, we take the actual train speed profiles of the Yizhuang line as an example. The black line represents the train speed profile recorded by the train. The train speed profile is irregular, especially in the discrete control system.

In this study, we use speed, acceleration, and gear condition in the train speed profile data to calculate the kinetic energy. The gear condition includes states 1, 0, and -1 (representing forward rotation, stalling, and reverse rotation of the engine). The blue curve in Fig. 11 is an example of the gear condition during operations. Compared with using acceleration to detect
the engine state, using the gear condition, we can present a more accurate interpretation of the engine state. For example, when the engine overcomes the resistance (speed remains constant), acceleration cannot detect the forward rotation of the engine at this time.

The adopted notation is listed in Table 14. Next, we introduce the model for calculating the electric power by using speed $V_{l,i}(\tau)$, acceleration $A_{l,i}(\tau)$, and gear conditions $G_{l,i}(\tau)$. First, we calculate the state of traction force $f_{k,i,l}^{(a)}(\tau)$ and braking force $f_{k,i,l}^{(b)}(\tau)$ over time corresponding to each train speed profile mode at each segment, as follows:

$$f_{k,i,l}^{(a)}(\tau) = \begin{cases} G_{l,i}(\tau) > 0 & (M^{(i)} + m_{k,i}^{(i)}) \cdot \left[ A_{l,i}(\tau) + (R_{k,i}^{(g)}(\tau) + R_{l,i}^{(c)}(\tau)) \cdot g \right] + R_{l,i}^{(b)}(\tau), \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in S \setminus \{S\}, k \in \mathbb{N}, l \in L_i, \tau \in T_{l,i} \text{ or } \forall i \in S \setminus \{2S\}, k \in \mathbb{N}, l \in L_i, \tau \in T_{l,i}.
$$

$$f_{k,i,l}^{(b)}(\tau) = \begin{cases} G_{l,i}(\tau) \geq 0 & (M^{(i)} + m_{k,i}^{(i)}) \cdot \left[ A_{l,i}(\tau) - (R_{k,i}^{(g)}(\tau) + R_{l,i}^{(c)}(\tau)) \cdot g \right] - R_{l,i}^{(b)}(\tau), \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in S \setminus \{S\}, k \in \mathbb{N}, l \in L_i, \tau \in T_{l,i} \text{ or } \forall i \in S \setminus \{2S\}, k \in \mathbb{N}, l \in L_i, \tau \in T_{l,i}.
$$

In formulas (49) and (50), we use the set $\mathcal{T}$ instead of $T$ to illustrate discrete times, because the train speed profiles belong to a microscopic level, and the time unit of the train speed profiles usually reaches the millisecond level, which is more accurate than the other three procedures (i.e., service frequency planning, timetabling, and rolling stock planning). Therefore, we introduce a more accurate discrete-time set $\mathcal{T}$ to handle the train speed profiles data. The gradient resistance coefficient $R_{k,i}^{(g)}(\tau)$ in formulas (49) and (50) is calculated based on the track slope. The curve resistance coefficient $R_{l,i}^{(c)}(\tau)$ is calculated based on the lateral friction between the train and the track. The basic resistance $R_{l,i}^{(b)}(\tau)$ is calculated according to the well-known Davis formula (Davis, 1926),

$$R_{l,i}^{(b)}(\tau) = \theta_1(V_{l,i}(\tau))^2 + \theta_2 V_{l,i}(\tau) + \theta_3.$$
(G_{i,l}(\tau) < 0) and the corresponding train is accelerating (decelerating). A_{i,l}(\tau) is negative if this is in a different state from G_{i,l}(\tau). For example, A_{i,l}(\tau) < 0 if G_{i,l}(\tau) < 0 (G_{i,l}(\tau) > 0) and the corresponding train is accelerating (decelerating).

m_{k,l}^{(p)} is the mass of onboard passengers after service k departs from station i. In this study, we consider the impact of headway on m_{k,l}^{(p)}, while we omit the impact of varying the train speed profiles on m_{k,l}^{(p)}. In other words, we assume that the values of m_{k,l}^{(p)} corresponding to timetables with the same departure times from stations 1 and 25, are the same. m_{k,l}^{(p)} is calculated according to the headway (between this service and its front service) and the corresponding passenger arrival rate. In our problem, headway at each station will not be changed when adjusting train speed profiles. Moreover, passenger arrival rates do not change evidently within a short time period (e.g., within 10 minutes), without considering special events (such as concerts and football games). Thus, adjusting train speed profiles has a relatively insignificant impact on the mass of onboard passengers, especially when compared to the mass of rolling stock. In the solution method, for each SOP, in which the departure times of effective services at stations 1 and 25 are fixed, we can use a feasible solution of the SOP (also a feasible solution of the SP-PME) to calculate the corresponding m_{k,l}^{(p)}. Please refer to Section 4.2 for detailed definitions of the SOP and SP-PME. The calculation method we use to handle m_{k,l}^{(p)} can avoid the computational burden and the curse of dimensionality caused by considering the dynamic passenger weight.

We next calculate electric power utilization in traction/braking operations, which depends on the traction/braking force of the current mode, speed, traction efficiency, and braking efficiency. Moreover, to match the regenerative energy utilization among multiple rolling stock units, we must calculate the total electric power of all services and embed the electric power into time set \( \mathcal{T} \) by considering the timetable. The traction/braking electric power in each segment is calculated by the following formulas.

\[
 r_{l}^{(a)}(\tau) = \begin{cases} 
 \frac{1}{\eta_1} \sum_{k \in \mathcal{K}_l} \sum_{i \in \mathcal{I}_l} x_{i,l} \cdot f_{k,l,l}^{(a)}(\tau - \frac{d_{i,l} \\delta}{\sigma}) \cdot V_{i,l}(\tau - \frac{d_{i,l} \\delta}{\sigma}), \\
 \text{if } i \in \mathcal{S} \setminus \{S\}, \quad \overline{N}_{i,l}(\tau) = \{k \in \overline{\mathcal{N}} : \tau - \frac{d_{i,l} \\delta}{\sigma} \in \tau_{l,l}\} \\
 0, \text{ otherwise} 
\end{cases}
\]

\[
 r_{l}^{(b)}(\tau) = \begin{cases} 
 \frac{1}{\eta_2} \sum_{k \in \mathcal{K}_l} \sum_{i \in \mathcal{I}_l} x_{i,l} \cdot f_{k,l,l}^{(b)}(\tau - \frac{d_{i,l} \\delta}{\sigma}) \cdot V_{i,l}(\tau - \frac{d_{i,l} \\delta}{\sigma}), \\
 \text{if } i \in \mathcal{S} \setminus \{S\}, \quad \overline{N}_{i,l}(\tau) = \{k \in \overline{\mathcal{N}} : \tau - \frac{d_{i,l} \\delta}{\sigma} \in \tau_{l,l}\} \\
 0, \text{ otherwise} 
\end{cases}
\]

B2. Formulation of \( \tilde{r}_{l,l} \)

For \( i \in \mathcal{S} \setminus \{S, 25\} \), \( l \in \mathcal{L}_i \), \( \tilde{r}_{l,l} \) is calculated as follows:

\[
 \tilde{r}_{l,l} = \begin{cases} 
 \sum_{k \in \overline{\mathcal{N}}} \sum_{r \in \mathcal{R}_l} (\tilde{r}_{k,l,l}^{(a)}(\tau) - \tilde{r}_{k,l,l}^{(b)}(\tau)) \quad & i \in \mathcal{S} \setminus \{S\}, \\
 \sum_{k \in \overline{\mathcal{N}}} \sum_{r \in \mathcal{R}_l} (\tilde{r}_{k,l,l}^{(a)}(\tau) - \tilde{\eta}_{l,l} \cdot \tilde{r}_{k,l,l}^{(b)}(\tau)) & i \in \mathcal{S} \setminus \{25\}, 
\end{cases}
\]

where

\[
 \tilde{r}_{k,l,l}^{(a)}(\tau) = \frac{1}{\eta_1} f_{k,l,l}^{(a)}(\tau) V_{i,l}(\tau) \sigma
\]

and

\[
 \tilde{r}_{k,l,l}^{(b)}(\tau) = \eta_2 f_{k,l,l}^{(b)}(\tau) V_{i,l}(\tau) \sigma.
\]

\( \tilde{r}_{k,l,l}^{(a)}(\tau) \) and \( \tilde{r}_{k,l,l}^{(b)}(\tau) \) represent the traction energy consumption and the regenerative energy, in which \( f_{k,l,l}^{(a)}(\tau) \) and \( f_{k,l,l}^{(b)}(\tau) \) are calculated by formulas (49) and (50). \( \tilde{\eta}_{l,l} \) represents an upper bound of the energy reuse rate from the regenerative energy and can be calculated by simulation. We next give a calculation method of \( \tilde{\eta}_{l,l} \) in the condition that there
is at most one rolling stock unit at any one time in each segment. Specifically, we obtain \( \bar{\eta}_{l,i} \) by simulating the combinations of train speed profile \( l \) over segment \( i \) and another speed profile \( l' \in \mathcal{L}_{2S-i} \). Each train speed profile \( l \) over segment \( i \) thus has \(|\mathcal{L}_{2S-i}|\) combinations with the train speed profiles over segment \( 2S-i \). For each combination, as shown in Fig. 12, we can obtain the maximum energy reuse rate by enumerating all overlap relationships between the two involved train speed profiles (i.e., between \( \mathcal{P}_{s,l}(\tau) \) and \( \mathcal{P}_{s',2S-i,l'}(\tau) \), where \( s = \arg \max_{k \in \mathcal{N}} m_{k,l}^{(p)} \) and \( s' = \arg \max_{k \in \mathcal{N}} m_{k,2S-i,l}^{(p)} \)), in which we generate the overlap relationships by shifting train speed profile \( l' \) (i.e., \( \mathcal{P}_{s',2S-i,l'}(\tau) \)) a time interval \( (\delta) \) sequentially. For example, we shift \( \mathcal{P}_{s',2S-i,l'}(\tau) \) one interval \( \delta \) and use \( \mathcal{P}_{s',2S-i,l'}(\tau) \) to denote the shifted \( \mathcal{P}_{s',2S-i,l'}(\tau) \). A new reused energy is obtained according to \( \min(\mathcal{P}_{s',2S-i,l'}(\tau), \mathcal{P}_{s',2S-i,l'}(\tau)) \), and then a new energy reuse rate is obtained. We set \( \bar{\eta}_{l,i} \) as the maximum energy reuse rate of train speed profile \( l \) over segment \( i \) among all overlap relationships of its \(|\mathcal{L}_{2S-i}|\) combinations. \( \bar{\eta}_{l,i} \) can also be obtained by other simulation methods. In addition, \( \bar{\eta}_{l,i} \) does not exceed 100%.

Appendix C. Complexity analysis of IOP model

The IOP variables are provided in Table 15. The complexity of the IOP depends on the number of services in the candidate sets (i.e., \( \bar{N} \max \) and \( \bar{N} \max \)), number of effective services (i.e., \( \bar{N} \) and \( \bar{N} \)), and the various train speed profile modes (i.e., \( \bar{L} i \)).

In the proposed model, the constraints must be formulated by considering two or more categories of decision variables, thus coupling these four decision variables together. To simply and clearly characterize the coupling relationship, we present a graphical description in Fig. 13.

As shown in Fig. 13, we use green, red, blue, and yellow circles to represent the formulas related to the timetable, rolling stock plan, train speed profiles, and service frequency, respectively. The four circles form four regions, and the overlaps between the circles create new hybrid regions, with each representing a coupling relationship. We use Roman numbers I to VII to label the coupling relationships in our model. For example, II indicates the coupling relationship between only the service frequency variables, while VII indicates the coupling relationship between the variables with respect to service frequency, timetable, and train speed profiles. We next classify the formulas proposed in Section 3 according to these seven coupling relationships, and the specific classification results are shown in Table 16.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Variables</th>
<th>Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service frequency</td>
<td>Integer variable</td>
<td>( \bar{N} ) ( \bar{N} \max )</td>
</tr>
<tr>
<td>Service frequency</td>
<td>Integer variable</td>
<td>( \bar{N} ) ( \bar{N} \max )</td>
</tr>
<tr>
<td>Timetable</td>
<td>Integer variables</td>
<td>( \bar{N} ) ( \bar{N} ) ( \bar{N} ) ( \bar{N} \max )</td>
</tr>
<tr>
<td>Timetable</td>
<td>Integer variables</td>
<td>( \bar{N} ) ( \bar{N} ) ( \bar{N} ) ( \bar{N} \max )</td>
</tr>
<tr>
<td>Train speed profiles</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{Y}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{Y}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
<tr>
<td>Rolling stock plan</td>
<td>Binary variables</td>
<td>( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} ) ( \bar{X}<em>{i,j} ) ( \bar{X}</em>{i,j} )</td>
</tr>
</tbody>
</table>

312
Appendix D. Data preparation

The Yizhuang line is the key subway line that connects downtown Beijing and the residential zones along the studied line. The metro operations direction from the Yizhuang Station to the Songjiazhuang station is considered as the upstream direction, while the opposite is considered as the downstream direction. Evidently, passenger demand on the Yizhuang line is obviously asymmetric in both directions (i.e., up and down) during different times of the day, especially during the morning peak hours (i.e., from 6:00 am to 10:00 am). In this study, we focus on the passenger demand metric from 6:00 am to 10:00 am and we use real passenger demand data on the Yizhuang line collected on April 24, 2015, with the aim of calibrating the time-dependent OD matrix $P_{i,j}$. In addition, $\delta$ and $\sigma$ (the time units to discretize the planned time horizon) are set as 15 seconds and 1 second, respectively, such that the planned time horizon is split into 960 and 14,400 units, that is, $T = 960$ and $T = 14400$.

As shown in Fig. 14, the Yizhuang line is a bidirectional subway line with 14 physical stations, with 6 underground and 8 above-ground stations. Additionally, the Taihu and Songjiazhuang depots are located near the Ciqiu and Songjiazhuang stations, respectively. Considering that the Yizhuang railway station is not operated, the operating section of the Yizhuang line only includes 13 physical stations from the Songjiazhuang station to the Ciqiu station, that is, $S = 13$. As each physical station is modeled as two separate stations, that is, one for each direction, there are 26 stations and 26 segments (including two turn-back segments) on the Yizhuang line. In each segment, rolling stock units operate by following the train speed profiles, and we consider that each segment (excluding the turn-back segments) has three modes of train speed profiles (i.e., $L = 3$), as shown in Table 17. The three modes of train speed profiles over each segment correspond to different train travel times. For the turn-back segments, we set the upper and lower time limits for the turn-back operations. The parameters of travel and dwell times in Table 17 are all set according to the actual operational data.

Table 18 lists the other parameter values. According to the daily operations in 2015, the headways during the morning peak hours and off-peak hours are set as 6 and 9 minutes, respectively. We pre-determine the departure times of the first

Table 16
Classification of formulas for the integrated optimization model.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Formulas</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(28), (32)</td>
</tr>
<tr>
<td>II</td>
<td>(3)</td>
</tr>
<tr>
<td>III</td>
<td>(12)</td>
</tr>
<tr>
<td>IV</td>
<td>(1)-(2), (8)-(10), (21)-(23)</td>
</tr>
<tr>
<td>V</td>
<td>(13)-(14), (17)-(18), (20), (25)-(27), (41)-(42), (45)-(46), (48)</td>
</tr>
<tr>
<td>VI</td>
<td>(11), (24), (29)-(31)</td>
</tr>
<tr>
<td>VII</td>
<td>(15)-(16), (19), (43)-(44), (47)</td>
</tr>
</tbody>
</table>

Fig. 13. The coupling relationships in the IOP.

Fig. 14. Illustration of the Beijing subway Yizhuang line.
effective services from stations 1 and 26 (nearby the Taihu depot) according to the April 2015 timetable, since the parking and dispatching of rolling stock units on the Yizhuang line in 2015 are mainly related to the Taihu depot. In the April 2015 timetable, after 6:00 am, the departure time of the first service at station 1 is 6:03, while the departure time of the first service at station 25 is 6:24. In this timetable, the maximum number of services in each direction (i.e., $N_{\text{max}}$ and $\bar{N}_{\text{max}}$) is set as 37, and the maximum difference between the numbers of effective services in the two directions (i.e., $\bar{N}^{(e)}$) is set as 5. The maximum number of rolling stock units is 12. We next introduce the parameters related to the performance of rolling stock units. Rolling stock units running on the Yizhuang line consist of 6 DKZ32 units, with a total mass of 84 t. Energy conversion efficiencies $\eta_1$, $\eta_2$, $\eta_3$ are calibrated to 0.96, 0.44, and 0.9, respectively. According to Fan et al. (2016), passengers overestimate the waiting time (no more than 10 minutes) by 1.5 times, on average. We thus set $w_0$ as 1.5. We set the utilization costs of each rolling stock unit as 368.04 ¥, which may include crew costs during the planned time horizon. Operators can set rolling stock utilization price to other values according to their specific preferences. The price of energy is 0.5 ¥/kWh. The maintenance costs of each service executed by the rolling stock unit are 21.15 ¥, while the maintenance costs of moving to depots 1 and 2 are 3.93 ¥ and 0.78 ¥, respectively. Based on the distance from the station to depots 1 and 2, the amount of energy consumption of each rolling stock unit returning to depot 1 and depot 2 are estimated to be 7.5 kWh and 5 kWh, respectively.

Appendix E. Illustration of practical operation strategy and four optimized solutions

For further analysis, we draw the timetable, rolling stock plan, and train speed profiles of the practical operation strategy and the four optimized solutions in Figs. 15, 16, 17, 18, and 19, respectively. In the representation of these timetables, the red lines indicate effective services in the upstream direction, while the blue lines indicate effective services in the downstream direction. In addition, we use shadows to distinguish services during peak hours from all effective services. Moreover, each green return line indicates that two services are connected by a rolling stock unit turn-back on the line, while each gold dotted line indicates that a rolling stock unit returns to or departs from the corresponding depots. In our representation of these rolling stock plans, the horizontal axis represents time, while the vertical axis represents the index of rolling stock unit. On the horizontal line corresponding to each rolling stock unit, squares represent the effective services that are required to be executed. Through these figures, we can clearly identify the specific services executed by each rolling stock unit. In the proposed representation of train speed profiles, the curve shows the selected mode of train speed profiles at each segment,
Fig. 15. Practical operation strategy.

while the bar graph shows the sectional passenger data for each segment. It should be noted that segments 13 and 26 are turn-back segments without passengers.

Comparing the practical timetable with the proposed optimization-based timetables, we can summarize the following managerial insights. First, concerning the timetable of asymmetric operation strategy (i.e., IOP and IOP-noCU), a reduction of the train travel time is only a necessary condition for reducing the fleet size, while it is also necessary to reasonably arrange the start times of the peak periods in the two directions. Taking the timetable of IOP as an example, the peak period in the upstream direction starts from the 3rd effective service, while the peak period in the downstream direction starts from the 5th effective service. If the downstream peak period also starts from the 3rd service, this 3rd service cannot
be executed by rolling stock units from the upstream direction. The Songjiazhuang depot must then dispatch the third rolling stock unit to execute this service.

Second managerial insight, changing the duration of peak periods in both directions may affect the rolling stock plan but can meet the asymmetry of passenger demand. The duration of peak periods (shaded parts) in the optimization results are shorter than in the practical operation strategy. It is also worth noting that the duration of peak periods in the two directions are not the same. The duration of upstream peak period is longer than that of the downstream peak period. This corresponds to the asymmetry in passenger demand, in which the pressure of passenger demand in the upstream direction is greater than that in the downstream direction. Additionally, the difference in peak period duration between the two directions
can cause a difference in the number of effective services, making the number of rolling stock units departing from each depot inconsistent with the number of rolling stock units returned. As shown in Fig. 17(b), five rolling stock units depart from the Taihu depot and four rolling stock units return, while two rolling stock units depart from the Songjiazhuang depot and three rolling stock units return. In fact, this difference meets the daily operational needs. Specifically, the downstream direction will have more passengers during the evening peak hours and, ideally, more effective services will be assigned to this direction during that period. The additional rolling stock units in the Songjiazhuang depot will not only return to the Taihu depot during the evening peak hours but will also be used as rolling stock units required for dispatch from the Taihu depot.
A third managerial insight is that the selection of the train speed profiles is evidently connected to the passenger demand, the rolling stock plan, and the operating environment at each segment. Compared with the selection of practical train speed profiles (Fig. 15(c)), there is a tendency to use the faster train speed profiles (e.g., mode 1) on segments with more passengers (such as segments 12 and 14), such that the travel time of more number of passengers can be reduced. Similarly, the slower train speed profiles (e.g., mode 3) are used over segments with lower passenger demand (such as segments 2 and 23) to save energy. In addition, the selection of train speed profiles is also related to the rolling stock plan. As mentioned above, a speedier train travel time can accelerate the circulation of rolling stock. To reduce the fleet size, the selection of train speed profiles in Figs. 16(c) and 18(c) use more profiles with faster train travel times than those in Figs. 17(c) and 19(c). The selection of train speed profiles fluctuates on some intermediate segments (such as segments 4–10 and 16–18),
possibly due to (i) the fine-tuning of train speed profiles to increase the reused energy; and (ii) the different energy costs for reducing the train travel time at each segment.

These three managerial insights can be used in specific operation strategies to adjust the effective services, rolling stock plans, and train speed profiles. Specifically, for further optimization of the practical operation strategy of the Yizhuang line in 2015, we can adjust both the duration and start time of peak-hour timetables in both directions according to the asymmetry in passenger demand to cancel some services, which can save operating costs. Moreover, adjusting the train speed profiles can reduce passenger travel time and energy consumption, thereby improving the service quality and saving operating costs. In addition, the adjustment of both timetables and train speed profiles must follow these three managerial insights, which
demonstrate the relationship between the service frequency planning, timetabling, train speed profile selection, and rolling stock planning in the asymmetric strategy.

References


Liu, R., Li, S., Yang, L., 2020. Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy. Omega (Westport) 90, 101990.


