Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches

Jiateng Yin [a], Lixing Yang [a,*], Tao Tang [a], Ziyou Gao [a], Bin Ran [b]

[a] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China
[b] Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI, 53706, USA

ARTICLE INFO

ABSTRACT

In the daily operation of metro systems, the train scheduling problem aims to find a set of space-time paths for multiple trains that determine their departure and arrival times at metro stations, while train operations are in charge of selecting the best operational speed to satisfy the punctuality and operation costs. Different from the most existing researches that treat these two problems separately, this paper proposes an integrated approach for the train scheduling problem on a bi-direction urban metro line in order to minimize the operational costs (i.e., energy consumption) and passenger waiting time. More specifically, we simultaneously consider (1) the train operational velocity choices that correspond to the energy consumption of trains on each travelling arc, and (2) the dynamic passenger demands at each station for the calculation of total passenger waiting time in the planning horizon. By employing a space-time network representation in the formulations, this complex train scheduling and control problem with dynamic passenger demands is rigorously formulated into two optimization models with linear forms. The first model is an integer programming model that jointly minimizes train traction energy consumption and passenger waiting time. The second model, which is formulated as a mixed-integer programming model, further considers the utilization of regenerative braking energy on the basis of the first model. Due to the computational complexity of these two models, especially for large-scale real-world instances, we develop a Lagrangian relaxation (LR)-based heuristic algorithm that decomposes the primal problem into two sets of subproblems and thus enables to find a good solution in short computational time. Finally, two sets of numerical experiments, involving a relatively small-scale case and a real-world instance based on the operation data of Beijing metro are implemented to verify the effectiveness of the proposed approaches.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

With the increasing social concerns on carbon emission and global warming problems in big cities (e.g., Tokyo, Beijing, Shanghai, New York), sustainable and environmental friendly transportation modes, involving public transport

* Corresponding author.
  E-mail address: lxyang@bjtu.edu.cn (L. Yang).

(Daganzo, 2010) and electric vehicles (Li et al., 2016), are paid more and more attention in recent years. As an indispensable part of public transport, urban metro traffic is regarded as an environmental friendly transportation mode with high carrying capacity and low carbon emissions (Yang et al., 2016). For example, Beijing metro owns 18 lines with a total length of 527 km, which can carry as many as 10 million passengers to their destinations within a single day. Meanwhile, the carbon emission for transporting one passenger per kilometer by the metro line is only about 10% compared with that of a single occupancy car (Kennedy, 2002), which enables to release the environmental pressure for carbon emissions and improve the social benefits through the development of urban metro systems.

In general, the daily operations of a metro system are formed on the basis of an extensive daylong planning process, which typically involves train scheduling, rolling stock plan and crew scheduling. In particular, the metro train scheduling process aims to find a set of feasible space-time paths for all the in-service trains to determine their departure and arrival times at the stations of a metro corridor. As indicated in Cordeau et al. (1998), Zhou and Zhong (2007) and Wang et al. (2014), the train schedule has a direct affect on the service quality and operational costs of a metro system. Specifically, the service quality is commonly associated with the train travel time and passenger waiting time, while the operational costs usually refer to the total energy consumption for the movement of trains. In essence, the service level and operational costs are two objectives that may conflict with each other in the metro train scheduling process. For example, the high train departure frequency and short travel time can reduce passenger waiting and travelling time, but will inevitably increase the energy consumption for train operations.

Nevertheless, the current researches that simultaneously consider the above two objectives through a microscopic view are very few, which is partially due to the following two reasons. (1) On one hand, the calculation of passenger waiting time (or travelling time) in a metro line is directly subjected to the complex passenger flow characteristics through a day-long planning horizon. Compared with mainline railways or high-speed railways, in which most passengers make their trip plans according to the given timetable, passengers in a metro system usually do not care about the train timetable before their trips, leading to the evident dynamic (or time-variant) features of passenger demands (Freyss et al., 2013; Yang et al., 2015a). In this sense, a non-periodic timetable, i.e., a timetable with irregular headway between adjacent trains, can be more appropriate to handle the dynamics of passenger demands. However, a non-periodic timetable is usually much more difficult to obtain due to the consideration of time-variant demands, resulting in the nonlinearity of the mathematical formulations (e.g., Barrena et al., 2014b; Niu et al., 2015b); (2) on the other hand, the optimization of energy consumption for train operations is inherently a complex nonlinear problem subject to many external factors, e.g., travel time in each segment, parameters of vehicle dynamic model and rail tracks (e.g., Howlett and Pudney, 1995; Howlett et al., 2009; Ye and Liu, 2016; Yin et al., 2014; 2016a). Additionally, if the regenerative braking energy, which is an effective approach to reduce the total energy consumption by converting train braking energy for the utilization of accelerating trains, is further considered, the metro train scheduling problem will become a highly nonlinear optimization model that is much more challenging to handle.

To address these issues, this paper aims to develop a unified metro train scheduling framework that simultaneously considers the dynamic passenger demands, train traction energy consumption and the utilization of regenerative braking energy. In particular, by using a space-time network analysis, we formulate the problem into two (mixed) integer linear programming models which aim to minimize the total passenger waiting time and energy consumption (i.e., the difference between the train traction energy consumption and regenerative energy). To solve the models more efficiently, we also develop an efficient Lagrangian relaxation-based heuristic algorithm that can obtain a good solution in an acceptable computational time.

1.1. Literature review

The train scheduling problem, which aims to obtain a timetable that is to be carried out by a scale of given fleet, is a very active research field in railway operations (e.g., Cacchiani and Toth, 2012; Cacchiani et al., 2016; Wong et al., 2008; Xu et al., 2015; Kroon et al., 2014; Niu and Zhou, 2013; Niu et al., 2015b; Barrena et al., 2014b; 2014a). In general, the current solution methodologies for the rail train scheduling problem can be divided into three categories: (1) optimization approaches which use the commercial optimization softwares (e.g., CPLEX, GAMS (Yang et al., 2014)) or accurate algorithms (e.g., branch and bound algorithms (Corman et al., 2012; D'Ariano et al., 2007)), (2) heuristic algorithms (e.g., Lagrangian relaxation (Cacchiani et al., 2016; Meng and Zhou, 2014; Zhou and Zhong, 2007)) and (3) simulation methods (e.g., Dorfman and Medanic, 2004; Mu and Dessouky, 2013; Xu et al., 2015). In particular, these researches in rail train scheduling problem mainly focus on the minimization of three types of objectives, involving train delay time (Higgins et al., 1996; Cacchiani et al., 2014; Corman and Quaglietta, 2014; Corman et al., 2010), total travel time (e.g., Zhou and Zhong, 2005; 2007) and fuel consumption costs (e.g., Yang et al., 2015b). For example, Higgins et al. (1996) developed a non-linear integer programming model for the real-time scheduling of trains on a single line corridor. The objective was to minimize the total delay time in case of train conflicts and the operational costs. For the minimization of total travel time, Zhou and Zhong (2007) formulated this problem into a job shop scheduling model, which was solved efficiently by a Lagrangian relaxation combined with branch and bound algorithms.

In a passenger railway, namely a metro system or urban rail transit, it is practically desirable to design a passenger-centric timetable to highlight the convenience, reliability and reduction of passenger waiting times (Niu et al., 2015b). In general, the researches of passenger-oriented timetable design can be divided into two categories. The first kind of re-

searches focus on the design of a periodic schedule throughout the planning horizon (Carey and Crawford, 2007; Cordone and Redaelli, 2011; Li and Lo, 2014), by which the train schedules are repeated in every time interval (e.g., 1 h). The periodic timetable model is formulated on the basis of a periodic event scheduling problem, and it can yield the minimum passenger waiting times in case of constant demands (Barrena et al., 2014b). To be more realistic, recent studies begin to focus on a non-periodic timetable that explicitly considers the time-dependent origin-destination (OD) passenger demands to reduce their waiting times and travelling times, e.g., Sun et al. (2014), Niu and Zhou (2013), Niu et al. (2015b), Barrena et al. (2014b), Wang et al. (2014) and Yin et al. (2016b). For example, Niu and Zhou (2013) derived a demand-oriented timetabling approach and formulated a nonlinear 0–1 integer model to minimize the total number of waiting passengers and weighted remaining passengers. An effective passenger loading time window was introduced to deal with the time-dependent waiting time and the number of passengers who fail to board the first coming train due to train capacity constraints. The model was solved by a genetic algorithm (GA) through a special binary coding method. For the non-periodic timetable design with time-variant passenger demands in a metro corridor, Barrena et al. (2014b) developed two non-linear mathematical formulations to minimize the passenger waiting time at the stations. Due to the computational intensity of these two models, they proposed a fast adaptive large neighborhood search (ALNS) algorithm to obtain a good solution. In addition, Niu et al. (2015b) further considered the skip-stop patterns in metro train scheduling problem with time-dependent OD demands, and the problem was then transformed into an algorithmically tractable non-linear mixed integer programming model.

Due to the rising concerns on carbon emission and environmental problems, the energy efficiency of urban metro systems is becoming one of the active research areas in order to pursue a convenient, efficient and environmental friendly transport mode for citizens travelling. Typically, the energy consumption for train operations accounts for nearly a half of the total energy usage in a metro system, and thus it is very promising to effectively reduce the energy consumption by the optimization of train schedules (e.g., Yang et al., 2015b; 2015c; Wang et al., 2015) and operation strategies (e.g., Albrecht et al., 2016; Howlett and Pudney, 1995; Liu and Golovitcher, 2003). More recently, the technology of using regenerative energy, which can be generated by the electric motor when a train is braking and then fed back to the contact lines, has been well applied to some metro systems. Therefore, the main challenge is to coordinate the train operations through the adjustment of train timetables and driving strategies such that the total energy consumption (i.e., the difference between the total traction energy and regenerative energy) is minimized. For example, Ramos et al. (2007) considered a particular case of metro train timetabling during off-peak hours, which aims to maximize the overlapping time between speed-up and slow-down trains in the same power supply substation, such that the regenerative energy can be utilized more effectively. Pena-Alcaraz et al. (2012) developed a mixed-integer nonlinear optimization model to synchronize the braking trains with the accelerating trains to better utilize the regenerative energy. Nasri et al. (2010) proposed a simulation-based train timetable optimization approach in order to completely utilize the regenerative energy from braking trains. Using a GA method, they studied the influence of train headway on the utilization of regenerative energy. Li and Lo (2014) formulated an approach to reduce the total energy consumption by optimizing the timetable and driving strategies (i.e., accelerating, coasting and braking points) by GA. Yang et al. (2015c) utilized the real-world train speed profiles and derived a nonlinear integer model under the assumption of fixed cycle time and headway between adjacent trains.

1.2. Focus of this study

To the best of our knowledge, the majority of existing literature devotes to one of objectives, i.e., energy-efficiency or service quality in the design of passenger-oriented metro train timetables. For comparative convenience, we list the detailed characteristics of some closely related references in Table 1, including decision variables, objectives, formulated models and solution algorithms. For example, Niu and Zhou (2013), Niu et al. (2015b) and Barrena et al. (2014b) addressed the time-variant OD passenger demands in a metro corridor, and aimed to improve the objectives that are only related to the service quality (passenger waiting time, number of waiting passengers, etc). Meanwhile, the other researches (e.g., Li and Lo, 2014; Yang et al., 2015c and Yang et al., 2015b) focus on the minimization of energy consumption for metro train scheduling, in which the passenger demand characteristics are simplified or neglected. More importantly, most of these formulations are nonlinear models that might not be easily handled by the commercial optimization softwares or variants of some classical algorithms. With these concerns, we are particularly interested in developing a unified metro train scheduling framework that synchronously considers these two aspects in order to jointly improve the service quality and operational economy. Specifically, this paper aims to make the following contributions to the study of metro train scheduling problems.

(1) By using a space-time network representation, we transform the metro train scheduling problem into a coordinated routing problem through defining a set of available train speed profiles with different train driving strategies. This method enables to handle both the train traction energy consumption and regenerative energy by linear formulas. Based on the dynamic OD passenger demands, we rigorously formulate two linear programming models. The first model is an integer optimization model, which aims to jointly minimize the train traction energy consumption and passenger waiting time at the stations in a bi-direction metro corridor. The second model, which is formulated as a mixed-integer linear programming (MILP) model, further considers the utilization of regenerative energy by braking trains in the same power supply substation. Compared with existing researches, the proposed approaches synchronously consider the dynamic passenger characteristics, train traction energy consumption and utilization of regenerative energy in a metro train scheduling problem. Since these two models are formulated as an integer linear programming and MILP, respectively, they can be expected to be solved with

Table 1
Recent publications on train scheduling problems in comparison with our work.

| Publication	Decision variable	Objective	Model	Solution algorithm
Barrena et al. (2014b)	Train timetable	Passenger waiting time	Nonlinear	Adaptive large neighborhood search
Barrena et al. (2014a)	Train timetable	Passenger waiting time	Linear	Branch and bound
Canca et al. (2014)	Train timetable	Passenger waiting time	Nonlinear	GAMS software
Niu and Zhou (2013)	Train timetable	Waiting passengers and remaining passengers	Nonlinear	GA
Niu et al. (2015b)	Train timetable	Passenger waiting time	Nonlinear integer programming	GAMS software
Ramos et al. (2007)	Train timetable	Overlapping time	Linear	GAMS software
Pena-Alcaraz et al. (2012)	Train timetable	Overlapping time	Nonlinear	GAMS software
Nasri et al. (2010)	Train timetable	Utilized regenerative energy	Simulation-based	GA
Li and Lo (2014)	Timetable and speed profile	Net energy consumption	Nonlinear	GA
Yang et al. (2015c)	Train dwelling time	Net energy consumption	Nonlinear	GA
Wang et al. (2014)	Train timetable	Traction energy consumption and passenger waiting time	Nonlinear	GA and simulation
This paper	Train timetable and speed profile	Net energy consumption and passenger waiting time	0–1 linear and MILP models	CPLEX and Lagrangian-based heuristic algorithm
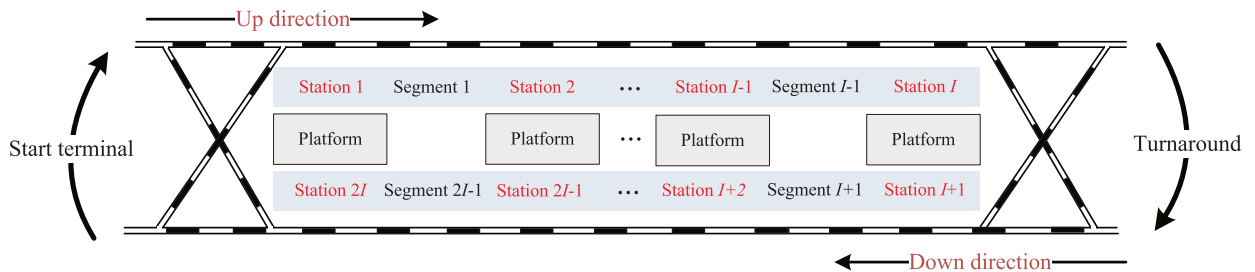
Fig. 1. Bi-directional urban rail line.

the aid of commercial optimization softwares if the problem scale is not too large, and more importantly, they can offer a good benchmark for the researches afterwards in developing heuristic solution algorithms.

(2) Since the complexity of large-scale models makes it difficult to use the integer programming (IP) solvers to search for the optimal train schedules within an acceptable computational time, this research especially designs an efficient Lagrangian relaxation (LR)-based heuristic algorithm to find a good solution as soon as possible. Specifically, we firstly dualize the passenger boarding constraints into the objective function by introducing the corresponding Lagrangian multipliers. Then, the relaxed model is decomposed into two sets of tractable sub-problems. One sub-problem corresponds to a coordinated routing optimization problem in the given space-time network, and the other is essentially a cutting stock problem, which are much easier to be handled. Additionally, we further design a heuristic algorithm to adjust the solution of relaxed models into a feasible solution of the primal problem in order to update the upper bounds. Finally, two sets of numerical experiments, which are respectively based on a simple metro corridor and a real-world instance in Beijing metro, are implemented to illustrate the applications and performance of the proposed approaches.

The rest of this paper is organized as follows. In Section 2, we give a detailed description for the train scheduling problem in a bi-direction metro corridor. In Section 3, two mathematical models are formulated to generate the train scheduling plan and speed profiles in each segment. Then, we propose the LR-based heuristic algorithm to solve the developed models in Section 4. In Section 5, we design two case studies, i.e., an illustrative small case and a large-scale case based on the real-world operation data of Beijing metro, to demonstrate the effectiveness of the proposed approaches. Finally, some conclusions and further studies are presented in Section 6.

2. Problem description

This paper considers the train scheduling problem for a bi-direction urban rail line with $2I$ stations. As shown in Fig. 1, the stations are numbered as 1, 2, $\cdots$, $2I$. The start terminal and return terminal are indexed by station 1 and station I, respectively. Let $\mathcal{I} = \{1, 2, \cdots, 2I\}$ be the set of stations. A group of service trains, denoted by $\mathcal{K} = \{1, 2, \cdots, K\}$, begin their journeys at station 1 in up-direction until reaching station I. After a given turn-around time at the turnaround terminal, they go back to the origin station (i.e., station $2I$) and then wait for the next cycle.
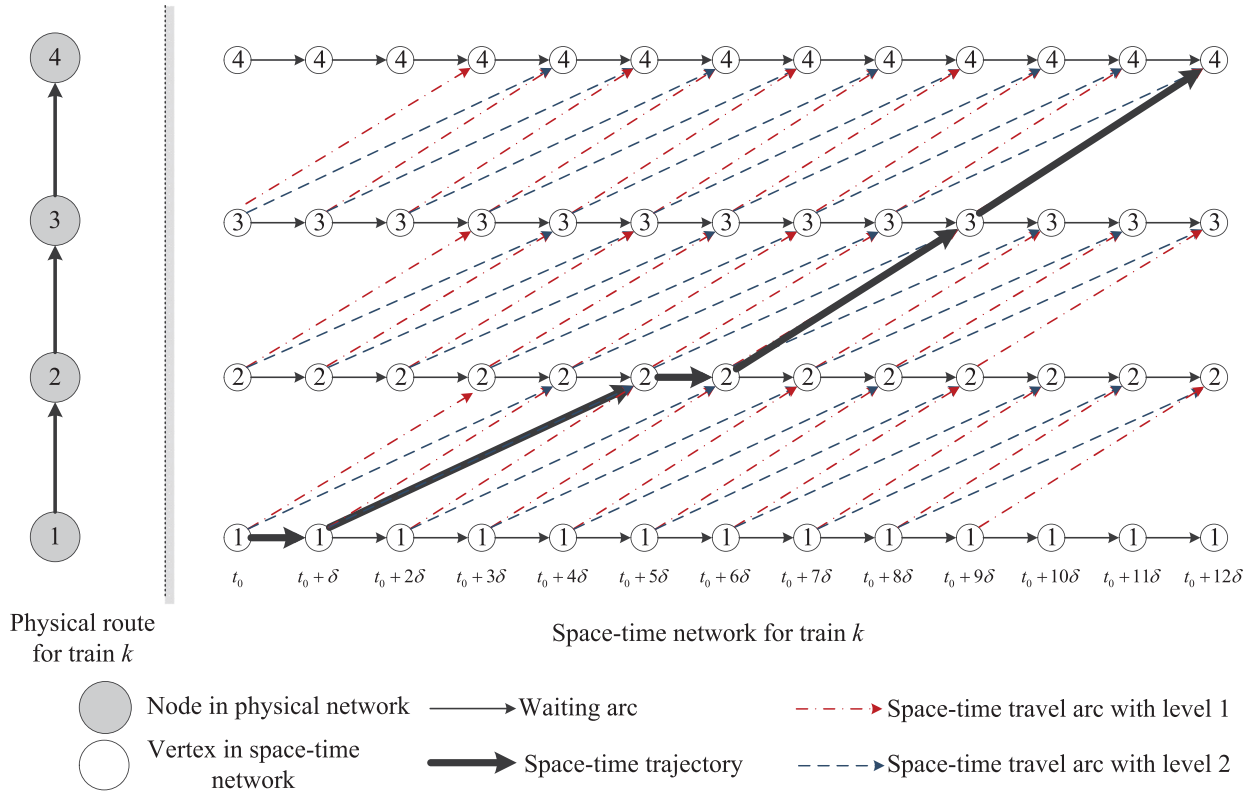
Fig. 2. Space-time network for representation of train trajectory.

Then, the purpose of metro train scheduling problem in this paper is to find an optimal schedule for the considered trains in the planning time horizon $[t_0, t_{end}]$, such that a trade-off between the operational costs and passenger waiting time can be optimized. One practical issue for the metro train scheduling is the consideration of dynamic passenger demands. For example, the passenger demands may vary in different time periods throughout a single-day time horizon, e.g., high demands in peak hours and low demands in off-peak hours. Meanwhile, since the amount of green house emissions is directly related to the energy consumption of trains, it is suggested that there are opportunities for reducing CO_2 emissions with proper train scheduling strategies to account for wider environmental and social impacts (e.g., Bektas and Laporte, 2011; Zhou et al., 2017). Therefore, it is practically significant to obtain a train timetable in order to realize a good balance between these two aspects.

• **Space-time network representation**

In what follows, we first introduce the space-time networks for representation of train scheduling problem, and then we develop two linear formulations for solving this problem with the consideration of both dynamic passenger demands and train energy consumption. Space-time networks, which aim to integrate physical transportation networks with vehicle time-dependent movements/trajectories, are effective representation methods for capturing and analyzing spatial and temporal characteristics of dynamic traffic systems (e.g., see Kennington and Nicholson, 2010; Li et al., 2015; Tong et al., 2015; Shi et al., 2014; Boland et al., 2015; Yang et al., 2014; Mahmoudi and Zhou, 2016; Liu and Zhou, 2016; Lu et al., 2016; Yang and Zhou, 2017). For instance, Yang and Zhou (2014) constructed a space-time network to capture spatial and temporal travel time correlations for finding a *priori* least expected time path in a time-dependent and stochastic network. Kennington and Nicholson (2010) proposed a novel strategy through a tight linear programming (LP) relaxation for space-time fixed-charge network flow problems to determine the artificial arc capacities. By using a space-time network, Tong et al. (2015) formulated the urban network design problem into a linear integer programming model that maximizes the system-wide transportation accessibility. More recently, Mahmoudi and Zhou (2016) solved the vehicle routing problem with pickup and delivery with time windows effectively based on a state-space-time network and dynamic programming algorithm embedded into a Lagrangian relaxation framework. Since a train timetable in essence determines space-time positions of in-service trains, we can explicitly express the trajectories of trains in a space-time network.

To construct a space-time network, the considered time horizon $[t_0, t_{end}]$ is first divided into a series of intervals with the same time length δ, denoted by a set of timestamps $\mathcal{T} = \{t_0, t_0 + \delta, t_0 + 2\delta, \cdots, t_0 + M\delta\}$, at which some important operations, e.g., arriving at stations and departing from stations, are allowable to carry out (see Fig. 2). Theoretically, a space-time network is the combination of the physical network and time horizon. As shown in Fig. 2, each node in the space-time
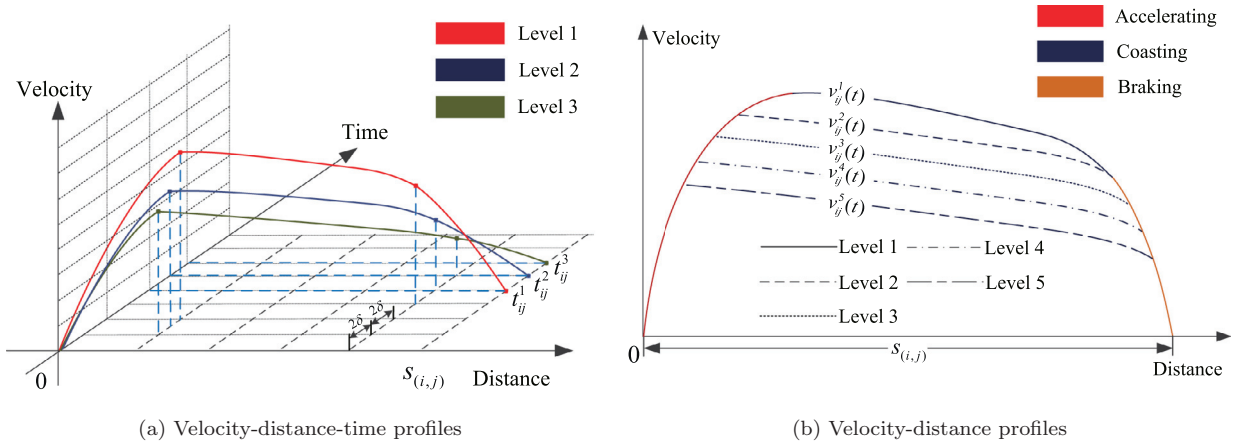
(a) Velocity-distance-time profiles   (b) Velocity-distance profiles

Fig. 3. An illustration of train operation levels and the corresponding velocity profiles.

network denotes the spatial and temporal state of a physical node, and a space-time arc essentially denotes the entering and departure times of a physical link for a given train. Once the time horizon is pre-given, we can easily deduce the set of space-time nodes in the entire space-time network. However, in determining the space-time arcs, we need to specify all the velocity profiles on each travel link, which correspond to different link travel times and space-time arcs. For clarity, we next introduce the concept of velocity profiles considered in this study.

- **Definition of velocity profiles on each travel link**

In many previous studies (e.g., Barrena et al., 2014b; Yang et al., 2015c; 2014; Niu and Zhou, 2013) for train scheduling, the train travelling velocity in each segment is usually considered as a predetermined value. Nevertheless, in practical operations, the velocity and travel time of a metro train on travel links are always not constant, but can be adjusted according to the operation level of on-board train control systems. In general, the operation levels are discrete speed levels (e.g., level 1 to level 5, see Fig. 3b) and preseted by metro signal suppliers. The aim is to provide more choices for metro dispatchers to improve the flexibility and efficiency of scheduling (or rescheduling) in daily operations. In particular, the widely used automatic train operation (ATO) systems in metro lines are commonly embedded with this important function (IEEE Standard). We consider the operation levels on each travel link (i, j), which are defined by a set of realistic train speed profiles, i.e.,

$$v_{ij}^l(t) \in \left\{ v_{ij}^1(t), v_{ij}^2(t), \cdots, v_{ij}^L(t) | \forall 1 \leq l \leq L, \int_{t=0}^{t_{ij}^l} v_{ij}^l(t)dt \equiv s_{(i,j)} \right\},$$

in which t represents the time index after train k departs from station i (here, we set departure time as 0), $s_{(i, j)}$ is the length of link (i, j), $l \in \{1, 2, \cdots, L\}$ indicates the operation level of each train, and t_{ij}^l refers to the preseted travel time of train k on link (i, j) corresponding to operation level l. Here, we note that the time interval in the space-time network is assumed to be short enough, so that all the departure times and arrival times in different operation levels for the trains can be represented by space-time nodes in this network. Moreover, if we consider the entering time t on link (i, j) (i.e., the departure time from station i), we then can use a unified representation $v_{ijtt'}^k(y)$ to denote different speed profiles for each train k, where $v_{ijtt'}^k(y) = v_{ij}^l(y - t)$ with $t' = t + t_{ij}^l$ and $y \in [t, t']$. According to the theoretical train motion equations (see Howlett and Pudney, 1995; Albrecht et al., 2016), the train movement in each segment typically contains acceleration, braking, coasting and cruising. In this study, considering the entering time $t = 0$, we use four sets, i.e., T_{ij}^{la}, T_{ij}^{lb}, T_{ij}^{lc} and T_{ij}^{lr}, each of which is a series of discrete timestamps, to respectively denote the train acceleration, braking, coasting and cruising timestamps on each physical link (i, j) and operation level l. Furthermore, if we consider the entering time t, the timestamp of each operation state should be added by the entering time t. That is, we can use $t + T_{ij}^{la}$, $t + T_{ij}^{lb}$, $t + T_{ij}^{lc}$, $t + T_{ij}^{lr}$ to represent the timestamp sets of train operation states (i.e., acceleration, braking, coasting and cruising) on the link (i, j) with entering time t and operation level l.

Next, we explicitly show the time-position-velocity profiles for the train operation levels in Fig. 3a (we only illustrate three curves for simplicity). A train departs from station i at time $t = 0$, uses one of these profiles to drive for the next station j, and then arrives at any timestamp $t_{ij}^l = Q_{ij}^l \cdot \delta$, where Q_{ij}^l is a link-and-operation-level-dependent integer, i.e., $Q_{ij}^l \in \mathbb{N}^+$. Additionally, since each time interval corresponds to a given train operation state (acceleration, braking, coasting or cruising), we can easily derive from this figure that, if the train employs operation level 1, the timestamp sets for train operation states are $T_{ij}^{1a} = \{0, \delta, 2\delta, 3\delta\}$, $T_{ij}^{1c} = \{4\delta, 5\delta\}$ and $T_{ij}^{1b} = \{6\delta, 7\delta\}$, respectively. In particular, each timestamp in the above sets actually represents an operation time interval in the discretized time horizon (for instance, timestamp δ represents that the train should accelerate in time interval $[\delta, 2\delta)$).

Next, we further give an illustration to clarify the concept of operation levels in practical metro train control systems. In practice, different operation levels will necessarily correspond to different operation modes. For instance, in the Chengdu metro system, a total of five operation levels are predetermined for train operations in each travel segment, and each operation level practically corresponds to a specific link-dependent speed profile. As illustrated in Fig. 3b, each higher operation level (e.g., level 5 versus level 4) corresponds to a unique velocity profile (i.e., the projection of the time-position-velocity profile on the plane of coordinates position and velocity) that increases the link travel time by about $5 - 10$ s, and in contrast, the velocity of each lower level reduces the link travel time by about $5 - 10$ s (Hua and Lin, 2014). In daily operations, the dispatchers can use the lower operation levels to reduce the train travel times in peak hours, or reschedule the current timetable in case of disturbances, which can effectively release the transport pressure and improve the service quality of metro systems. Meanwhile, the travel speed of trains in off-peak hours can be scheduled to be relatively slow with higher operation levels in order to save energy for train operations.

- **Illustration of space-time network formulation for metro train scheduling**

We use Fig. 2 to demonstrate the space-time network formulation with multiple operation levels, in which the physical route of train k consists of four nodes and three links. A total of 13 timestamps, i.e., t_0, $t_0 + \delta$, $\cdots$, $t_0 + 12\delta$, are considered to represent the whole time horizon. In this space-time network, each space-time node represents the state of a physical node at the current timestamp. In addition, two types of velocity profiles, which correspond to two types of space-time arcs, are considered for generating trajectory of train k. Operation levels 1 and 2 subject to the arc travel times 3δ and 4δ in each segment, respectively. Particularly, we define the waiting arcs to imply the dwelling times of trains, and each waiting arc indicates that the train is scheduled to dwell at this station for time length δ. Here, we give a space-time trajectory of train k from origin to destination to demonstrate the schedule process. As can be seen in this figure, train k first dwells at station 1 at timestamp t_0, and departs from this station at timestamp $t_0 + \delta$. Travelling with operation level 2, train k arrives at station 2 at timestamp $t_0 + 5\delta$. After one time interval dwelling, train k departs from station 2 by using operation level 1. Likewise, the train finally reaches the destination at timestamp $t_0 + 12\delta$. In this sense, the train scheduling problem can be transformed into a coordinated routing problem with the consideration of multiple trains in this space-time network. In order to formulate mathematical models for metro train scheduling problem by using the space-time network representation, we first make the following assumptions throughout this paper:

Assumption 1. First, we assume that the provided total train capacity is greater than or equal to the total passenger demand, and the last train stops at every station and departs from the first station at the exact ending time T of the study period. This assumption is to ensure that all passengers arrive at their destination stations during the planning horizon.

Assumption 2. At each station, we further assume that the oversaturation situation is not permitted, and each train should use its capacity c_j to accommodate all waiting passengers. If oversaturation is allowed at stations, then we refer readers to our previous paper (Niu and Zhou, 2013) for various models on calculating the effective loading time periods.

Assumption 3. In this study, we shall not track the detailed number of passengers getting on/off each train, but consider the estimated loading capacity. It is particularly required that the loading capacity of each train stopping at each station is larger than the aggregated total passenger demands during the departure of two successive trains.

Assumption 4. Finally, we assume that the considered metro trains only consume energy in acceleration and cruising phases. This is a practical case in metro systems that has been widely adopted in many previous studies (e.g., Yang et al., 2015b; Pena-Alcaraz et al., 2012; Wang et al., 2014).

3. Mathematical formulation

In this section, we rigorously formulate two mathematical models in order to solve the metro train scheduling problem. The following discussion will focus on specifying each part of the models, including parameters, decision variables, objective functions and systematic constraints.

3.1. Notations

3.1.1. Symbols and parameters

$\mathcal{I} = \{1, 2, \ldots, 2I\}$	Set of involved metro stations
$\mathcal{K} = \{1, 2, \ldots, K\}$	Set of in-service trains to be scheduled
$\mathcal{U} = \{1, 2, \ldots, U\}$	Set of power supply substations
$\mathcal{L} = \{1, 2, \cdots, L\}$	Set of operation levels
$\mathcal{T} = \{t_0, \cdots, t_0 + M\delta\}$	Set of timestamps in the planning horizon
$\mathcal{A}$	Set of space-time arcs
$\mathcal{S}$	Set of involved links

i, j	Index of metro stations, $i, j \in \mathcal{I}$
(i, j)	Index of directed links, $(i, j) \in \mathcal{S}$
t, t', τ, τ', y	Time index
(i, i, t, t')	Index of space-time waiting arcs indicating waiting time at station i, $t' = t + \delta$
(i, j, t, t')	Index of space-time arcs indicating the actual movement at entering time t and leaving time t' on link (i, j), $(i, j, t, t') \in \mathcal{A}$, determined by different profiles
k	Service index of in-service trains, $k \in \mathcal{K}$
u	Index of power supply substations, $u \in \mathcal{U}$
l	Operation level index, $l \in \mathcal{L}$
K_p	Fleet size of interest
O	Dummy origin station
D	Dummy destination station
$s_{(i, j)}$	Length of link (i, j)
C_k	Loading capacity of train k
m_k	Mass of train k
$p_{ij}(t)$	Time-variant demand ratio from stations i to j in time interval $(t - \delta, t]$
$b^i_{\max}$	Allowable passenger boarding speed at station i
$\overline{d_i}$	Maximum dwelling time at station i
$\underline{d_i}$	Minimum dwelling time at station i
t^l_{ij}	Travel time on link (i, j) with operation level l
$v^k_{ijtt'}(y)$	Speed profile for train k on space-time arc (i, j, t, t'), $y \in [t, t']$
$F^k_{ijtt'}(y)$	Traction force for train k on space-time arc (i, j, t, t'), $y \in [t, t']$ (per unit mass)
$B^k_{ijtt'}(y)$	Braking force for train k on space-time arc (i, j, t, t'), $y \in [t, t']$ (per unit mass)
T^{al}_{ij}	Set of timestamps for train acceleration on link (i, j) with operation level l
T^{bl}_{ij}	Set of timestamps for train braking on link (i, j) with operation level l
T^{cl}_{ij}	Set of timestamps for train coasting on link (i, j) with operation level l
T^{rl}_{ij}	Set of timestamps for train cruising on link (i, j) with operation level l
c_r	Coefficient from kinetic energy to regenerative energy
c_a	Available energy percentage after transmission
$\theta_a, \theta_b, \theta_c$	Davis parameters of friction forces
$\mathbb{N}^+$	Set of nonnegative integer variables

3.1.2. Decision variables

In this study, we consider three types of decision variables in scheduling trains with the time-variant passenger demands, which are detailed as follows.

$x^k_{ijtt'}$	Selection indicator for space-time arc, $=1$ if train k uses link (i, j) in time interval $[t, t']$; $=0$, otherwise (*binary variable*)
b^i_t	Number of boarding passengers at station i in time interval $(t - \delta, t]$ (*nonnegative integer variable*)
n^i_t	Number of waiting passengers at station i and time t (*nonnegative integer variable*)

In general, decision variable $x^k_{ijtt'}$ is employed to generate a space-time trajectory for each train k. The other two variables, i.e., b^i_t and n^i_t are introduced to represent the time-variant numbers of boarding passengers and waiting passengers at each station, which can be used as the passenger flow control variables in the scheduling process (A similar definition is also adopted by Barrena et al., 2014a). Typically, these two sets of variables are intermediate variables associated with $x^k_{ijtt'}$ in the given space-time network representation.

Remark 3.1. After the arrival of a physical train at the terminus (i.e., station 1), its service index will be augmented when this train departs. For example, a metro line has K_p physical trains in total (namely the fleet size), which are numbered as trains 1, 2, $\cdots$, K_p. For each physical train p with $1 \leq p \leq K_p$, the service index of train p is changed to $p + K_p$ when it finishes the first cycle and prepares to start the next cycle. Thus, the service indexes of trains can be successively numbered by

$$\underbrace{\overbrace{1, 2, \cdots, K_p}^{\text{Cycle 1}}, \overbrace{K_p + 1, \cdots, 2K_p}^{\text{Cycle 2}}, \cdots, \overbrace{(N_c - 1)K_p + 1, \cdots, N_cK_p}^{\text{Cycle } N_c}, N_cK_p + 1, \cdots, N_cK_p + N_p}_{\text{The total number is } |\mathcal{K}|}$$
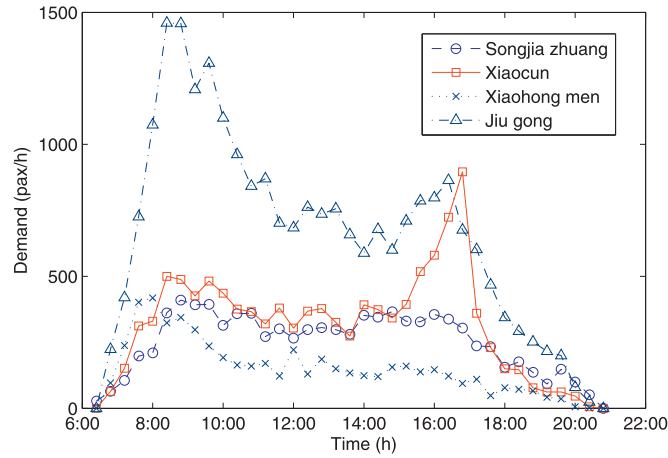
Fig. 4. Time-variant passenger demands at stations of Yizhuang line through a single day.

in which N_c and N_p are nonnegative integer values, and $N_p \leq K_p$. For simplicity, we use set $\mathcal{K}$ to represent the set of service indexes of trains in this paper.

As mentioned above, the aim of metro train scheduling process is to develop a balanced train timetable between service quality and operation cost with the limitation of metro infrastructures. In this study, the service quality refers to the total passenger waiting time at stations for boarding the on-coming trains, and the operation cost typically refers to the energy consumption for the travelling of trains. Note that, there are strong intrinsic relationships between these two indicators. For instance, the passenger waiting time can be reduced by increasing the train departure frequency or the train travel speed in each link. Meanwhile, the reduction of passenger waiting time may potentially increase the energy consumption, i.e., the operation costs of metro companies. Therefore, we explicitly analyze these two objectives in the train scheduling problem by using the space-time network representation, and rigorously formulate two linear optimization models in order to generate a service-oriented and energy-efficient metro train scheduling plan.

3.2. Objective function

This part analyzes two objectives in the problem of interest, i.e., passenger waiting time and train energy consumption. In the calculation of passenger waiting time, we particularly consider the dynamic passenger demands, which is an important feature of metro train scheduling process that is different from railway (mainline railway, high-speed railway, etc) scheduling problems (Barrena et al., 2014b; Yang et al., 2015a). Besides, we are interested in formulating the objectives into linear formulas that could make the model easier to be handled by some commercial solvers (e.g., CPLEX, Gurobi) or the variants of some classical algorithms (e.g., branch and bound, Lagrangian relaxation), and also provide a good benchmark for subsequent heuristic solution algorithms.

• **Passenger waiting time**

Due to the wide application of smart card technologies, the time-variant origin-destination (OD) passenger travel data become available in many new urban metro systems, which provide an effective method for us to analyze the dynamic passenger demand in a daylong time (see Niu et al., 2015a; Sun et al., 2014; Pelletier et al., 2011). For example, Fig. 4 demonstrates the variation of passenger arrival rates recorded at four stations, namely Songjia zhuang station, Xiaocun station, Xiaohong men station and Jiu gong station, in Beijing Yizhuang metro line on a weekend in Oct. 2014.

In essence, the dynamic passenger demand at every station can be expressed as a time-variant square matrix $\mathbf{p}(t)$, which has as many rows and columns as stations in the network. Each element in this matrix denoted by $p_{ij}(t)$ indicates the passenger demand ratio from station i to j in time interval $(t - \delta, t]$ for any $t \in \mathcal{T} \setminus \{t_0\}$. In particular, the initial value of $\mathbf{p}(t)$, i.e., $p_{ij}(t_0)$ is set as zero for any $i, j \in \mathcal{I}$. Since this study considers a bi-direction metro line, the passengers who board the train at an up-direction station i for $1 \leq i < I$ need to alight before the train's turnaround. Likewise, the passengers with down-direction travel will alight the train before its turnaround, which can be summarized by

$$p_{ij}(t) : \begin{cases} \geq 0, & \text{if } (1 \leq i \leq I, 1 \leq j \leq I, i < j) \cup (I < i \leq 2I, I < j \leq 2I, i < j) \\ = 0, & \text{otherwise.} \end{cases}$$

To model the passenger boarding process, we derive two additional decision variables (nonnegative integer variables), i.e., n_t^i and b_t^i, which represent the numbers of waiting passengers and boarding passengers at station i and the end of time interval $(t - \delta, t]$, respectively. The initial value of $n_{t_0}^i$ is set to be a constant value and $b_{t_0}^i$ is set to be zero. Then, as adopted in literature (Barrena et al., 2014a), the objective function with respect to passenger waiting time in planning time horizon

$\mathcal{T}$ can be given by

$$WT = \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \delta \cdot n_t^i, \tag{1}$$

where

$$n_t^i = n_{t-\delta}^i + \sum_{i \le j \le 2I} p_{ij}(t) \cdot \delta - b_t^i, \quad \forall t \in \mathcal{T} \setminus \{t_0\}, \quad i \in \mathcal{I}, \tag{2}$$

$$b_t^i \le \sum_{k \in \mathcal{K}} x_{iit't}^k b_{max}^i, \quad \forall t \in \mathcal{T} \setminus \{t_0\}, \ t' = t - \delta, \quad i \in \mathcal{I}, \tag{3}$$

and

$$n_t^i, \ b_t^i \ge 0, \quad \forall t \in \mathcal{T}, \quad i \in \mathcal{I}. \tag{4}$$

In the above formulations, Eq. (2) indicates that, the number of waiting passengers at timestamp t is equal to the number of waiting passengers at timestamp $t - \delta$ plus the arrival passengers and then subtract the number of boarding passengers in this time interval. In addition, Eq. (3) restricts the number of boarding passengers in one time interval, where b_{max}^i is the maximum allowable boarding passengers in one time interval at station i. Specifically, if there is a train $k \in \mathcal{K}$ that is dwelling at station i in time interval $[t - \delta, t]$, we have $0 \le b_t^i \le b_{max}^i$; otherwise, we have $b_t^i = 0$ since there is no train dwelling at station i in this time interval, and thus no passengers can board. For simplification, the maximum allowable boarding speed b_{max}^i is treated as a constant value in this study, i.e., it is not affected by the train remaining capacity or number of waiting passengers. We can also refer to Wang et al. (2014) and Lin and Wilson (1992) for more practical explanations on this point. Besides, to represent the time-variant passenger demands at station i, we define

$$d_t^i = \sum_{i \le j \le 2I} p_{ij}(t) \cdot \delta, \quad \forall t \in \mathcal{T}, \quad i \in \mathcal{I}$$

which indicates the number of arriving passengers at station i and timestamp t.

Proposition 1. *The objective function of passenger waiting time in Eq. (1) is essentially equivalent to the following formulation:*

$$WT = \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \delta \cdot (n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i)), \tag{5}$$

with the constraint of $\sum_{\bar{t} \in [t_0, t]} b_{\bar{t}}^i \le n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} d_{\bar{t}}^i$ for $\forall t \in \mathcal{T}$ and $i \in \mathcal{I}$. (Here, $[t_0, t]$ is an abbreviation for set $\{t_0, t_0 + \delta, t_0 + 2\delta, \cdots, t\}$.)

Proof. In essence, this proof aims to find an equivalent relationship between decision variables n_t^i and b_t^i. We note that, Eq. (2) holds for any $t \in \mathcal{T} \setminus \{t_0\}$ and $i \in \mathcal{I}$. We can list a sequence of equations for any $t \in \mathcal{T} \setminus \{t_0\}$, i.e.,

$$n_{t_0 + \delta}^i = n_{t_0}^i + d_{t_0 + \delta}^i - b_{t_0 + \delta}^i, \tag{6}$$

$$n_{t_0 + 2\delta}^i = n_{t_0 + \delta}^i + d_{t_0 + 2\delta}^i - b_{t_0 + 2\delta}^i, \tag{7}$$

$$\cdots \cdots \tag{8}$$

$$n_t^i = n_{t-\delta}^i + d_t^i - b_t^i. \tag{9}$$

By summing up the sequence of equations, we can get

$$n_t^i = n_{t_0}^i + \sum_{\bar{t} \in (t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i) \tag{10}$$

for each $t \in \mathcal{T} \setminus \{t_0\}$ and $i \in \mathcal{I}$. Since the values of $d_{t_0}^i$ and $b_{t_0}^i$ are set as zeros, Eq. (10) can be further represented by $n_t^i = n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i)$ for each $t \in \mathcal{T}$ and $i \in \mathcal{I}$. In this sense, it is clear that, the decision variable n_t^i is essentially an intermediate decision variable that can be deduced by b_t^i. Note that, since n_t^i is defined to be a nonnegative variable, i.e., $n_t^i \ge 0$, we also have the following constraint for b_t^i, i.e.,

$$\sum_{\bar{t} \in [t_0, t]} b_{\bar{t}}^i \le n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} d_{\bar{t}}^i, \quad \forall t \in \mathcal{T}, i \in \mathcal{I}. \tag{11}$$

Then, taking Eq. (10) into the original Eq. (1), we prove this proposition. □

Remark 3.2. In the above proof, Eqs. (6)–(11) are essentially the derivation of cumulative flow variables that were proposed by Daganzo (1997). We can refer to Chapter 2 in this book for a more detailed description.

• **Energy consumption**

Energy consumption is an important indicator to evaluate the quality of a train schedule in public transportation systems, since energy consumption is closely related to environmental concerns and profits of management companies (Daganzo, 2010). Here, we consider the formulation of total energy consumption for the in-service trains in the whole metro corridor, involving the traction energy consumption and regenerative braking energy. For formulating convenience, we first use TE_{ij}^k and TD_{ij}^k to represent the entering time and leaving time of train k on link (i, j) by the selected space-time arcs, given by

$$TE_{ij}^k = \sum_{(i,j,t,t') \in \mathcal{A}} t \cdot x_{ijtt'}^k, \quad \forall (i, j) \in \mathcal{S}, k \in \mathcal{K}, \tag{12}$$

$$TD_{ij}^k = \sum_{(i,j,t,t') \in \mathcal{A}} t' \cdot x_{ijtt'}^k, \quad \forall (i, j) \in \mathcal{S}, k \in \mathcal{K}. \tag{13}$$

In other words, TE_{ij}^k and TD_{ij}^k represent the departure time at station i and arrival time at station j for train k, respectively.

Then, we analyze the link traction energy consumption based on the link travel speed profile of a train. Given the time-dependent travelling arc $(i, j, t, t') \in \mathcal{A}$ of train $k \in \mathcal{K}$, the corresponding operation level l^* can be uniquely determined by

$$l^* = \operatorname{argmin}_l |t_{ij}^l - t' + t|. \tag{14}$$

For example, consider a travelling arc $(1, 2, 0, 5)$, indicating that the train departs from station 1 at timestamp 0 and arrives at station 2 at timestamp 5. There are totally three operation levels, i.e., $l = 1, 2, 3$, which correspond to arc travel times 3, 4 and 5, respectively. In this case, we can obviously derive from Eq. (14) that the train travels with operation level $l^* = 3$. Using this formulation, the link travel time of arc (i, j, t, t') is expressed by $t_{ij}^{l^*}$, which corresponds to accelerating timestamp set $t + T_{ij}^{l^*a}$, coasting timestamp set $t + T_{ij}^{l^*c}$, cruising timestamp set $t + T_{ij}^{l^*r}$ and braking timestamp set $t + T_{ij}^{l^*b}$.

Next, we analyze the energy consumption on each space-time travelling arc. Consider the space-time arc (i, j, t, t'), where the operation level is denoted by l^*. Then, the corresponding speed profile is given by $v_{ijtt'}^k(y)$ for $y \in [t, t + t_{ij}^{l^*}]$, indicating train k's speed with respect to each specific time point y on this arc. According to the mechanical power equation of train movement described in Wang et al. (2014) and Howlett and Pudney (1995), the traction and braking force (per unit mass) of train k can be denoted by

$$F_{ijtt'}^k(y) = \begin{cases} a_{ijk}^{l^*}(y) + \frac{1}{m_k}(\theta_a(v_{ijtt'}^k(y))^2 + \theta_b v_{ijtt'}^k(y) + \theta_c), & \text{if } y \in t + T_{ij}^{l^*a} \text{ or } y \in t + T_{ij}^{l^*r} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

$$B_{ijtt'}^k(y) = \begin{cases} -a_{ijk}^{l^*}(y) - \frac{1}{m_k}(\theta_a(v_{ijtt'}^k(y))^2 + \theta_b v_{ijtt'}^k(y) + \theta_c), & \text{if } y \in t + T_{ij}^{l^*b} \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

in which $a_{ijk}^{l^*}(y)$ is the acceleration rate at time y that can be derived by the corresponding speed profile on arc (i, j, t, t'), and $\theta_a(v_{ijtt'}^k(y))^2 + \theta_b v_{ijtt'}^k(y) + \theta_c$ is the Davis formula that represents the friction resistances. Particularly, since the train speed is constant in train cruising phase, i.e., $a_{ijk}^{l^*}(y) = 0$, $F_{ijtt'}^k(y)$ in Eq. (15) can essentially indicate the traction force in both accelerating and cruising phases. Besides, in train coasting phase there is no traction or braking force, and hence no energy consumption occurs. Then, according to the mechanical power equation, the actual energy consumption on this space-time arc can be calculated by the train mass times the integral of power over the operation time interval. Therefore, the required energy of train k on travelling arc (i, j, t, t') can be formulated as

$$E_k(i, j, t, t') = \int_t^{t+t_{ij}^{l^*}} m_k F_{ijtt'}^k(y) v_{ijtt'}^k(y) dy, \quad \forall (i, j, t, t') \in \mathcal{A}, k \in \mathcal{K}. \tag{17}$$

3.3. Model 1: energy-efficient metro train scheduling with dynamic passenger demands

Considering the above two objective functions, the energy-efficient metro train scheduling model with dynamic passenger demands (denoted by ETS model) can be formulated as follows:

$$(\text{ETS}) \quad \min_{\mathbf{x}, \mathbf{b}} Z_1(\mathbf{x}, \mathbf{b}) = w_T \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \delta \cdot (n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i)) + w_E \sum_{k \in \mathcal{K}} \sum_{(i,j,t,t') \in \mathcal{A}} E_k(i, j, t, t') x_{ijtt'}^k \tag{18}$$

$$\text{s.t.} \quad \sum_{(i,j,t,t') \in \mathcal{A}} x_{ijtt'}^k - \sum_{(j,i,t',t) \in \mathcal{A}} x_{jit't}^k = \begin{cases} 1 & i = O, t = t_0, \\ -1 & i = D, t = t_0 + M\delta, \quad \forall k \in \mathcal{K} \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

$$TE_{2I,D}^k + TR_k \leq TD_{O1}^{k+K_p}, \quad \forall k, k + K_p \in \mathcal{K} \tag{20}$$

$$TD^k_{I(I+1)} - TE^k_{I(I+1)} = TA_k, \qquad \forall k \in \mathcal{K} \tag{21}$$

$$TE^{k-1}_{ij} + h_{ij} \le TE^k_{ij}, \qquad \forall (i,j) \in \mathcal{S}, k \in \mathcal{K} \backslash \{1\} \tag{22}$$

$$TD^{k-1}_{ij} + h_{ij} \le TD^k_{ij}, \qquad \forall (i,j) \in \mathcal{S}, k \in \mathcal{K} \backslash \{1\} \tag{23}$$

$$TD^k_{j'i} + \underline{d}_i \le TE^k_{ij}, \qquad \forall (i,j) \in \mathcal{S}, (j',i) \in \mathcal{S}, k \in \mathcal{K} \tag{24}$$

$$TE^k_{ij} - \overline{d}_i \le TD^k_{j'i}, \qquad \forall (i,j) \in \mathcal{S}, (j',i) \in \mathcal{S}, k \in \mathcal{K} \tag{25}$$

$$TE^k_{ij} = \sum_{(i,j,t,t') \in \mathcal{A}} t \cdot x^k_{ijtt'}, \qquad \forall (i,j) \in \mathcal{S}, k \in \mathcal{K} \tag{26}$$

$$TD^k_{ij} = \sum_{(i,j,t,t') \in \mathcal{A}} t' \cdot x^k_{ijtt'}, \qquad \forall (i,j) \in \mathcal{S}, k \in \mathcal{K} \tag{27}$$

$$b^i_t \le \sum_{k \in \mathcal{K}} x^k_{iit't} b^i_{max}, \qquad \forall i \in \mathcal{I}, t \in \mathcal{T} \backslash \{t_0\}, t' = t - \delta \tag{28}$$

$$\sum_{\bar{t} \in [t_0,t]} b^i_{\bar{t}} \le n^i_{t_0} + \sum_{\bar{t} \in [t_0,t]} d^i_{\bar{t}}, \qquad \forall i \in \mathcal{I}, t \in \mathcal{T} \tag{29}$$

$$\sum_{i' \in [1,i]} \sum_{t \in \mathcal{T}} \left[\sum_{j>i}^{2I} \delta \cdot p_{i'j}(t+\delta) \sum_{\bar{t} \in [t_0,t]} \sum_{t' \in \mathcal{T}} (x^{k-1}_{i'(i'+1)\bar{t}t'} - x^k_{i'(i'+1)\bar{t}t'}) \right] \le C_k,$$
$$\forall i \in \mathcal{I}, k \in \mathcal{K} \backslash \{1\} \tag{30}$$

$$x^k_{ijtt'} \in \{0,1\}, \qquad \forall (i,j,t,t') \in \mathcal{A}, k \in \mathcal{K} \tag{31}$$

$$b^i_t \in \mathbb{N}^+, \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}. \tag{32}$$

In this model, the objective function (18) aims to determine the optimal train scheduling plan and speed profile in each segment such that the weighted total passenger waiting time and train energy consumption can be minimized within the given time horizon. Note that, to obtain a trade-off solution, we give two weight factors, i.e., w_T and w_E, to respectively indicate the importance of each objective. In this sense, we actually transform the passenger waiting time and energy consumption into their generalized costs with the same unit scales in the objective function. Constraint (19) is the space-time flow balance constraint, which ensures to generate a feasible space-time path for each train in the space-time network (see in Mahmoudi and Zhou, 2016; Yang and Zhou, 2014). In order to represent the beginning and ending of the generated space-time paths, a dummy origin and a dummy destination, denoted by O_{t_0} and $D_{t_0+M\delta}$, respectively, are added into the space-time network. Moreover, to show the connection between the dummy origin and physical origin, a total of $M+1$ dummy time-dependent arcs are also added into the space-time network without actual link travel times, denoted by $(O,1,t_0,t_0+m\delta), 0 \le m \le M$. In the scheduling process, if an arc $(O,1,t_0,t_0+m\delta)$ is selected by a train, it indicates that the actual arrival time of this train at station 1 is $t_0+m\delta$. Likewise, we also add $M+1$ dummy time-dependent arcs into the space-time network to show the connection between the physical destination and dummy destination without actual link travel times, denoted by $(2I,D,t_0+m\delta,t_0+M\delta), 0 \le m \le M$. If an arc $(2I,D,t_0+m\delta,t_0+M\delta)$ is selected, it shows that the actual departure time of this train at station $2I$ is $t_0+m\delta$. For clarity, we illustrate the travel routes of five trains in a space-time network with 3 physical stations and 13 timestamps. As shown in Fig. 5, by using the dummy arcs, all the trains start their journeys from the dummy origin and finally arrive at the dummy destination.

Constraint (20) defines the minimum turnaround time TR_k for train k at the terminus to prepare for the next cycle. Constraint (21) fixes the turnaround time TA_k for each train $k \in \mathcal{K}$ at the turnaround station I. Constraints (22) and (23) are the headway constraints that guarantee the least time distance h_{ij} between two adjacent trains when they enter or leave the common link (i,j). Constraints (24) and (25) set the lower and upper bounds for dwelling times of the trains at each station, respectively. Constraints (26) and (27) define the entering time TE^k_{ij} and departure time TD^k_{ij} of train k on link (i,j) in the space-time network. Constraint (28) links the decisions variables **x** and **b**. Specifically, if there is no train dwelling at station i at time t, we have $b^i_t = 0$; otherwise, the number of boarding passengers b^i_t in one time interval should be bounded by an allowable passenger boarding speed, i.e., b^i_{max}. Constraint (29) is derived in Proposition 1, indicating that the number of waiting passengers at any time and station should be nonnegative. Constraint (30) is the train
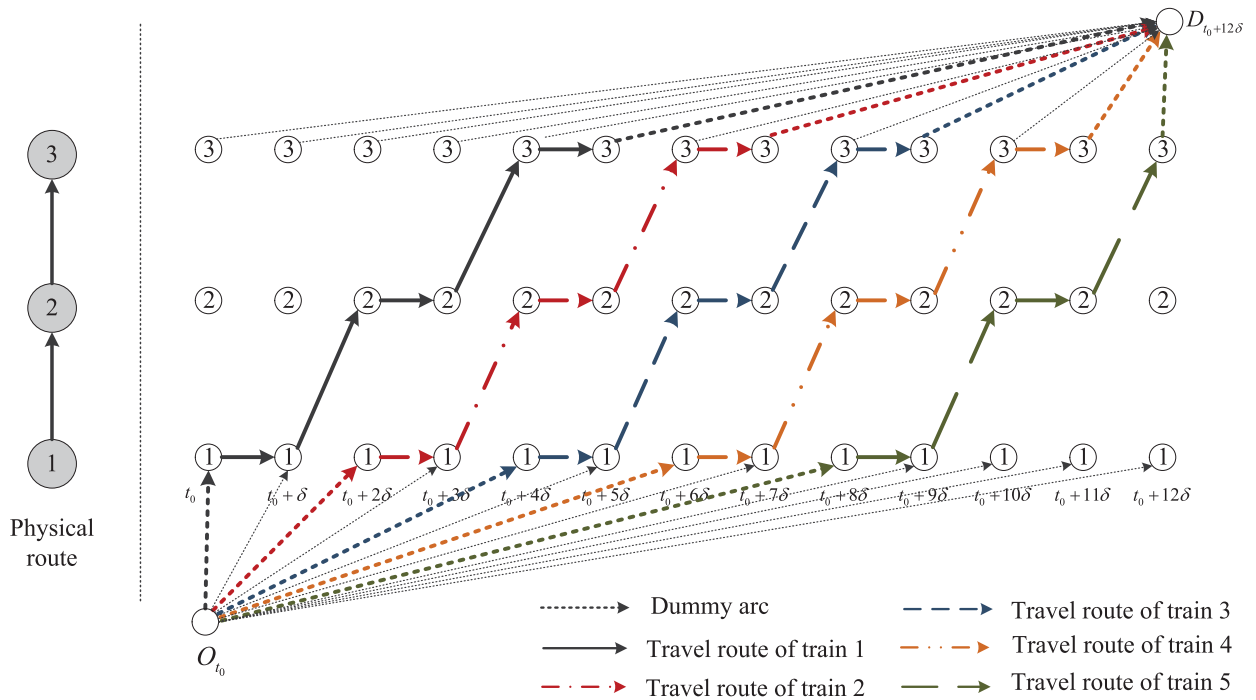
Fig. 5. Dummy origin and destination for trajectories generation of all the trains.

capacity constraint to guarantee that the total passenger demands at one station are no more than the train loading capacity. $\sum_{\tilde{t}\in[t_0,t]}\sum_{t'\in\mathcal{T}}(x^{k-1}_{i'(i'+1)\tilde{t}t'} - x^k_{i'(i'+1)\tilde{t}t'})$ equals to 1 only when train $k-1$ departs from station i' at or before timestamp t while train k has not; otherwise, it equals to 0. This formulation is particularly derived to represent the time period between the departures of train k and $k-1$ from station i', which is similar to the definition given by Niu and Zhou (2013). $\sum^{2I}_{j>i}\delta \cdot p_{i'j}(t+\delta)$ represents the passenger demands in time interval $(t, t+\delta]$ with origin $i' \le i$ and destination $j > i$. The last two constraints, i.e., constraints (31) and (32) denote that the decision variables **x** are binary variables while elements of **b** are nonnegative integer values.

Remark 3.3. Here, we particularly refer to a nonlinear model that was proposed by Niu et al. (2015b) to minimize the passenger waiting time with dynamic demands. This nonlinear model adopts less variables and can sometimes achieve faster computational speed compared with the linear model in this study. Nevertheless, we note that, the nonlinear model is actually difficult to model the train operation states and the energy consumption. Therefore, in this study we derive the linear formulation with the space-time network representation method for the integrated metro train scheduling problem.

3.4. Model 2: ETS with consideration of train regenerative energy

In the former ETS model, we focus on the minimization of passenger waiting time and train traction energy consumption in condition of time-variant passenger demands. We note that, with the increasing concerns on environmental issues, train regenerative energy, which accounts for nearly one third in the train energy consumption of a metro system, has received significant attention in order to reduce the total energy consumption and achieve a green public transport mode (see Albrecht et al., 2016; Huang et al., 2016; Yang et al., 2015c). The regenerative energy is generated by braking trains through converting the kinetic energy into electricity. Typically, one of the efficient approaches to utilize the regenerative braking energy is to coordinate the operations of metro trains properly such that the regenerated energy by a braking train can be immediately used by other accelerating trains in the same power supply substation. The main advantage of this approach is that it has no need of any additional energy storage equipments that will cause extra infrastructure or maintenance costs. Nevertheless, if the regenerative energy cannot be utilized timely, i.e., there is no accelerating train simultaneously, the energy will be wasted by heating resistors. Hence, one of the critical problems is how to schedule the timetable and control the speed profiles of trains such that the regenerative energy from braking trains can be used by nearby accelerating trains.

Due to the complexity of dynamic characteristics of the train operation problem, for instance the "real-time" property of regenerative energy, the most existing approaches for maximizing the use of regenerative braking energy are formulated as complex nonlinear models that can only be solved by some evolutionary heuristic algorithms (e.g., GA, ACO, TS). To this end, we are particularly interested in formulating the regenerative braking energy into a linear formula that can be potentially
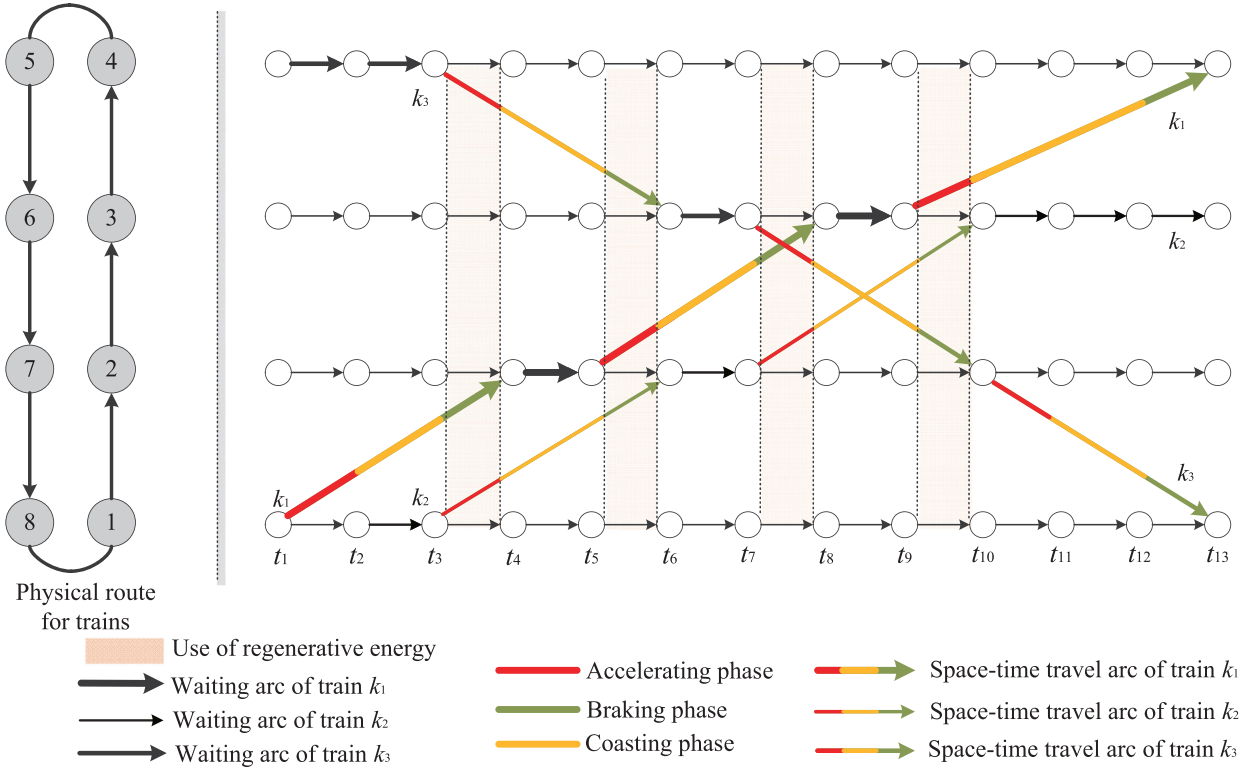
Fig. 6. Regenerative energy transmission in a space-time network.

handled by some optimization solvers or exact algorithms. In the following discussion, we further consider the utilization of train regenerative energy with multiple substations on the basis of the ETS model formulated above.

For illustrative convenience, an example is first given here to state the regenerative energy transformation among train k_1, train k_2 and train k_3 in the space-time network. As shown in Fig. 6, we illustrate the space-time trajectories of trains k_1, k_2 and k_3 that travel in the metro corridor with 8 stations, with the assumption that these eight stations are in the same substation. It can be seen that, trains k_1 and k_2 are travelling in the up-direction, while train k_3 is travelling in the down-direction. For simplicity, each travel arc is assumed to be composed of accelerating, coasting and braking phases, which are illustrated with different colors in this figure. In each time interval, the regenerative energy can be used only when a train is braking and another train is accelerating at the same time. We can see clearly that there are a total of four overlapping regions where some trains brake and the remainder trains accelerate. The first overlapping region is between timestamps t_3 and t_4, when train k_1 brakes on link (1, 2) and train k_2 and k_3 can use the regenerative energy as a part of accelerating energies to travel on their corresponding links. Likewise, the regenerative energy transformations also take place between time intervals $[t_5, t_6]$, $[t_7, t_8]$ and $[t_9, t_{10}]$, respectively.

In this study, we take advantage of the space-time network formulation to model the utilization of regenerative energy among all the trains in the same power supply substation. Specifically, we first consider one time interval $[t, t + \delta]$ in the formulated space-time network. To model the regenerative energy transmission in time interval $[t, t + \delta]$ for each substation $u \in \mathcal{U}$, we need to calculate the total acceleration energy consumption and regenerative braking energy for all the trains in this interval.

In particular, we derive a unified formulation of train energy consumption in time interval $[t, t + \delta]$ and substation u as follows.

$$F_a(u, t) = \sum_{k \in \mathcal{K}} \sum_{(i,j,\tau,\tau') \in \mathcal{A}, \tau \le t, \tau' > t} \left[m_k \phi(i, j, u) x^k_{ij\tau\tau'} \int_{y=t}^{t+\delta} F^k_{ij\tau\tau'}(y) v^k_{ij\tau\tau'}(y) dy \right], \quad \forall u \in \mathcal{U}, t \in \mathcal{T}^M,$$

in which $\mathcal{T}^M = \mathcal{T} \backslash \{t_0 + M\delta\}$, and $\phi(i, j, u)$ denotes whether link (i, j) is located in substation u. In other words, if segment (i, j) is located in substation u, we have $\phi(i, j, u) = 1$; otherwise, $\phi(i, j, u) = 0$. Specifically, in the real-world operations, if $x^k_{ij\tau\tau'} = 0$ or $[t, t + \delta] \cap [\tau, \tau'] = \varnothing$, the traction energy of train k in $[t, t + \delta]$ is typically zero. On the other hand, if $x^k_{ij\tau\tau'} = 1$ and $[t, t + \delta] \cap [\tau, \tau'] \ne \varnothing$, the traction energy of train k may occur in this time interval over the power supply substation. For instance, consider the time interval $[t_7, t_8]$ in Fig. 6. We need to find all the possible travel arcs for the considered trains

that go through this time interval, i.e., space-time arcs $(2, 3, t_5, t_8)$, $(6, 7, t_7, t_{10})$ and $(2, 3, t_7, t_{10})$. Then we sum up their corresponding traction energy of these arcs in time interval $[t_7, t_8]$ to obtain the total traction energy consumption.

Similarly, the regenerative braking energy in time interval $[t, t+\delta]$ and substation u can also be denoted by

$$R_b(u, t) = c_r \sum_{k \in \mathcal{K}} \sum_{(i,j,\tau,\tau') \in \mathcal{A}, \tau \leq t, \tau' > t} \left[m_k \phi(i, j, u) x_{ij\tau\tau'}^k \int_{y=t}^{t+\delta} B_{ij\tau\tau'}^k(y) v_{ij\tau\tau'}^k(y) dy \right], \quad \forall u \in \mathcal{U}, t \in \mathcal{T}^M,$$

in which c_r is the regenerative energy coefficient from kinetic energy to regenerative braking energy.

Here, we note that, both $F_a(u, t)$ and $R_b(u, t)$ have a linear form with respect to the decision variable $\mathbf{x}$. In this sense, the utilization of regenerative energy in time interval $[t, t+\delta]$ over substation u can be given by the minimum of regenerative energy by braking trains (after transmission loss) and energy consumption of accelerating trains, i.e.,

$$G_u(t) = \min\{F_a(u, t), c_a R_b(u, t)\}, \tag{33}$$

for each $t \in \mathcal{T}^M$ and $u \in \mathcal{U}$, in which c_a is the available percentage of regenerative energy after transmission loss. Considering the utilization of regenerative energy, the Model 2, i.e., energy-efficient train scheduling model with train regenerative energy (RTS) is formulated as follows.

$$\min_{\mathbf{x}, \mathbf{b}} Z(\mathbf{x}, \mathbf{b}) = w_E \left[\sum_{k \in \mathcal{K}} \sum_{(i,j,t,t') \in \mathcal{A}} E_k(i, j, t, t') x_{ijtt'}^k - \sum_{t \in \mathcal{T}^M} \sum_{u \in \mathcal{U}} G_u(t) \right]$$
$$+ w_T \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \delta(n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i)) \tag{34}$$

s.t.   constraints $(19) - (32)$.

In this formulation, the objective function further considers the utilization of regenerative energy, and the constraints (19)–(32) are the same as those in Model 1.

Nevertheless, we note that, the objective function in this model has a *min* function, i.e., $G_u(t)$. To transform this formulation into the linear form, we derive a more general formulation by introducing a set of nonnegative variables r_t^u for each $t \in \mathcal{T}^M$ and $u \in \mathcal{U}$, to represent the utilized energy in time interval $[t, t+\delta]$ over power supply substation u. Then, the RTS model can be uniformly formulated as a mixed-integer linear programming (MILP) model, given as follows.

$$\begin{cases} \min_{\mathbf{x}, \mathbf{b}, \mathbf{r}} Z(\mathbf{x}, \mathbf{b}, \mathbf{r}) = w_E \left[\sum_{k \in \mathcal{K}} \sum_{(i,j,t,t') \in \mathcal{A}} E_k(i, j, t, t') x_{ijtt'}^k - \sum_{t \in \mathcal{T}^M} \sum_{u \in \mathcal{U}} r_t^u \right] \\ \qquad + w_T \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \delta(n_{t_0}^i + \sum_{\bar{t} \in [t_0, t]} (d_{\bar{t}}^i - b_{\bar{t}}^i)) \\ \text{s.t.} \quad \text{constraints } (19) - (32) \\ \qquad r_t^u \leq F_a(u, t), \qquad \forall u \in \mathcal{U}, t \in \mathcal{T}^M \\ \qquad r_t^u \leq c_a R_b(u, t), \quad \forall u \in \mathcal{U}, t \in \mathcal{T}^M \\ \qquad r_t^u \in \mathbb{R}^+, \qquad\qquad \forall u \in \mathcal{U}, t \in \mathcal{T}^M. \end{cases} \tag{35}$$

Proposition 2. *The developed RTS model (35) is essentially equivalent to model (34).*

Proof. By the introduction of new variables r_t^u, i.e., the value of utilized regenerative energy in time interval $[t, t+\delta]$ over substation u, these two models can be equivalent only if we guarantee that $r_t^u = \min\{F_a(u, t), c_a R_b(u, t)\}$ is always true in the process of minimizing objective functions. In essence, adding constraints $r_t^u \leq F_a(u, t)$ and $r_t^u \leq c_a R_b(u, t)$ will be sufficient if the model ensures that the largest value of r_t^u will always be chosen. In particular, we note that, the coefficients of variable r_t^u are all negative values, and the variable r_t^u is not connected to other variables except for these new added constraints. It is obvious that, r_t^u will always choose either $F_a(u, t)$ or $c_a R_b(u, t)$ such that the model can be optimized. Therefore, by adding these two constraints, it is ensured that the condition $r_t^u = \min\{F_a(u, t), c_a R_b(u, t)\}$ is always true in the optimal solution, which guarantees the equivalence of these two models. □

3.5. Complexity of formulation

Typically, all the variables in the proposed formulations can be divided into three categories. The first type, i.e., $\mathbf{x}$, is associated with binary decision variables that determine the routing trajectories of scheduled trains in the space-time network.

Table 2
Numbers of variables and constraints in ETS and RTS models.

Variable or constraints	Total number at most for ETS model	Total number at most for RTS model
Binary variable **x**	$|\mathcal{I}| \cdot |\mathcal{T}| \cdot (|\mathcal{L}| + 1) \cdot |\mathcal{K}|$	$|\mathcal{I}| \cdot |\mathcal{T}| \cdot (|\mathcal{L}| + 1) \cdot |\mathcal{K}|$
Integer variable **b**	$|\mathcal{T}| \cdot |\mathcal{I}|$	$|\mathcal{T}| \cdot |\mathcal{I}|$
Continuous variable **r**	0	$|\mathcal{T}| \cdot |\mathcal{U}|$
Flow balance constraints (19)	$(|\mathcal{T}| \cdot |\mathcal{I}| + 2) \cdot |\mathcal{K}|$	$(|\mathcal{T}| \cdot |\mathcal{I}| + 2) \cdot |\mathcal{K}|$
Turn-around constraints	$2|\mathcal{K}| - K_p$	$2|\mathcal{K}| - K_p$
Headway constraints	$2(|\mathcal{K}| - 1) \cdot |\mathcal{A}|$	$2(|\mathcal{K}| - 1) \cdot |\mathcal{A}|$
Dwelling time constraints	$2|\mathcal{I}| \cdot |\mathcal{K}|$	$2|\mathcal{I}| \cdot |\mathcal{K}|$
Passenger boarding constraints	$2|\mathcal{I}| \cdot |\mathcal{T}|$	$2|\mathcal{I}| \cdot |\mathcal{T}|$
Capacity constraints	$|\mathcal{I}| \cdot |\mathcal{K}|$	$|\mathcal{I}| \cdot |\mathcal{K}|$
Regenerative energy constraints	0	$2|\mathcal{U}| \cdot |\mathcal{T}|$

The second refers to integer variables, i.e., **b**, which represent the number of boarding passengers at each timestamp and station. Decision variable **r** is nonnegative decision variables, which are employed in RTS model to formulate the utilization of regenerative energy at each timestamp and substation. It is obvious that there only exist linear formulations in the objectives of two developed models, i.e., ETS model and RTS model. Besides, all the constraints (i.e., constraints (19)–(32) and last three constraints in (35)) are linear equalities or inequalities. We conclude that, ETS and RTS models are linear integer programming model and mixed integer linear programming (MILP) model, respectively.

Next, we are particularly interested in analyzing the complexity of the proposed formulations. The following discussion will detail the total numbers of variables and some critical constraints involved in the problem, listed in Table 2, where $|\cdot|$ represents the cardinality of a set $\{ \cdot \}$. In particular, the number of binary variables **x** is given by $|\mathcal{I}| \cdot |\mathcal{T}| \cdot (|\mathcal{L}| + 1) \cdot |\mathcal{K}|$, in which $|\mathcal{L}| + 1$ represents one waiting arc and $|\mathcal{L}|$ travelling arcs, and each travelling arc corresponds to one operation level l.

It is clear that, the complexity of the ETS and RTS formulations is fully dependent on the physical scale of the scheduled metro corridor, the numbers of time intervals, trains and operation levels of the train control system. For illustrative convenience, an example is given to clarify the total number of decision variables for these two models. We consider 10 trains in a metro corridor with 10 stations (5 up-direction and 5 down-direction stations) covered by 2 power supply substations, and there are 3 operational levels (i.e., 3 alternative speed profiles) for the trains on each physical link. We use 200 time intervals as the time horizon to set up space-time networks, which will produce 80000 binary variables with respect to **x** and 2000 integer variables with respect to **b**. Moreover, if we consider the utilization of regenerative energy over each substation among the trains, a total of 400 nonnegative continuous variables r_t^u are also needed to specify how much regenerative braking energy can be used in each time interval and substation.

4. Solution approaches

The energy-efficient metro train scheduling model with dynamic demands is a complex problem with integer, binary and continuous variables, and it is very computationally intensive in practice, especially for the real-world large-scale problems. For example, for the instance referred in Secion 3.5, there are totally 80,000 binary variables, 2000 integer variables and 400 continuous variables for RTS model. By using the CPLEX solver on a personal computer, it takes as long as 8 hours to solve this model with only 2 trains, and for 4 trains, it takes about 20 hours to obtain a solution with 5% gap (please see Section 5 for details). Therefore, in this section, we aim to develop a solution algorithm that can efficiently obtain a good schedule in a much shorter time. In the following, Section 4.1 proposes a Lagrangian relaxation method that decomposes model (35) (i.e., RTS model) into two relatively easy subproblems, i.e., subproblem 1 with only binary variables $x_{ijtt'}^k$ and continuous variables r_t^u, and subproblem 2 with only integer variables b_t^i. Section 4.2 utilizes the relaxed solution to regenerate a feasible solution that yields an upper bound to the primal objective. Section 4.3 designs a sub-gradient search approach that iteratively updates both relaxed and feasible solutions to reduce the optimality gap between the upper bound (the best feasible solution) and the estimated lower bound to an acceptable tolerance.

Remark 4.1. Note that, the only difference between models ETS and RTS is the introduction of new decision variable r_t^u and the corresponding constraints. Therefore, in the following content, we only present the solution methodology for model RTS (i.e., model (35)), and model ETS (i.e., model (18)) can also be solved in a similar way.

4.1. Lagrangian relaxation and decomposition

It can be seen from the RTS model that, the Eq. (28) includes a set of hard constraints that couple variables **x** and **b**, which may seriously affect the computational speed. If we relax this set of constraints, the model can be decomposed into two separate problems that can be handled more efficiently. Therefore, we first dualized constraints (28) with a set of Lagrangian multipliers $\boldsymbol{\rho} := \{\rho_t^i \geq 0\}_{i \in \mathcal{I}, t \in \mathcal{T} \setminus \{t_0\}}$ to obtain the following relaxed problem:

$$\Delta(\boldsymbol{\rho}) := \min_{\mathbf{x},\mathbf{b},\mathbf{r}} \left\{ \left[w_E \sum_{k \in \mathcal{K}} \sum_{(i,j,t,t') \in \mathcal{A}} E_k(i,j,t,t') x_{ijtt'}^k - w_E \sum_{t \in \mathcal{T}^M} \sum_{u \in \mathcal{U}} r_t^u - \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \rho_t^i \sum_{k \in \mathcal{K}} x_{iit't}^k b_{max}^i \right] \right.$$
$$\left. + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \left[w_T \delta n_{t_0}^i + w_T \sum_{\bar{t} \in [t_0,t]} \delta(d_{\bar{t}}^i - b_{\bar{t}}^i) + \rho_t^i b_{\bar{t}}^i \right] \right\}$$

(36)

subject to constraints (19)–(27), (29)–(32) and the last three constraints in model (35).

Note that because the relaxed problem (36) no longer contains any connections between the variables **x** and variables **b**, we may decompose the relaxed problem into two subproblems. The first subproblem contains only binary variables **x** and continuous variables **r**, given below,

$$\Gamma(\boldsymbol{\rho}) := \min_{\mathbf{x},\mathbf{r}} \left[\sum_{k \in \mathcal{K}} \sum_{(i,j,t,t') \in \mathcal{A}} w_E E_k(i,j,t,t') x_{ijtt'}^k - \sum_{t \in \mathcal{T}^M} \sum_{u \in \mathcal{U}} w_E r_t^u - \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \rho_t^i \sum_{k \in \mathcal{K}} x_{iit't}^k b_{max}^i \right]$$

(37)

subject to constraints (19)–(27), (30)–(31) and the last three constraints in model (35). Obviously, the objective function in model $\Gamma(\boldsymbol{\rho})$ can be rewritten in the following formula:

$$\Gamma(\boldsymbol{\rho}) := \min_{\mathbf{x},\mathbf{r}} \sum_{k \in \mathcal{K}} \left[\sum_{(i,j,t,t') \in \mathcal{A}} w_E E_k(i,j,t,t') x_{ijtt'}^k - \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \rho_t^i x_{iit't}^k b_{max}^i \right] - w_E \sum_{t \in \mathcal{T}^M} \sum_{u \in \mathcal{U}} r_t^u.$$

(38)

Remark 4.2. In essence, the first sub-problem corresponds to a metro train scheduling problem without considering passenger flow, in which no meeting & passing (M&P) and meeting & overtaking (M&O) constraints need to be taken into consideration due to the operation environment of the metro line of interest. In general, M&P and M&O constraints are usually regarded as the difficult constraints in solving the train scheduling problem in the main line railway. Without these two constraints, the CPLEX solver can solve the sub-problem 1 much more easily and faster.

The second subproblem then can be given as follows.

$$\Phi(\boldsymbol{\rho}) := \min_{\mathbf{b}} \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} w_T \delta \left(n_{t_0}^i + \sum_{\bar{t} \in [t_0,t]} (d_{\bar{t}}^i - b_{\bar{t}}^i) \right) + \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} \rho_t^i b_t^i$$
$$\text{s.t.} \quad \sum_{\bar{t} \in [t_0,t]} b_{\bar{t}}^i \leq n_{t_0}^i + \sum_{\bar{t} \in [t_0,t]} d_{\bar{t}}^i, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}$$
$$b_t^i \in \mathbb{N}^+, \qquad\qquad \forall i \in \mathcal{I}, t \in \mathcal{T}.$$

(39)

Here, the constraints in model (39) are essentially the cumulative arrive count relation that is given in Eq. (10). Additionally, we further add the following set of constraints to guarantee that the number of boarding passengers at each time interval and station is restricted by the allowable passenger boarding speed. Typically, these constraints also hold for the primal model, i.e.,

$$b_t^i \leq b_{max}^i, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}.$$

To be more specific, model (39) consists of $|\mathcal{I}|$ small subproblems $\Phi_i(\boldsymbol{\rho})$ for each $i = 1, 2, \ldots, |\mathcal{I}|$, i.e.,

$$\Phi_i(\boldsymbol{\rho}) := \min_{\mathbf{b}} \sum_{t \in \mathcal{T} \setminus \{t_0\}} w_T \delta \left(n_{t_0}^i + \sum_{\bar{t} \in [t_0,t]} (d_{\bar{t}}^i - b_{\bar{t}}^i) \right) + \sum_{t \in \mathcal{T} \setminus \{t_0\}} \rho_t^i b_t^i$$
$$\text{s.t.} \quad \sum_{\bar{t} \in [t_0,t]} b_{\bar{t}}^i \leq n_{t_0}^i + \sum_{\bar{t} \in [t_0,t]} d_{\bar{t}}^i, \quad \forall t \in \mathcal{T}$$
$$b_t^i \leq b_{max}^i, \qquad\qquad \forall t \in \mathcal{T}$$
$$b_t^i \in \mathbb{N}^+, \qquad\qquad \forall t \in \mathcal{T},$$

(40)

in terms of *station-based flow waiting process*, since it essentially propagates the cumulative passenger arrival and departure flows at each station. Clearly, this set of models only contain integer variables b_t^i, which are in the form of cutting stock problems. Since the scale of this subproblem $\Phi_i(\boldsymbol{\rho})$ is much smaller than the original problem, it can be handled more efficiently.

Theoretically, by solving problem $\Gamma(\boldsymbol{\rho})$ and a series of subproblems $\Phi_i(\boldsymbol{\rho})$ with the given $\boldsymbol{\rho}$, we can obtain a lower bound to the original problem (35). Then, the tightest lower bound can be obtained by solving the following Lagrangian dual problem.

$$\max_{\boldsymbol{\rho} \geq 0} \Delta(\boldsymbol{\rho}) := \Gamma(\boldsymbol{\rho}) + \Phi(\boldsymbol{\rho}).$$
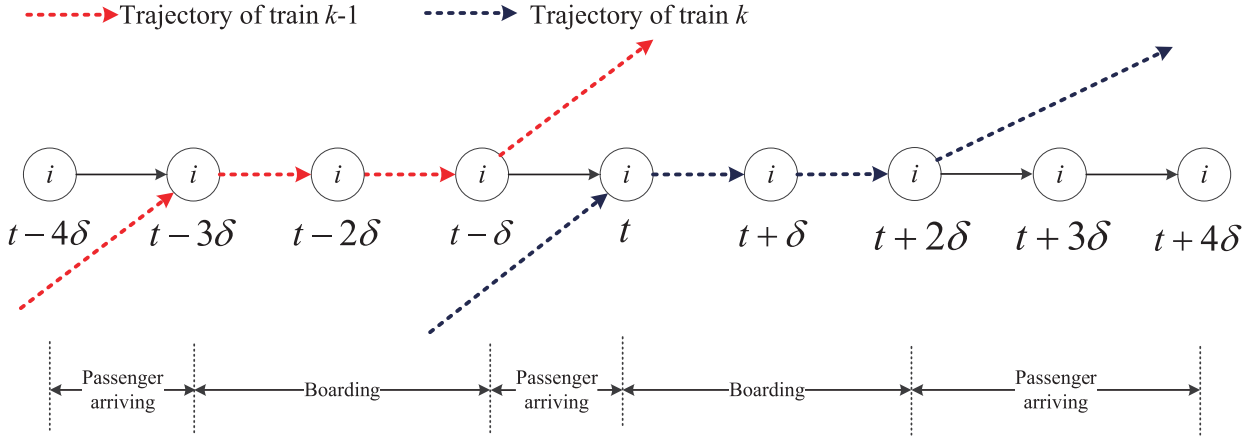
(41)

Fig. 7. Boarding process of waiting passengers.

For each set of Lagrangian multiplier $\{\rho_t^i\}$, if we can find the exact optimal objectives for these two sub-problems, we can produce a lower bound to the original problem. On the other hand, if the sub-problems are still hard to solve their exact optimal solutions (for instance, the real-world case study in the following section), some heuristic algorithms or the optimization solvers (e.g., CPLEX) can be adopted to produce a potential lower bound. The obtained solution information can also be used to update the upper bound if we correct it to a feasible solution to the original problem and act as the initial solution in the next iteration. In the literature, this type of approach is also adopted by some researches, for instance Rajagopalan et al. (2004), Cacchiani et al. (2016) and Dauzère-Pérès et al. (2015).

4.2. Estimating prevailing number of waiting passengers

In the primal model, the aim is to minimize the total train energy consumption and passenger waiting time with time-variant demands. Here, the set of feasible solution typically refers to three categories of variables, i.e., the space-time trajectories of trains denoted by variable $\mathbf{x}$, utilized regenerative energy denoted by $\mathbf{r}$, and the passenger flow control variable given by $\mathbf{b}$.

If the optimal solution obtained by solving relaxed subproblems (38) and (40) is a feasible solution to the primal problem (35), its corresponding objective in the primal problem can be regarded as an upper bound to the optimal objective. Otherwise, we need to develop a heuristic method based on the solutions of subproblems (38) and (40) to reconstruct a feasible solution in order to update the upper bound in the solution process. To obtain a feasible solution, an intuitive way is to fix the feasible trajectories of trains (i.e., $\mathbf{x}$) and regenerative energy (i.e., $\mathbf{r}$), while updating $\mathbf{b}$ values. For simplicity, we denote $\bar{\mathbf{x}} = \{\bar{x}_{ijtt'}^k\}$, $\bar{\mathbf{b}} = \{\bar{b}_t^i\}$ and $\bar{\mathbf{r}} = \{\bar{r}_t^u\}$ to represent the optimal solutions of relaxed model, which are respectively train space-time trajectories, number of boarding passengers and utilization of regenerative energy. Since we dualize constraints (28), we only need to judge if the solutions $\bar{\mathbf{x}}$ and $\bar{\mathbf{b}}$ satisfy the relaxed constraints (28), i.e.,

$$b_t^i \leq \sum_{k \in \mathcal{K}} x_{iit't}^k b_{max}^i$$

for $\forall i \in \mathcal{I}$, $t \in \mathcal{T} \backslash \{t_0\}$ and $t' = t - \delta$. Specifically, this constraints rigorously consider two different cases for passenger boarding process. In the first case, there is no train that dwells at station i and time t, i.e., $\sum_{k \in \mathcal{K}} x_{iit't}^k = 0$, then the constraints guarantee that the number of boarding passenger at this station must equal to 0. In the second case, there is a train that arrives and dwells at station i and time t, i.e., $\sum_{k \in \mathcal{K}} x_{iit't}^k = 1$. In this case, the number of boarding passengers in this time interval is a nonnegative value that should be less than or equal to the number of current waiting passengers and the allowable boarding speed b_{max}^i. To clearly show the passenger boarding process, we adopt Fig. 7 to illustrate the proposed approach, in which we can see that the passengers' boarding is only allowed in dwelling time periods of trains. Note that, due to the headway constraints, there is at most one train that dwells at a station in the same time interval, indicating that $\sum_{k \in \mathcal{K}} x_{iit't}^k \leq 1$.

Therefore, we can consider to find a set of feasible solution $\hat{\mathbf{b}} = \{\hat{b}_t^i\}$ iteratively from the beginning of time horizon $t_0 + \delta$ to $t_0 + M\delta$ for each station $i \in \mathcal{I}$ by using the solution $\bar{\mathbf{x}}$ (note that, the initial values for $b_{t_0}^i$ and $d_{t_0}^i$ are set as zeros, and each $n_{t_0}^i$ is set as a constant value). For example, in the first time interval $[t_0, t_0 + \delta]$, we firstly calculate $\sum_{k \in \mathcal{K}} \bar{x}_{iit_0(t_0+\delta)}^k$ to specify if there is a train that dwells at station i. If no train is dwelling at this station, we set $\hat{b}_{t_0+\delta}^i = 0$. Otherwise, we set $\hat{b}_{t_0+\delta}^i = \min\{n_{t_0}^i + d_{t_0+\delta}^i, b_{max}^i\}$. This indicates that, if the number of boarding passengers is larger than the allowable boarding speed, then only b_{max}^i passengers can board the train in this time interval; otherwise, the waiting passengers at

the initial time t_0 (i.e., $n_{t_0}^i$) and the new arrival passengers (i.e., $d_{t_0+\delta}^i$) will board the dwelling train. In the second time interval, we derive the similar calculation, and we set $\hat{b}_{t_0+2\delta}^i = 0$ if there is no dwelling train; otherwise, we have $\hat{b}_{t_0+2\delta}^i = \min\{b_{max}^i, n_{t_0}^i + d_{t_0+\delta}^i + d_{t_0+2\delta}^i - \hat{b}_{t_0+\delta}^i\}$. The latter formula is the number of cumulative passengers at this station minus the number of passengers who have already boarded the train in the former time intervals. We iteratively make the calculation and assign each $\hat{b}_t^i$ until the end of time horizon $t_0 + M\delta$. It is obvious that the obtained solution $\hat{\mathbf{b}}$ satisfies all the constraints (28), (29), (32) in the primal problem (35) that are relevant to the decision variable $\mathbf{b}$. Then, we summarize the procedure of this heuristic algorithm in Algorithm 1.

Algorithm 1 A heuristic algorithm to construct a feasible solution.

    Step 1. Input the solutions $\bar{\mathbf{x}}$, $\bar{\mathbf{b}}$, and the time-variant passenger demands $\mathbf{p}(t)$.
    Step 2. If $\bar{\mathbf{x}}$ and $\bar{\mathbf{b}}$ satisfy constraint (28), go to Step 4; otherwise, go to Step 3.
    Step 3. Do for $i = 1, \ldots, 2I$;
      Step 3.1 Do for $t = t_0, \ldots, t_0 + M\delta$;
      Step 3.2 Make the following judgement:
        (a) if $t = t_0$, set $\hat{b}_t^i = 0$;
        (b) else if $\sum_{k\in\mathcal{K}} \bar{x}_{iit't}^k = 0$, set $\hat{b}_t^i = 0$;
        (c) else if $t = t_0 + \delta$ and $\sum_{k\in\mathcal{K}} \bar{x}_{iit't}^k = 1$, calculate $\hat{b}_t^i$ by

$$\hat{b}_t^i = \min\{b_{max}^i, n_{t_0}^i + \sum_{j=i}^{2I} \delta \cdot p_{ij}(t_0 + \delta)\}; \tag{42}$$

        (d) else if $t \neq t_0 + \delta$ and $\sum_{k\in\mathcal{K}} \bar{x}_{iit't}^k = 1$, calculate $\hat{b}_t^i$ by

$$\hat{b}_t^i = \min\{b_{max}^i, n_{t_0}^i + \sum_{\tau=t_0+\delta}^{t} \sum_{j=i}^{2I} \delta \cdot p_{ij}(\tau) - \sum_{\tau=t_0}^{t-\delta} \hat{b}_\tau^i\}. \tag{43}$$

    Step 4. Return the set of feasible solutions $\{\hat{b}_t^i\}$ and the corresponding objective of the primal problem.

The developed Algorithm 1 takes a polynomial solution time of O($|\mathcal{I}||\mathcal{T}|$), which is very time-efficient for searching a feasible solution to the primal problem. By inserting these feasible decision variables into the objective function of the primal model (35), we can obtain a feasible objective value, which serves as an upper bound of the true optimal objective.

4.3. Updating Lagrangian multipliers

Theoretically, if the objective of the Lagrangian dual problem equals to the upper bound of the primal problem, we can return the corresponding feasible solution, which is guaranteed to be the optimal solution according to the duality theory. Otherwise, we need to iteratively update the multipliers $\boldsymbol{\rho}$ based on the sub-gradient algorithm and the solutions of the relaxed model. Here, we introduce two methods in this Lagrangian relaxation-based heuristic to terminate the solution process: (i) if $w > W_{max}$, in which w and W_{max} denote the iteration index and the maximum pre-given iteration number, the algorithm can be stopped; (ii) if the computational time exceeds the pre-determined time threshold T_{lim}, the solution process can be terminated.

In this study, we first give a set of initial Lagrangian multipliers $\{(\rho_t^i)^{w=0}\}_{i\in\mathcal{I}, t\in\mathcal{T}\setminus\{t_0\}}$ with values of 0. Then, we derive a sub-gradient algorithm, which is widely used in Lagrangian relaxation approaches (e.g., Xie et al., 2016; Yang and Zhou, 2014) in order to iteratively update the Lagrangian multipliers. By using the sub-gradient algorithm, each Lagrangian multiplier $(\rho_t^i)^w$ for $t \in \mathcal{T}\setminus\{t_0\}$ and $i \in \mathcal{I}$ is updated to $(\rho_t^i)^{w+1}$ by

$$(\rho_t^i)^{w+1} = \max\left\{(\rho_t^i)^w + \lambda^w((\bar{b}_t^i)^w - \sum_{k\in\mathcal{K}}(\bar{x}_{iit't}^k)^w b_{max}^i), 0\right\}, \quad \forall i \in \mathcal{I}, t \in \mathcal{T}\setminus\{t_0\}, \tag{44}$$

in which $\bar{\mathbf{b}}^w$ and $\bar{\mathbf{x}}^w$ are the optimal solutions of relaxed models in the wth iteration, and λ^w is the step size that is computed in each iteration by

$$\lambda^w = \frac{\alpha^w |UB^w - LB^w|}{\sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}\setminus\{t_0\}}((\bar{b}_t^i)^w - \sum_{k\in\mathcal{K}}(\bar{x}_{iit't}^k)^w b_{max}^i)^2}. \tag{45}$$

In the above equation, α^w is the step parameter, UB^w is the best known upper bound up to iteration w, and LB^w is the best potential lower bound up to iteration w found by CPLEX solver in the searching process.
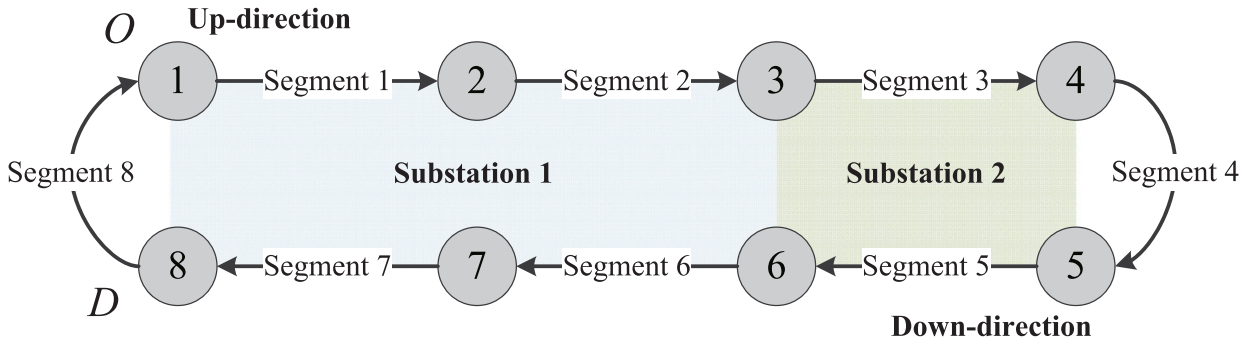
Fig. 8. A simple bi-direction metro corridor.

4.4. Procedure of the LR-based algorithm

Based on the introduction of Lagrangian relaxation scheme, the developed heuristic algorithm and the Lagrangian multiplier updating methods, we summarize the general procedure of LR-based solution process in Algorithm 2.

Algorithm 2 LR-based algorithm for energy-efficient metro train scheduling problem.

Step 1. Input initial information, including the maximum number of fleet size, metro corridor parameters, set of train speed profiles, time-variant passenger demands, dwelling time constraints, planning time horizon, etc.

Step 2. Set the initial Lagrangian multipliers $(\rho_t^i)^0 = 0$ for each $t \in \mathcal{T} \setminus \{t_0\}$ and $i \in \mathcal{I}$.

Step 3. Solve the subproblems $\Phi(\boldsymbol{\rho}^w)$ and $\Gamma(\boldsymbol{\rho}^w)$ and obtain solutions $\bar{\mathbf{x}}^w$, $\bar{\mathbf{b}}^w$ and $\bar{\mathbf{r}}^w$.

Step 4. Calculate the feasible solutions by using Algorithm 1, and update the upper bound.

Step 5. Update the Lagrangian multipliers by the subgradientalgorithm using Eq. (44) and Eq. (45).

Step 6. Go to Step 7 if satisfying one of the following conditions; otherwise, set $w \leftarrow w + 1$ and go to Step 3:

   (i) if $w > W_{\max}$, in which $W_{\max}$ is the maximum iteration number;

   (ii) if the computational time exceeds the pre-determined time limit T_{lim}.

Step 7. Output the best upper bound and the corresponding feasible solutions.

---

5. Numerical examples

In this section, a series of numerical experiments, involving a small-scale case and a real-world case study on Beijing Metro Yizhuang line, are implemented to illustrate the applications of the proposed models and algorithms. The proposed Lagrangian relaxation-based heuristic framework is coded in Visual C++ 2012 on a Window 8 personal computer. The subproblems in the relaxed model are solved by IBM CPLEX 12.3 Academic Version on the same platform.

5.1. A small-scale example

In this example, we consider a bi-direction metro corridor with 8 stations, as shown in Fig. 8. This metro corridor consists of 2 power supply substations. Substation 1 covers segment 1, segment 2, segment 6 and segment 7, while substation 2 covers segment 3 and segment 5. The trains leave from station 1, arrive at station 4, then turn around and finally end at station 8. For simplicity, the distance of each segment between any two adjacent stations is assumed to be 2.0 km. There are two types of speed profiles for the running trains in every segment. Level 1 and level 2 speed profiles correspond to the link travel times of 120 s and 150 s, respectively. In addition, the minimum headway between two adjacent trains is 180 s, and the minimum and maximum dwelling times at each station are set as 30 s and 90 s, respectively.

In order to test the effectiveness of the proposed models and solution algorithms, we derive a set of instances in this small-scale metro corridor with different numbers of trains and timestamps. To better illustrate the parameter settings of each instance, we use an instance index, i.e., N-K-S-M, in which K, S and M respectively represent the maximum numbers of scheduled trains, stations and time intervals in the formulated space-time diagram. In this set of experiments, the minimum time interval is set as 30 s and we suppose that the considered time horizon for train scheduling is [0, 30 × M] (unit: second). For example, the instance N-7-8-100 indicates that we schedule 7 trains in a metro corridor with 8 stations in time horizon [0, 3000] (unit: second). For illustrative purpose, Table 3 shows the average time-variant passenger demands adopted in this set of experiments under the maximum length of time horizon 9000 s. It is clear that the average passenger demands are higher in time periods [300, 600] and [4200, 5400], while the average passenger demands are relatively lower in the middle of the planning horizon.

Table 3
The average time-variant passenger arrival rates.

Time period (s)	Average passenger arrival rates (passenger/min)
0–300	1.50
300–600	4.50
600–1200	3.00
1200–2400	1.50
2400–3600	0.00
3600–4200	1.50
4200–4800	3.00
4800–5400	4.50
5400–6000	1.50
6000–7200	3.00
7200–8100	1.50
8100–9000	0.00

In the experiments, we first use CPLEX solver to solve these two developed models, i.e., ETS model that ignores the utilization of regenerative energy, and RTS model, which is a more complex model that further considers the utilization of regenerative energy among the running trains. Thus, in the results, there is no regenerative energy for ETS model. Then, we use the developed Lagrangian relaxation-based heuristic to solve the RTS model, and compare its performances, including objective value (OB), total passenger waiting time (WT), energy consumption (EA), utilization of regenerative energy (RE), relative gap and CPU times, with the results obtained by CPLEX solver. In this set of experiments, the two coefficients, i.e., w_E and w_T are set as 10.0 and 1.0, respectively. In addition, the maximum CPU time is limited to 7200 s, and after that time, we stop CPLEX or LR algorithm to output the best results.

Table 4 shows the computational results for 14 instances with different numbers of trains and planning time horizons, in which we use CPLEX solver to solve both ETS and RTS models, and use the developed LR-based algorithm to solve the RTS model. These three approaches are expressed in this table as ETS-CPLEX, RTS-CPLEX and RTS-LR, respectively. From experimental results, we can see clearly that, under the same planning time horizon, when we add more trains (for example, from instances N-2-8-110 to N-8-8-110), the total passenger waiting time is evidently decreased while the energy consumption increases at the same time. This fact is consistent with the practical experiences that higher train departure frequency (or shorter headway time) reduces passenger waiting time at the stations, which improves the service quality, but will potentially increase the train operational costs. The gap returned by CPLEX solver for both ETS and RTS models with limited CPU time (i.e., 7200 s) presents drastic variations for the tested instances (i.e., from about 3% to 92%). This is essentially because of the large-scale decision variables due to the complexity of formulated models. In particular, the returned gap by solving RTS model is larger than that of ETS model, since RTS model further considers the regenerative energy by introducing some additional continuous decision variables, which makes the RTS model even more difficult to be solved. Interestingly, when the planning time horizon is shorter than 4500 s (i.e., $M \leq 150$), RTS model by CPLEX solver can usually achieve better objective values than those of ETS model with the same computational time (i.e., 7200 s), since RTS model can reasonably utilize the regenerative braking energy so as to reduce the objective value. Nevertheless, when we continue to increase the planning time horizon, CPLEX solver gradually becomes unable to solve RTS model, leading to the worse result than that of ETS model (e.g., see instances N-8-8-210, N-9-8-210, N-10-8-210). In addition, the returned gaps of these two models with long time horizon become very large, indicating that CPLEX solver is very sensitive to the scale of planning time horizon for the formulated models. Finally, when we continue to increase the planning time horizon, CPLEX solver fails to solve the formulated models with 300 timestamps, and returns no feasible solution (represented by "NAN", e.g., see instance N-15-8-300 for example).

Then, we analyze the performance of the developed LR-based heuristic algorithm in solving RTS model with these tested instances. It is easy to see that, LR-based algorithm takes much shorter computational time and obtains the results that are very close to or better than those of CPLEX solver in the test instances with $M = 110$. In particular, when we consider longer time horizons (i.e., $M \geq 150$), the LR-based algorithm performs much better than the CPLEX solver. For example, in instance N-5-8-150, LR-based algorithm improves the objective values for about 10% in comparison with that of RTS model solved by CPLEX solver, and the computational time of LR-based algorithm is still shorter than that of CPLEX solver. Moreover, when we further consider larger instances with more trains and long time horizon (e.g., $M \geq 200$), the RTS model cannot return a feasible solution by using CPLEX in the limited computational time (i.e., 7200 s). Nevertheless, when $M \geq 200$, the developed LR algorithm can still be very effective in solving the RTS model, which obtains a very good solution within the limited computational time. We take instance N-15-8-300 as an illustration, which schedules 15 trains in time horizon [0, 9000] s. By running CPLEX solver for 7200 s, it still cannot find a feasible solution for both ETS and RTS models, while LR algorithm returns a good solution within an acceptable computational time. In summary, these experiment results verify that the effectiveness and efficiency of the LR heuristic algorithm are significantly better than those of CPLEX solver in solving the developed models.

For illustrative convenience, Fig. 9 plots the variation trends of the lower and best upper bounds for instance N-5-8-110 in the searching process with 60 iterations (here, lower bound is produced by CPLEX solver in each iteration). It is obvious

Table 4
Performance comparison for different instances with CPLEX and LR-based heuristic.

| Instance index		WT (s)	EA (J/Kg)	RE (J/Kg)	OB	CPU time (s)	Gap
N-2-8-110	ETS-CPLEX	370440	1890	–	389340	7200.14	14.88%
	RTS-CPLEX	375390	1860	224	391750	7200.51	39.31%
	RTS-LR	375570	1890	0	394770	182.3	–
N-3-8-110	ETS-CPLEX	274000	2910	–	303100	7200.20	42.34%
	RTS-CPLEX	275820	2820	80	303220	7200.19	47.44%
	RTS-LR	255340	2900	80	283540	561.1	–
N-4-8-110	ETS-CPLEX	204660	3750	–	242160	7014.69	30.34%
	RTS-CPLEX	213000	3930	1020	242100	7234.82	42.07%
	RTS-LR	212310	3930	242	249190	623.3	–
N-5-8-110	ETS-CPLEX	167430	4590	–	213330	7305.74	22.16%
	RTS-CPLEX	167250	4710	746	206890	7150.28	31.98%
	RTS-LR	167760	4710	314	211720	1323.27	–
N-7-8-110	ETS-CPLEX	125700	6480	–	190500	7200.33	9.04%
	RTS-CPLEX	132900	6630	1436	184840	7159.41	23.01%
	RTS-LR	130830	6630	1310	184030	833.3	–
N-8-8-110	ETS-CPLEX	120360	7530	–	195660	7200.46	2.89%
	RTS-CPLEX	123030	7590	1312	185810	7200.34	31.72%
	RTS-LR	120220	7460	402	190800	1123.6	–
N-5-8-150	ETS-CPLEX	329240	4690	–	376140	7200.18	38.21%
	RTS-CPLEX	323580	4620	468	365100	7200.05	62.69%
	RTS-LR	288750	4470	242	331030	4899.83	–
N-7-8-150	ETS-CPLEX	226130	6500	–	291130	7200.46	37.32%
	RTS-CPLEX	235380	6240	890	288880	7201.99	51.28%
	RTS-LR	217350	6500	410	278250	5342.83	–
N-8-8-210	ETS-CPLEX	440260	7700	–	517260	7200.92	45.66%
	RTS-CPLEX	1154490	7640	638	1224510	7200.18	85.18%
	RTS-LR	437700	7700	427	510430	2323.9	–
N-9-8-210	ETS-CPLEX	442380	8400	–	526380	7200.99	48.89%
	RTS-CPLEX	579990	9240	1060	661790	7201.97	74.21%
	RTS-LR	428580	8310	332	508360	2968.9	–
N-10-8-210	ETS-CPLEX	456330	9060	–	546930	7202.55	52.02%
	RTS-CPLEX	NAN	NAN	NAN	NAN	NAN	NAN
	RTS-LR	416670	9130	388	504090	4811.1	–
N-11-8-210	ETS-CPLEX	403260	9780	–	501060	7201.19	69.76%
	RTS-CPLEX	NAN	NAN	NAN	NAN	NAN	NAN
	RTS-LR	370500	9960	988	460220	6154.28	–
N-7-8-300	ETS-CPLEX	1434180	6660	–	1500780	7202.23	91.59%
	RTS-CPLEX	9783000	6480	852	9839280	7200.09	99.16%
	RTS-LR	1338090	6450	224	1400350	5446.6	–
N-15-8-300	ETS-CPLEX	NAN	NAN	NAN	NAN	NAN	NAN
	RTS-CPLEX	NAN	NAN	NAN	NAN	NAN	NAN
	RTS-LR	510330	13470	970	635330	6888.3	-
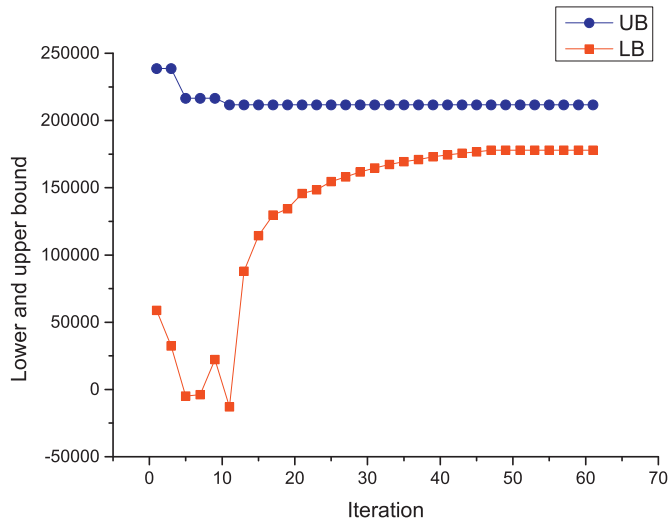
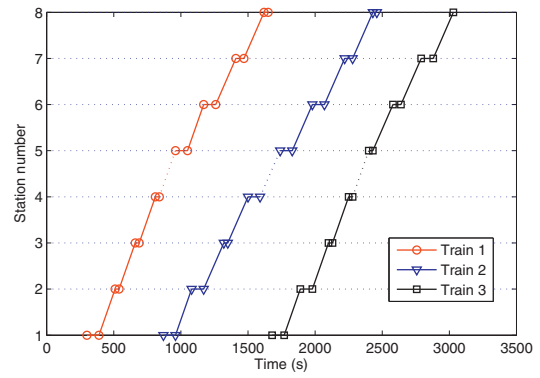**Fig. 9.** Lower and upper bound variations.
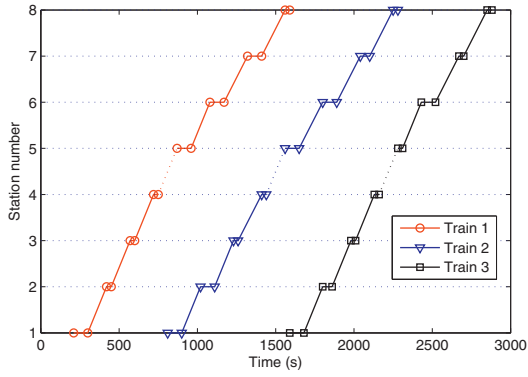
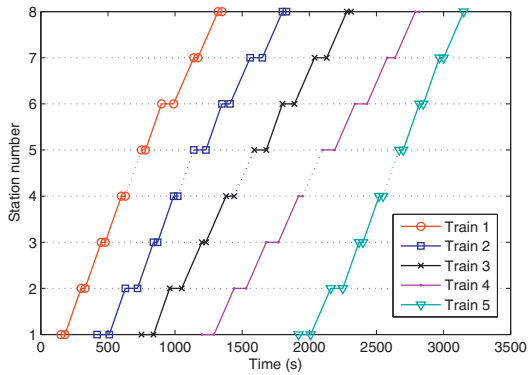(a) Instance N-3-8-110: Obtained timetable by ETS model
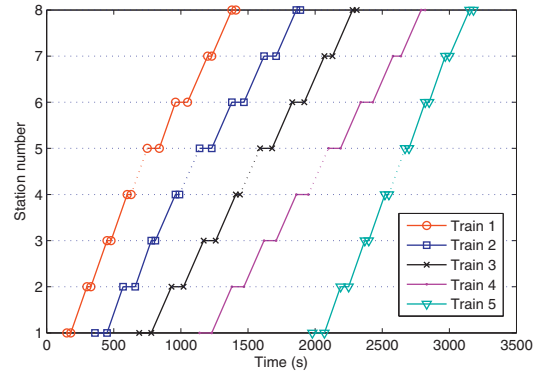
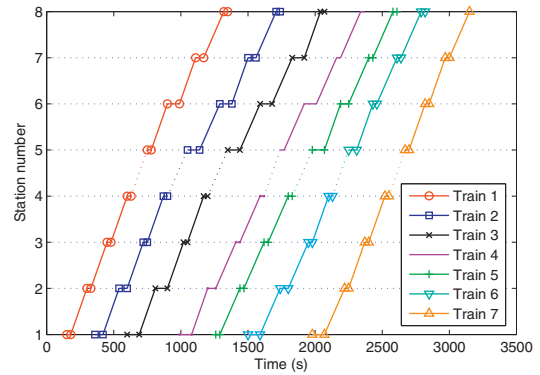(b) Instance N-5-8-110: Obtained timetable by ETS model
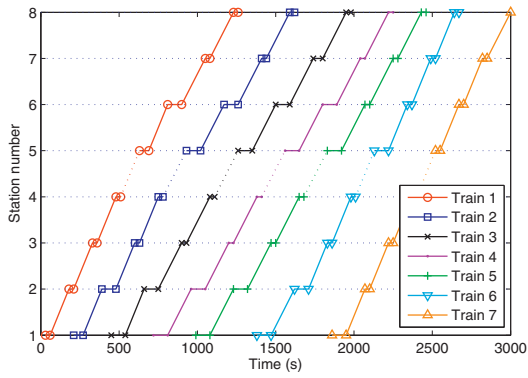
(c) Instance N-7-8-110: Obtained timetable by ETS model

(d) Instance N-3-8-110: Obtained timetable by RTS model
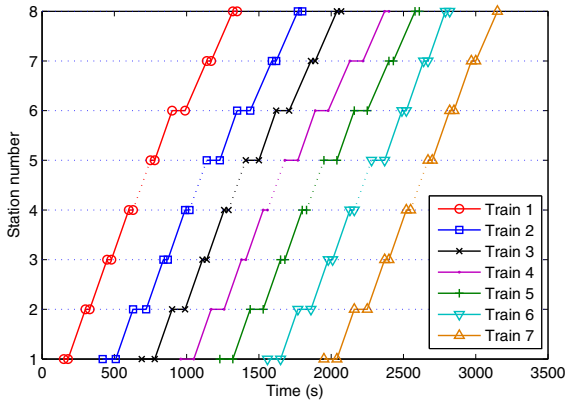
(e) Instance N-5-8-110: Obtained timetable by RTS model

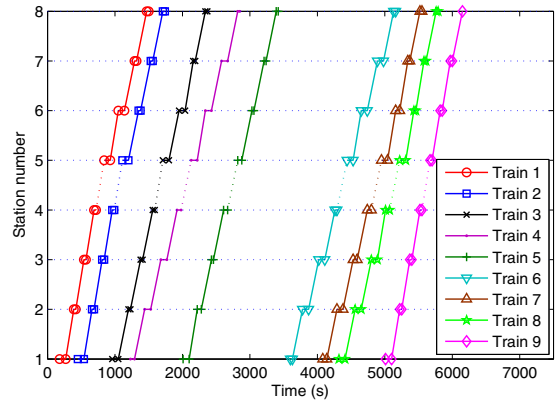(f) Instance N-7-8-110: Obtained timetable by RTS model

Fig. 10. Train schedule by CPLEX solver for different instances.

that, the developed heuristic algorithm iteratively improves the solution quality and the best upper bound is achieved after about 20 iterations.

Fig. 10 depicts the obtained scheduling plans by CPLEX solver corresponding to instances N-3-8-110, N-5-8-110 and N-7-8-110 in Table 4. We illustrate the timetables of ETS and RTS models with maximum number of 3, 5, and 7 trains, respectively. It is shown that, all the trains depart from the origin station at different timestamps one by one, turn-around at station 4 and finally go back to station 1. In this process, the headway constraints are strictly guaranteed between adjacent trains. We can see obviously from Fig. 10b–e that, the departure frequencies are essentially consistent with the time-variant passenger demands. For example, the passenger demands are set to be higher in time period [300, 1200] s (see Table 3), which correspond to the higher departure frequency and shorter departure headway. Apparently, dispatching more trains
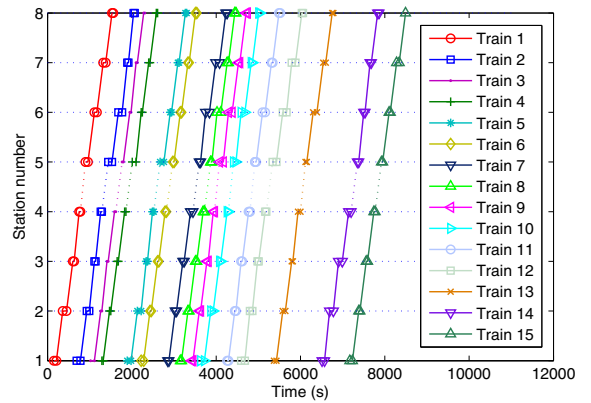
(a) Instance N-7-8-110: Obtained timetable with 7 trains

(b) Instance N-9-8-210: Obtained timetable with 9 trains
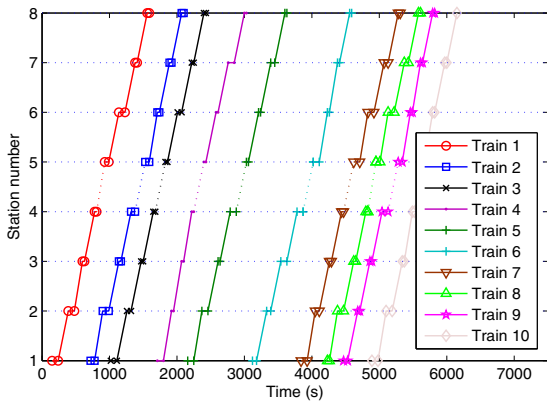
(c) Instance N-10-8-210: Obtained timetable with 10 trains

(d) Instance N-15-8-300: Obtained timetable with 15 trains

Fig. 11. Train schedule by LR-based algorithm with different numbers of trains and time horizons.

in high-demand periods can reduce the total waiting time of passengers. This also verifies that our formulated models can adjust the train departure times in order to reduce the passenger waiting time at the stations.

Fig. 11 shows the train scheduling plans by LR-based algorithm with different train scales and time horizons with respect to instances N-7-8-110, N-9-8-210, N-10-8-210 and N-15-8-300 in Table 4. These figures indicate the similar conclusion that the larger passenger demands are consistent with higher train departure frequencies. For example, in Fig. 11b and c, higher departure frequency is associated with higher demands in time periods [300, 1200] s and [4200, 5400] s, while the lower departure frequency corresponds to the low demands in time period [2400, 3600] s.

5.2. Numerical experiments on Beijing metro Yizhuang line

In this example, we consider a real-world case study on the Beijing metro Yizhuang line. In the implementations, the dynamic parameters (i.e., passengers' time-variant origin-destination demands) are all taken from historical detected operation data. We next execute this set of experiments with different planning time horizons and numbers of in-service trains to verify the effectiveness of the proposed LR-based approach for metro train scheduling problem.

5.2.1. Experiment description and parameter settings

The Yizhuang line is a bi-directional metro line with a total length of 23.3 km, and each direction consists of 13 stations and 12 running segments (see Fig. 12). Some basic data settings in this set of experiments, including the distance of each segment, the maximum and minimum dwelling time at each station, the planned running time in each segment, and the covered segments of each power supply substation, are all illustrated in Table 5. In addition, the other parameter settings in the experiments are set as: the fleet size is 10, the minimal headway is $h_{\min} = 180$ s and the turnaround times at the turnaround station and start terminus are set as $TR_k = 90$ s and $TA_k = 150$ s, respectively. In Beijing metro system, a type of train named DKZ32 is widely used and its Davis parameters are listed in Table 6 for the following experiments (Chen et al., 2013). In this set of numerical experiments, we use the real-world passenger demand data at each station collected on
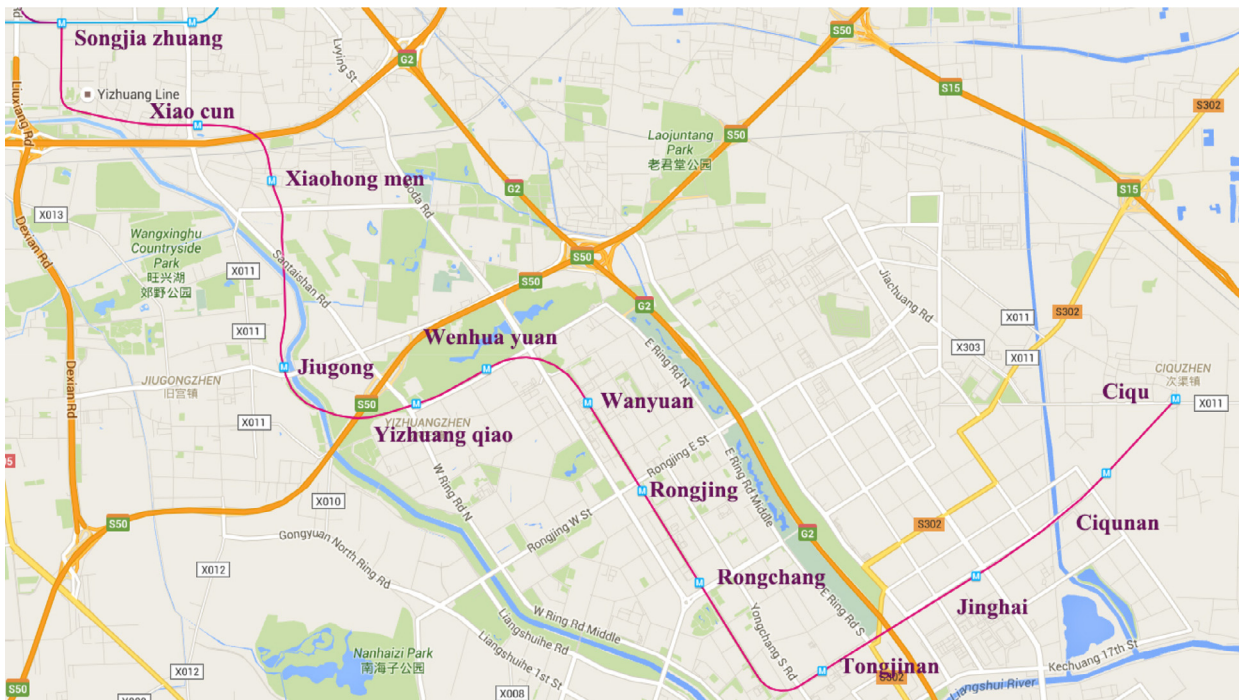
Fig. 12. Illustration of Beijing metro Yizhuang line (Source: Google map).

Table 5
Some numerical characteristics in this case study.

Station index & segment	Length (m)	Minimum dwelling time (s)	Maximum dwelling time (s)	Running time options (s)	Substation index
1 Songjia zhuang - Xiao cun	2641	30	90	210, 180, 150	1
2 Xiao cun - Xiaohong men	1337	30	90	150, 120, 90	1
3 Xiaohong men - Jiugong	2377	30	90	180, 150, 120	2
4 Jiugong - Yizhuang qiao	1993	30	90	180, 150, 120	2
5 Yizhuang qiao - Wenhua yuan	998	30	90	120, 90, 60	3
6 Wenhua yuan - Wanyuan	1543	30	90	150, 120, 90	3
7 Wanyuan - Rongjing	1285	30	90	150, 120, 90	3
8 Rongjing - Rongchang	1360	30	90	150, 120, 90	4
9 Rongchang - Tongjinan	2348	30	90	180, 150, 120	4
10 Tongjinan - Jinghai	2274	30	90	180, 150, 120	5
11 Jinghai - Ciqunan	2096	30	90	180, 150, 120	5
12 Ciqunan - Ciqu	1290	30	90	150, 120, 90	6
13 Turnaround	-	30	90	90, 90, 90	-
14 Ciqu - Ciqunan	1290	30	90	150, 120, 90	6
15 Ciqunan - Jinghai	2096	30	90	180, 150, 120	5
16 Jinghai - Tongjinan	2274	30	90	180, 150, 120	5
17 Tongjinan - Rongchang	2348	30	90	180, 150, 120	4
18 Rongchang - Rongjing	1360	30	90	150, 120, 90	4
19 Rongjing - Wanyuan	1285	30	90	150, 120, 90	3
20 Wanyuan - Wenhua yuan	1543	30	90	150, 120, 90	3
21 Wenhua yuan - Yizhuang qiao	998	30	90	120, 90, 60	3
22 Yizhuang qiao - Jiugong	1993	30	90	180, 150, 120	2
23 Jiugong - Xiaohong men	2377	30	90	180, 150, 120	2
24 Xiaohong men - Xiao cun	1337	30	90	150, 120, 90	1
25 Xiao cun - Songjia zhuang	2641	30	90	210, 180, 150	1
26 Terminal	–	30	90	120, 120, 120	–

a weekend of October, 2014. The demand profiles at different stations are illustrated in Fig. 13. It can be seen that the passenger demand curve varies throughout a day, which is typically related to the different stations and service times. For example, this figure explicitly expresses the peak-hours in the metro corridor with high demands around 8: 00 am and 17: 00 pm, while the demands are relatively low in the afternoon, for instance, from 13:00 pm to 16:00 pm.

In the following experiments, we first consider different periods of planning time horizons with a fixed number of in-service trains. For example, an instance with planning time period [7: 00 am, 10: 05 am] with 16 trips means that, the first

Table 6
Parameter settings in the numerical experiments.

Parameters	Symbol	Value
Train mass (*Kg*)	M	1.99×10^5
Train capacity	C_k	1468
Davis parameter	θ_a	1.36×10^{-4}
Davis parameter	θ_b	1.45×10^{-2}
Davis parameter	θ_c	1.244
Coefficient from kinetic energy to regenerative energy	c_r	0.75
Available energy percentage after transmission	c_a	0.8
Allowable passenger boarding speed (person/min)	$b^i_{\max}$	200
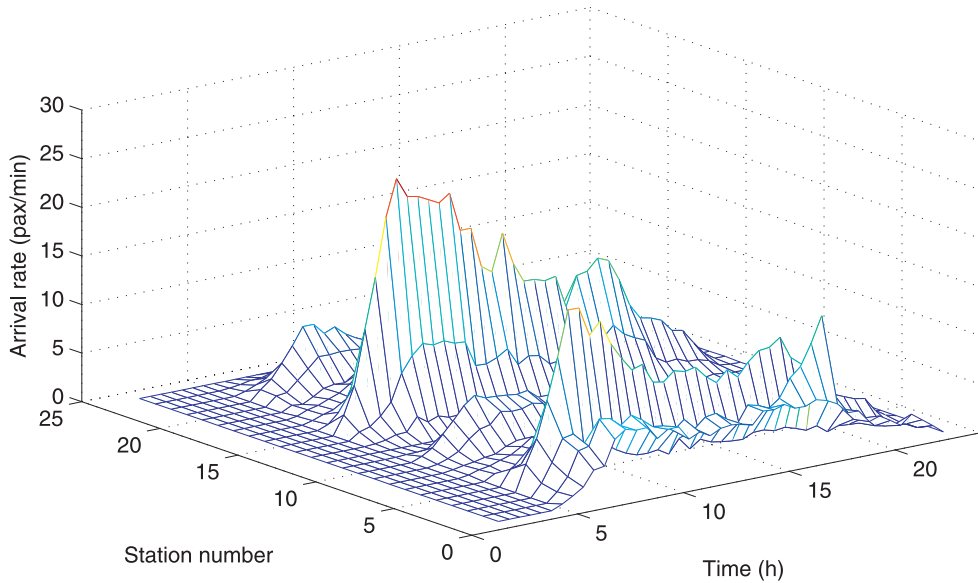
Fig. 13. Passenger demand variations of Yizhuang line on a weekend.

trip must begin after time 7: 00 am, and the last trip will end before time 10: 05 am at its destination. In this time period, the passengers arrive at the stations according to the given time-variant passenger arrival rates mentioned above. Besides, the discretized time interval in the space-time network is set as 30 s to balance the real-world conditions and computational efficiency.

5.2.2. Computational results

In the experimental implementations for the real-world instance, we first try to use CPLEX solver to solve the developed ETS and RTS models. But, when the planning time horizon is longer than one hour, e.g., the timestamp exceeds 200 with more than 10 trains, the CPLEX solver can no longer provide even a feasible solution for the developed two models through hours of computation because of the explosion of decision variables (0–1 variables, integer variables and continuous variables). Due to this fact, we only use the developed LR-based heuristic to solve the RTS model for scheduling trains in the following experiments. In order to evaluate the performance of LR-based approach, we specifically derive two additional timetables as benchmarks. The first timetable (denoted by E-timetable) focuses on the minimization of energy consumption by scheduling all the trains with the lowest speed in every segment. The other timetable (denoted by T-timetable), in contrast, makes more beneficial to the passengers, and the trains are scheduled with the fastest speed in every segment. In addition, the departure frequency of these two timetables are set as constant values. Here, we consider three instances for the train scheduling problem in three distinctive time periods, i.e., 7:00 to 9:55 am, 12:30 to 15:25 pm and 16:30 to 19:25 pm, which are indexed by instance 1, instance 2 and instance 3, respectively. Each time period consists of 350 time intervals in the space-time network. For each instance, we schedule 16 trips, and the last trip (i.e., trip 16) needs to arrive at the destination (i.e., start terminal) before the end of time horizon.

Remark 5.1. We note that, since the trains are scheduled with the lowest speed, the E-timetable is essentially unavailable in practical applications due to the fleet size constraints. In other words, train $k + K_p$ will possibly depart before train k arrives, which violates the fleet size constraints. In this set of experiments, we only aim to compare the performances of E-timetable with our proposed approach. So we neglect the fleet size constraints in the process of generating E-timetable.

Table 7
Performance comparison of E-timetable, T-timetable and LR-based approach for different instances.

| Instance index		WT (s)	EA (KJ)	RE (KJ)	OB	CPU time (h)
Instance 1	E-timetable	$8.87 \cdot 10^6$	$9.57 \cdot 10^6$	$1.04 \cdot 10^6$	$1.74 \cdot 10^7$	–
7:00-9:55	T-timetable	$7.27 \cdot 10^6$	$1.49 \cdot 10^7$	$4.68 \cdot 10^5$	$2.17 \cdot 10^7$	–
K=16	LR-based algorithm	$6.28 \cdot 10^6$	$1.054 \cdot 10^7$	$7.60 \cdot 10^5$	$1.61 \cdot 10^7$	5.9
Instance 2	E-timetable	$5.82 \cdot 10^6$	$9.57 \cdot 10^6$	$1.04 \cdot 10^6$	$1.44 \cdot 10^7$	–
12:30-15:25	T-timetable	$3.69 \cdot 10^6$	$1.49 \cdot 10^7$	$4.68 \cdot 10^5$	$1.81 \cdot 10^7$	–
K=16	LR-based algorithm	$3.63 \cdot 10^6$	$9.73 \cdot 10^6$	$8.80 \cdot 10^5$	$1.25 \cdot 10^7$	5.2
Instance 3	E-timetable	$5.11 \cdot 10^6$	$9.57 \cdot 10^6$	$1.04 \cdot 10^6$	$1.36 \cdot 10^7$	–
16:30-19:25	T-timetable	$2.70 \cdot 10^6$	$1.49 \cdot 10^7$	$4.68 \cdot 10^5$	$1.71 \cdot 10^7$	–
K=16	LR-based algorithm	$2.68 \cdot 10^6$	$9.89 \cdot 10^6$	$7.90 \cdot 10^5$	$1.18 \cdot 10^7$	5.3
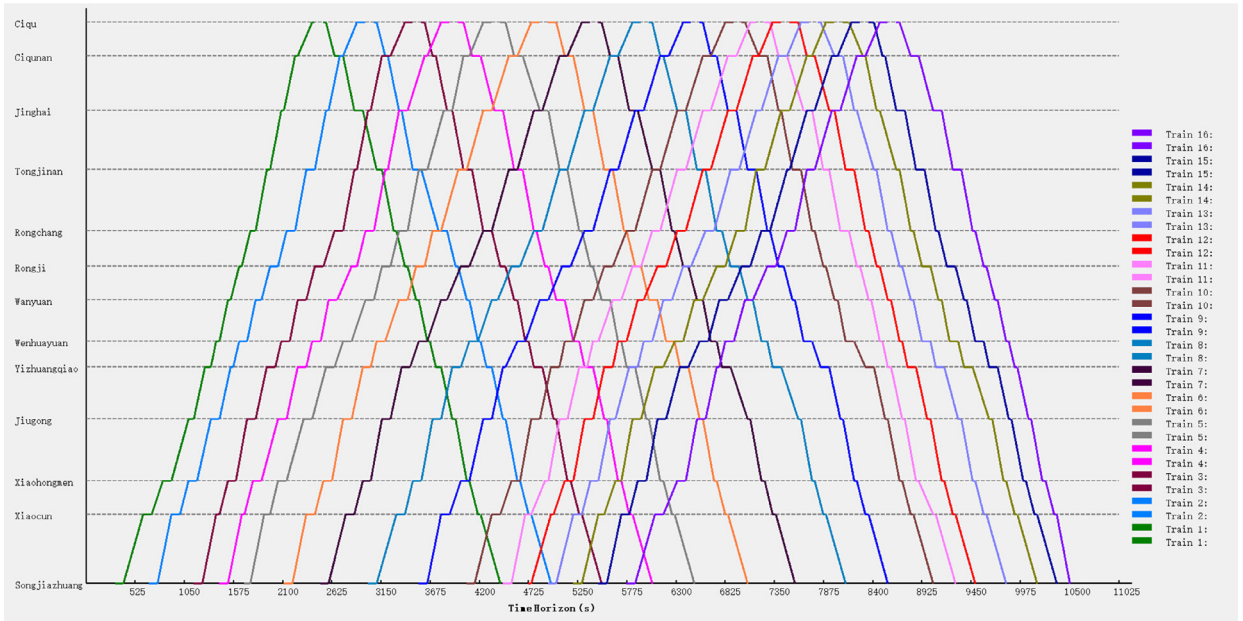
Fig. 14. The obtained timetable by our approach.

Table 7 demonstrates the computational results for T-timetable, E-timetable and LR-based algorithm for these three tested scenarios. Among all these tested scenarios, we can see that, T-timetable consumes the most traction energy, while E-timetable has the minimum traction energy consumption. This is obviously because that, T-timetable and E-timetable define train running speed to be the highest and lowest, respectively, and higher running speed will increase the energy consumption. With higher running speed, T-timetable typically can reduce the passenger waiting time, which is also consistent with our practical experiences. In particular, the obtained timetables by using the LR-based algorithm yield the least passenger waiting times and the best objective values in all these tests. The results also demonstrate the effectiveness of the proposed metro train scheduling approaches with the consideration of time-variant passenger demands. Besides, it is interesting to note that, the reduction of WT by LR-based algorithm is about 15% compared with T-timetable for instance 1, which is much more obvious than the other two instances. One possible reason is that, the passenger demand varies dramatically in morning peak-hours (like instance 1), while it keeps relatively flat in other hours of a day (like instances 2 and 3). In conditions of relatively constant demands, the train timetable with fixed departure headway is potentially a good solution, and therefore, the performance improvement of solutions by LR-based algorithm in such cases is not significantly evident.

For illustration, we display the obtained timetable by LR-based algorithm, E-timetable and T-timetable of instance 1 in Figs. 14, 15 and 16, respectively. It is obvious that, the headway times between adjacent trains for both E-timetable and T-timetable are constant, while the trains are operated with the lowest speed and highest speed by E-timetable and T-timetable, respectively. Differently, the LR-based algorithm derives an irregular timetable, in which the headway time is varied in different time periods. For instance, the headway is a little larger in the first hour (i.e., 7:00–8:00 am), but it becomes much shorter in the next hour (i.e., 8:00–9:00 am). This is essentially caused by the increased passenger demands in this time period. In addition, we note that, the train dwelling time at each station and running time in each segment for the obtained timetable are also different, which can potentially make benefits to the energy consumption and utilization of regenerative energy.
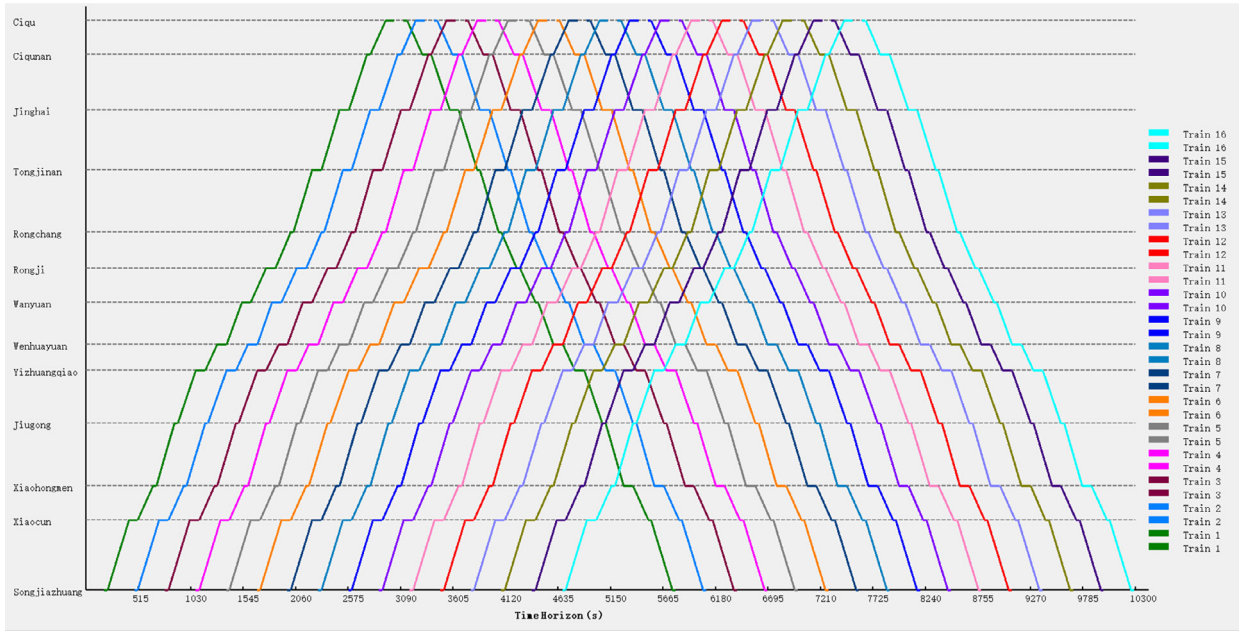
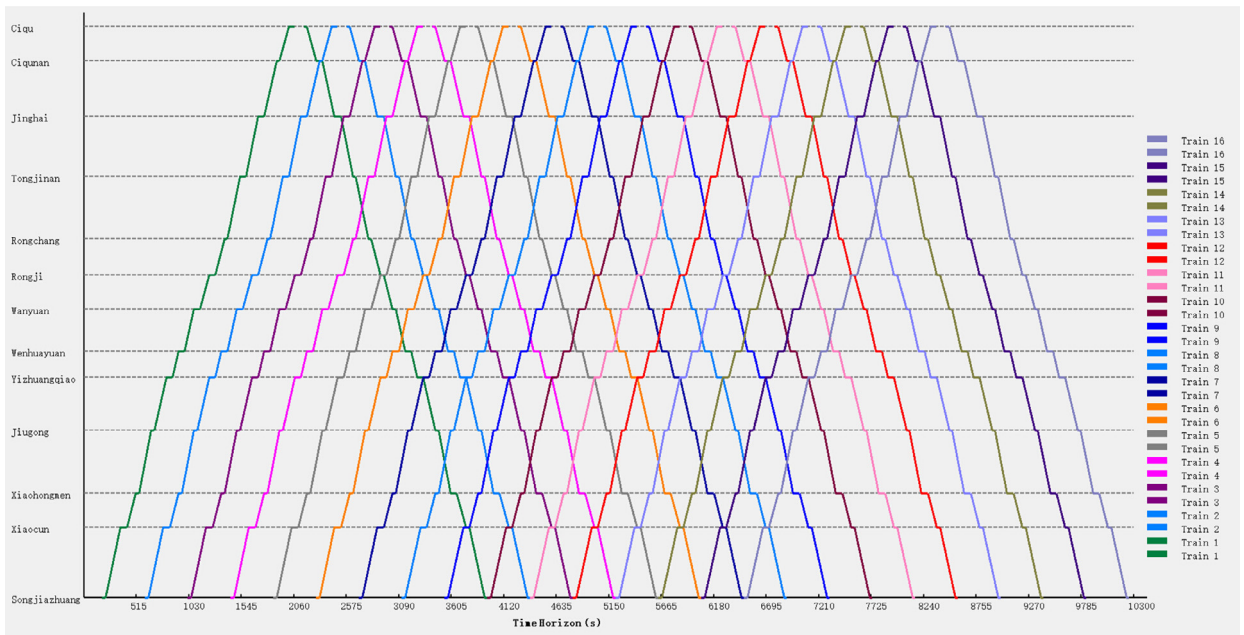**Fig. 15.** Regular timetable with lowest travel speed (E-timetable).

Fig. 16. Regular timetable with highest travel speed (T-timetable).

5.2.3. More experiments with different numbers of trips

In order to further investigate the performance of the generated metro train timetables, we use the LR-based approach to schedule trains with different numbers of trips, in which the parameter settings are the same to those in Section 5.2.1. Also, we are particularly interested in analyzing the variations of passenger waiting time and energy consumption with different numbers of trips. In what follows, we only present the results in the morning period (i.e., 7:00–9:55 am in Table 7) due to the page limitation.

Fig. 17a–d depict the values of total passenger waiting time, total energy consumption, objective functions and computational times, respectively, with different numbers of trips by E-timetable, T-timetable and the timetable developed by LR-based approach. First, we analyze the overall variation tendencies of waiting time and energy consumption with different
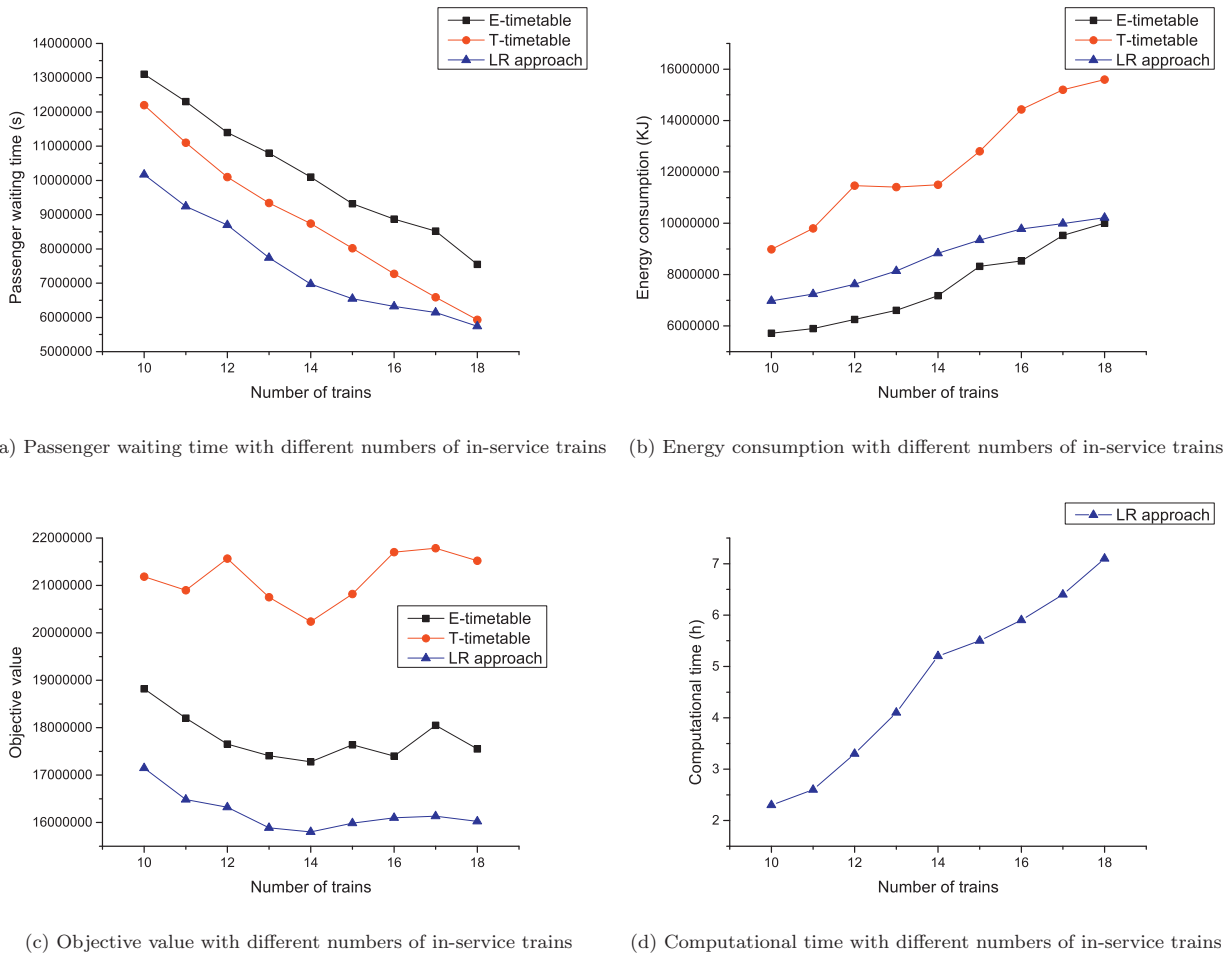
(a) Passenger waiting time with different numbers of in-service trains

(b) Energy consumption with different numbers of in-service trains

(c) Objective value with different numbers of in-service trains

(d) Computational time with different numbers of in-service trains

Fig. 17. Performance value comparison with different numbers of in-service trains.

numbers of trains. In Fig. 17a, it can be seen that, the passenger waiting times for all these three approaches are evidently reduced if we increase the number of in-service trains. This is because that, more trains can shorten the headway time between adjacent trains and thus reduce the passenger waiting time at the stations. Being consistent with our experiences, we can see a different tendency in Fig. 17b that more trains lead to the larger energy consumption. Additionally, the increase of energy consumption is not rigorously linear with respect to the number of in-service trains, as shown in this figure. This is due to the consideration of utilized regenerative energy in the energy consumption calculation.

In addition, comparing these three approaches (i.e., E-timetable, T-timetable and LR approach), we can see that, the value of passenger waiting time by LR-based heuristics is much lower than those of the other two timetables. This indicates that, by adjusting the train speed profiles and departure and arrival times of trains at each station, the passenger waiting time can be reduced in comparison with the timetables with constant headway (i.e., E-timetable and T-timetable). With respect to the energy consumption, it is clear that E-timetable achieves the best performance, since it is scheduled by the slowest speed in every segment. Moreover, the energy consumption of LR-based approach is only a little higher than that of E-timetable, and the energy consumption is nearly the same when the number of in-service trains increases to 17 and 18.

Also, Fig. 17c shows the objective values of E-timetable, T-timetable and the timetable generated by LR-based approach with different numbers of in-service trains. In these results, the performance of developed LR-based approach is typically better than those of E-timetable and T-timetable, which indicates the effectiveness of our proposed approaches. Besides, an interesting phenomenon is that, the best objective values are achieved when the number of in-service trains equals to 14. Thus, it is potentially concluded that, choosing 14 trains in this planning period can better balance the passenger waiting time and energy consumption under the current weight coefficients. The computational times of LR-based approach are shown in Fig. 17d for different instances, in which it is easy to see that the computational time is much short with 10 in-service trains, while it is increased to about 7 hours with 18 trains. This fact shows that, although the developed LR-based approach is able to obtain a good solution in an acceptable time, it is still sensitive to the scale of the model for some large instances.

6. Conclusion

In this paper, we studied the metro train scheduling problem with dynamic demands to reduce the passenger waiting time at stations and energy consumption for train operations. By using a space-time network diagram, we proposed two (mixed) integer linear programming models to capture the time-variant demands and different travel speeds in the calculation of train energy consumption. Specifically, the first model was an integer linear programming model which aims to jointly minimize the passenger waiting time and train traction energy consumption. The second model was formulated as an MILP model that further considers the utilization of regenerative barking energy among trains in the same power supply substation. Due to the complexity of these models, a Lagrangian relaxation-based heuristic algorithm was particularly developed in order to obtain a good solution within an acceptable computational time.

To verify the effectiveness of the proposed train scheduling approach, two sets of numerical examples were implemented, which include a small-scale example and a real-world instance of Beijing metro. The computational results showed that, for small-scale instances, LR-based algorithm can achieve very close or better results in solving the models compared with CPLEX solver, and the computational time is much shorter than that of CPLEX. For large-scale instances or real-world instances, CPLEX solver could not provide even a feasible solution through several hours of computation, while LR-based algorithm is able to obtain good solutions for these cases. Besides, the proposed metro train scheduling approach could evidently reduce the total passenger waiting time especially in peak-hours, and kept relatively low energy consumption in comparison with fixed-headway timetables.

Our future research will focus on the following four major aspects. (1) Although we consider the passenger demands to be dynamic in this study, the real-world passenger characteristics in metro systems can be even more complex, e.g., the day-to-day demand variations. Thus, further analyzing time-variant and random features of metro passenger demands, and consequently extending our work to stochastic optimization models can be one of future research directions. (2) Some constraints in the developed models (e.g., turnaround constraints, dwelling constraints) can be further embedded into the formulated space-time network through introducing relevant space-time arcs, which subsequently requires a more efficient algorithm to speed up the scheduling process for solving large-scale problems. (3) As the computational results have shown that, an appropriate scale of in-service trains can achieve better trade-off between service quality and energy consumption, more analyses on the performance indexes can be taken into consideration in the future researches. (4) In this study, we only consider the train scheduling problem in a single metro corridor. Our further research will also focus on the coordinated timetable scheduling problem for a metro network, in which some other important indicators, such as the waiting time of transfer passengers and utilization of regenerative braking energy from other metro corridors, can be also considered to realize a network-optimization-based operation strategy.

Acknowledgement

References

Albrecht, A., Howlett, P., Peter, P., Vu, X., Zhou, P., 2016. The key principles of optimal train control-part 1: formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points. Transp. Res. Part B 94, 482–508.

Barrena, E., Canca, D., Coelho, L., Laporte, G., 2014. Exact formulations and algorithm for the train timetabling problem with dynamic demand. Comput. Oper. Res. 44, 66–74.

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014. Single-line rail transit timetabling under dynamic passenger demand. Transp. Res. Part B 70, 134–150.

Bektas, T., Laporte, G., 2011. The pollution-routing problem. Transp. Res. Part B 45, 1232–1250.

Boland, N., Hewitt, M., Marshall, L., Savelsbergh, M., 2015. The continuous time service network design problem. Technical Report, Optimization on-line.

Cacchiani, V., Furini, F., Kidd, M., 2016. Approaches to a real-world train timetabling problem in a railway node. Omega 58, 97–110.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transp. Res. Part B 63, 15–37.

Cacchiani, V., Toth, P., 2012. Nominal and robust train timetabling problem. Eur. J. Oper. Res. 219, 727–737.

Canca, D., Barrena, E., Algaba, E., Zarzo, A., 2014. Design and analysis of demand-adapted railway timetables. J. Adv. Transp. 48, 119–137.

Carey, M., Crawford, I., 2007. Scheduling trains on a network of busy complex stations. Transp. Res. Part B 41, 159–178.

Chen, D., Chen, R., Li, Y., Tang, T., 2013. Online learning algorithms for train automatic stop control using precise location data of balises. IEEE Trans. Intell. Transp. Syst. 14 (3), 1526–1535.

Cordeau, J., Toth, P., Vigo, D., 1998. A survey of optimization models for train routing and scheduling. Transp. Sci. 32 (4), 380–404.

Cordone, R., Redaelli, 2011. Optimizing the demand captured by a railway system with a regular timetable. Transp. Res. Part B 45, 430–446.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2010. A tabu search algorithm for rerouting trains during rail operations. Transp. Res. Part B 44, 175–192.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2012. Bi-objective conflict detection and resolution in railway traffic management. Transp. Res. Part C 20, 79–94.

Corman, F., Quaglietta, E., 2014. Closing the loop in real-time railway control: framework design and impacts on operations. Transp. Res. Part C 54, 15–39.

Daganzo, C., 1997. Fundamentals of transportation and traffic operations. Oxford: Pergamon.

Daganzo, C., 2010. Public transportation systems: Basic principles of system design, operations planning and real-time control. No. UCB-ITS-CN-2010-1.

D'Ariano, A., Pacciarelli, D., Pranzo, M., 2007. A branch and bound algorithm for scheduling trains in a railway network. Eur. J. Oper. Res. 183, 643–657.

Dauzère-Pérès, S., Almeida, D.D., Guyon, O., Benhizia, F., 2015. A lagrangian heuristic framework for a real-life integrated planning problem of railway transportation resources. Transp. Res. Part B 74, 138–150.

Dorfman, M.J., Medanic, J., 2004. Scheduling trains on a railway network using a discrete event model of railway traffic. Transp. Res. Part B 38, 81–98.

Freyss, M., Giesen, R., Munoz, J., 2013. Continuous approximation for skip-stop operation in rail transit. Transp. Res. Part C 36, 419–433.

Higgins, A., Kozan, E., Ferreira, L., 1996. Optimal scheduling of trains on a single line track. Transp. Res. Part B 30 (2), 147–161.

Howlett, P.G., Pudney, P.J., 1995. Energy-efficient train control. Springer-Verlag, Berlin.

Howlett, P.G., Pudney, P.J., Vu, X., 2009. Local energy minimization in optimal train control. Automatica 45, 2692–2698.

Hua, Z., Lin, Y., 2014. The Computation Model of ATO-Level Profile. In: Proceedings of the 2013 International Conference on Electrical and Information Technologies for Rail Transportation (EITRT 2013) 1, Springer Berlin Heidelberg, pp. 583–596.

Huang, Y., Yang, L., Tang, T., Cao, F., Gao, Z., 2016. Saving energy and improving service quality: bicriteria train scheduling in urban rail transit systems. IEEE Trans. Intell. Transp. Syst. 17 (12), 3364–3379.

IEEE Standard,. Rail transit vehicle interface standards committee of the IEEE vehicular technology society. IEEE 1474.3^{TM} IEEE recommended practice for communication-based train control (CBTC) system design and functional allocations, USA: IEEE, Inc., 2(23).

Kennedy, C.A., 2002. A comparison of the sustainability of public and private transportation systems: study of the greater toronto area. Transportation 29, 459–493.

Kennington, J.L., Nicholson, C.D., 2010. The uncapacitated time-space fixed-charge network flow problem: an empirical investigation of procedures for arc capacity assignment. INFORMS J. Comput. 22 (2), 326–337.

Kroon, L.G., Peeters, L.W., Wagenaar, J.C., Zuidwijk, R.A., 2014. Flexible connections in PESP models for cyclic passenger railway timetabling. Transp. Sci. 48 (1), 136–154.

Li, P., Mirchandani, P., Zhou, X., 2015. Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork. Transp. Res. Part B 81, 103–130.

Li, X., Lo, H., 2014. An energy-efficient scheduling and speed control approach for metro rail operations. Transp. Res. Part B 64, 73–89.

Li, X.P., Ma, J., Cui, J., Ghiasi, A., Zhou, F., 2016. Design framework of large-scale one-way electric vehicle sharing systems: a continuum approximation model. Transp. Res. Part B 88, 21–45.

Lin, T.-M., Wilson, N.H.M., 1992. Dwell time relationships for light rail systems. Transp. Res. Rec. 1361, 287–295.

Liu, J., Zhou, X., 2016. Capacitated transit service network design with boundedly rational agents. Transp. Res. Part B 93, 225–250.

Liu, R., Golovitcher, I.M., 2003. Energy-efficient operation of rail vehicles. Transp. Res. Part A 37 (10), 917–932.

Lu, C.-C., Liu, J., Qu, Y., Peeta, S., Rouphail, N., Zhou, X., 2016. Eco-system optimal time-dependent flow assignment in a congested network. Transp. Res. Part B 94, 217–239.

Mahmoudi, M., Zhou, X., 2016. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: a dynamic programming approach based on state-space-time network representations. Transp. Res. Part B 89, 19–42.

Meng, L., Zhou, X., 2014. Simultaneous train rerouting and rescheduling on an n-track network: a model reformulation with network-based cumulative flow variables. Transp. Res. Part B 67, 208–234.

Mu, S., Dessouky, M., 2013. Efficient dispatching rules on double tracks with heterogeneous train traffic. Transp. Res. Part B 51, 45–64.

Nasri, A., Moghadam, M.F., Mokhtari, H., 2010. Timetable Optimization for Maximum Usage of Regenerative Energy of Braking in Electrical Railway Systems. In: Proceedings of the International Symposium on Power Electronics, Electrical Drives, Automation and Motion, IEEE Publisher, Pisa, Italy, pp. 1218–1221.

Niu, H., Tian, X., Zhou, X., 2015. Demand-driven train schedule synchronization for high-speed rail lines. IEEE Trans. Intell. Transp. Syst. 16 (5), 2642–2652.

Niu, H., Zhou, X., 2013. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. Transp. Res. Part C, 36, 212–230.

Niu, H., Zhou, X., Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints. Transp. Res. Part B 76, 117–135.

Pelletier, M.P., Trpanier, M., Morency, C., 2011. Smart card data use in public transit: a literature review. Transp. Res. Part C 19, 557–568.

Pena-Alcaraz, M., Fernandez, A., Cucala, A.P., Ramos, A., Pecharroman, R.R., 2012. Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy. Proc. Inst. Mech. Eng., Part F 226 (4), 397–408.

Rajagopalan, S., Heragu, S.S., Taylor, G., 2004. A lagrangian relaxation approach to solving the integrated pick-up/drop-off point and AGV flowpath design problem. Appl. Math. Model. 28, 735–750.

Ramos, A., Pena, M., Fernndez-Cardador, A., Cucala, A.P., 2007. Mathematical Programming Approach to Underground Timetabling Problem for Maximizing Time Synchronization. In: Proceedings of the International Conference on Industrial Engineering and Industrial Management, IEEE Publisher, Madrid, Spain, pp. 88–95.

Shi, N., Song, H., Powell, W.B., 2014. The dynamic fleet management problem with uncertain demand and customer chosen service level. Int. J. Prod. Econ. 148, 110–121.

Sun, L., Jin, J., Lee, D., Axhausen, K.W., Erath, A., 2014. Demand-driven timetable design for metro services. Transp. Res. Part C 46, 284–299.

Tong, L., Zhou, X., Miller, H.J., 2015. Transportation network design for maximizing space-time accessibility. Transp. Res. Part B 81, 555–576.

Wang, Y., Schutter, B.D., van den, B.T.J.J., Ning, B., Tang, T., 2014. Efficient bilevel approach for urban rail transit operation with stop-skipping. IEEE Trans. Intell. Transp. Syst. 15 (6), 2658–2670.

Wang, Y., Tang, T., Ning, B., van den, B.T.J.J., Schutter, B.D., 2015. Passenger-demands-oriented train scheduling for an urban rail transit network. Transp. Res. Part C 60, 1–23.

Wong, R., Yuen, T., Fung, K., Leung, J., 2008. Optimizing timetable synchronization for rail mass transit. Transp. Sci. 42 (1), 57–69.

Xie, W., Ouyang, Y., Wong, S., 2016. Reliable location-routing design under probabilistic facility disruptions. Transp. Sci. 50 (3), 1128–1138.

Xu, X., Li, K., Yang, L., 2015. Scheduling heterogeneous train traffic on double tracks with efficient dispatching rules. Transp. Res. Part B 78, 364–384.

Yang, L., Qi, J., Li, S., Gao, Y., 2016. Collaborative optimization for train scheduling and train stop planning on high-speed railways. Omega 64, 57–76.

Yang, L., Li, S., Gao, Y., Gao, Z., 2015. A coordinated routing model with optimized velocity for train scheduling on a single-track railway line. Int. J. Intell. Syst. 30, 3–22.

Yang, L., Zhang, Y., Li, S., Gao, Y., 2016. A two-stage stochastic optimization model for the transfer activity choice in metro networks. Transp. Res. Part B 83, 271–297.

Yang, L., Zhou, X., 2014. Constraint reformulation and a lagrangian relaxation-based solution algorithm for a least expected time path problem. Transp. Res. Part B 59, 22–44.

Yang, L., Zhou, X., 2017. Optimizing on-time arrival probability and percentile travel time for elementary path finding in time-dependent transportation networks: linear mixed integer programming reformulations. Transp. Res. Part B 96, 68–91.

Yang, L., Zhou, X., Gao, Z., 2014. Credibility-based rescheduling model in a double-track railway network: a fuzzy reliable optimization approach. Omega 48, 75–93.

Yang, X., Chen, A., Li, X., Ning, B., Tang, T., 2015. An energy-efficient scheduling approach to improve the utilization of regenerative energy for metro systems. Transp. Res. Part C 57, 13–29.

Ye, H., Liu, R., 2016. A multiphase optimal control method for multi-train control and scheduling on railway lines. Transp. Res. Part B 93, 377–393.

Yin, J., Chen, D., Li, L., 2014. Intelligent train operation algorithms for subway by expert system and reinforcement learning. IEEE Trans. Intell. Transp. Syst. 15 (4), 1561–1571.

Yin, J., Chen, D., Yang, L., Tang, T., Ran, B., 2016. Efficient real-time train operation algorithms under uncertain passenger demands. IEEE Trans. Intell. Transp. Syst. 17 (9), 2600–2612.

Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: an approximate dynamic programming approach. Transp. Res. Part B 91, 178–210.

Zhou, L., Tong, L., Chen, J., Tang, J., Zhou, X., 2017. Joint optimization of high-speed train timetables and speed profiles: a unified modeling approach using space-time-speed grid networks. Working paper, https://www.researchgate.net/publication/296160123.

Zhou, X., Zhong, M., 2005. Bicriteria train scheduling for high-speed passenger railroad planning applications. Eur. J. Oper. Res. 167 (3), 752–771.

Zhou, X., Zhong, M., 2007. Single-track train timetabling with guaranteed optimality: branch-and-bound algorithms with enhanced lower bounds. Transp. Res. Part B 41, 320–341.