

Research Article

Metro Timetabling for Time-Varying Passenger Demand and Congestion at Stations

Keping Li,¹ Hangfei Huang ,¹ and Paul Schonfeld²

¹State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

²Department of Civil & Environmental Engineering, University of Maryland, College Park, MD 20742, USA

Correspondence should be addressed to Hangfei Huang; 14114200@bjtu.edu.cn

Received 7 November 2017; Revised 4 March 2018; Accepted 11 March 2018; Published 3 July 2018

Academic Editor: Giulio E. Cantarella

Copyright © 2018 Keping Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For the train timetabling problem (TTP) in a metro system, the operator-oriented and passenger-oriented objectives are both important but partly conflicting. This paper aims to minimize both objectives by considering frequency (in the line planning stage) and train cost (in the vehicle scheduling stage). Time-varying passenger demand and train capacity are considered in a nonsmooth, nonconvex programming model, which is transformed into a mixed integer programming model with a discrete time-space graph (DTSG). A novel dwell time determining process considering congestion at stations is proposed, which turns the dwell times into dependent variables. In the solution approach, we decompose the TTP into a subproblem for optimizing segment travel times (OST) and a subproblem for optimizing departure headways from the shunting yard (OH). Branch-and-bound and frequency determining algorithms are designed to solve OST. A novel rolling optimization algorithm is designed to solve OH. The numerical experiments include case studies on a short metro line and Beijing Metro Line 4, as well as sensitivity analyses. The results demonstrate the predictive ability of the model, verify the effectiveness and efficiency of the proposed approach, and provide practical insights for different scenarios, which can be used for decision-making support in daily operations.

1. Introduction

With increasing concerns about urban congestion and climate change, urban metro rail transportation receives increasing attention due to its high capacity, punctuality and sustainability. Metro timetabling is the problem of assigning precise utilization times for infrastructure resources to every train in the metro system [1]. In large cities, a metro system is often the busiest public transportation system. To meet travel demand and passenger satisfaction, a timetable should be as compact and flexible as possible. Meanwhile, since many metro systems need government subsidies to cover operating expenses, the planners focus more on operating cost than on passenger factors when designing a timetable [2]. This motivates some recent studies to combine these two conflicting objectives in the train timetabling problem [3, 4].

Two types of timetables are used in metro systems: cyclic and noncyclic. In a cyclic timetable, the departure times of trains are scheduled in equally spaced cycles (e.g., a half hour

cycle). In a railway system, the cyclic timetables are assumed to be preferred by both operators and passengers because such timetables are easy to operate and remember. However, in a metro system, passengers usually arrive randomly at stations and wait for the next available train, without checking precise timetables beforehand. Besides, as pointed out by Robenek et al. [5], a cyclic timetable provides an inefficient operation as there is a mismatch between the supply (determined by the timetables) and the demand (characterized by the time-varying demand). On the other hand, a noncyclic timetable imposes no special rules on the departure times of trains [6]. The noncyclic timetables are more flexible regarding time-varying passenger demand, especially when the demand is large [2]. Thus, the noncyclic timetabling problem is worth exploring [2, 3, 7].

The planning process in public transportation can be generally split into three stages: line planning, timetabling, and vehicle scheduling [8]. The line planning stage determines the route/station layouts and frequency. Timetabling is based on the output of line planning, after which the vehicle scheduling

can be designed. It is pointed out by Schöbel [8] that going through all these stages sequentially (i.e., independently) leads to unsatisfactory solutions. Since they are interrelated, important factors and objectives at the line planning stage and the vehicle scheduling stage should be accounted for in the timetabling problem.

This paper proposes a rolling optimization algorithm to obtain noncyclic timetables considering time-varying passenger demand and the effects of congestion at stations. It integrates the objectives in the line planning (frequency), timetabling (conflicting objectives including passenger wait/in-vehicle time and energy), and vehicle scheduling (train cost). The main contributions of this paper are as follows.

(1) This paper specifically models the dynamic evolution of passenger loads on trains at each station, by considering passenger arrival rates, limited train capacity and actual passenger alighting/boarding rates associated with congestion. This is an extension of existing studies [9]. Particularly, Niu and Zhou [10] and Niu et al. [11] propose the concept of “effective loading time” that represents the actual time interval within which the arriving passengers can board a train. The detailed boarding and alighting processes are not discussed in their model. Wang et al. [3] set a lower bound (which accounts for details including the number of boarding/alighting passengers and their boarding/alighting rates) for dwell times but consider the dwell times as variables to be optimized, rather than parameters that depend on arrival/departure times and passenger demand. Impacts of gradually increasing passenger demand are analyzed in Robenek et al. [2]; however, in modelling they do not consider the effects of passenger congestion.

(2) This paper integrates objectives from different planning stages in the train timetabling problem, for which solution approaches are scarce [8]. Particularly, Schöbel [8] considers the line planning, timetabling and vehicle scheduling in an integrated way, with passenger- and cost-oriented objectives. However, since his eigenmodel is general and resulting approaches are generic, important objectives described in this paper are not considered in his work, that is, passenger wait time, dwell time, and train capacity. These objectives are included in [2, 5], whereas the frequency and dwell times are fixed in their studies. Besides, energy is not considered in [2, 5].

(3) This paper proposes a novel and effective solution approach, which decomposes the master timetabling model into two subproblems. The first subproblem is solved by branch-and-bound and frequency determining algorithms, and the second one is solved by a rolling optimization algorithm which optimizes the departure headways from the shunting yard (which are equivalent to arrival times of trains at their first station) and considers interdependent variables in each rolling step. The design of decision variables in this paper is similar to [2, 5], but the models and solution approaches are different. Compared to Wang et al. [3], which addresses the similar timetabling problem with traditional approaches such as sequential quadratic programming and a genetic algorithm, the approach proposed in this paper

performs better computationally in terms of run time and solution quality.

The remainder of this paper is as follows. Section 2 reviews relevant studies on the train timetabling problem and summarizes the research gap in the literature. Section 3 first describes the problem, introduces formulation of different cost functions, and constructs the train timetabling model. Then the discrete time-space graph is proposed to transform the model into a mixed integer program (MIP). Section 4 discusses the solution approach for the optimization model. Section 5 presents different numerical experiments and sensitivity analyses on a short metro line and on Beijing Metro Line 4, to demonstrate the predictive ability of the model as well as the effectiveness and efficiency of the proposed approach. Section 6 summarizes the works done in this paper and presents the limitations and potential topics in future studies.

2. Literature Review

The train timetabling problem (TTP) aims to schedule trains to transport passengers (or goods) without conflicts, by specifying train arrival and departure times at stations. It is interrelated with other planning stages, that is, the line planning stage and the vehicle scheduling stage. We organize the related literature from the perspectives of operator-oriented objectives, passenger-oriented objectives, and their integration.

2.1. Operator-Oriented Objectives. In the literature, operator-oriented objectives can be found only for the noncyclic version of the TTP [2]. Most studies focus on the operation indicators, for example, energy [17, 18], train delay [19–21], and train travel time [22–24].

The total train delay and train travel time are usually considered in the TTP for railway networks, which aims to find a feasible timetable by minimizing the profit loss resulting from changes to the ideal/planned timetable. Energy is the focus and main cost for metro system operators. Most energy is consumed in train operations [25, 26], and thus obtaining an energy-efficient train timetable receives considerable attention [17, 18, 27]. Note that energy is affected by the travel time in the segment, and the segment travel time is determined by the arrival and departure times at adjacent stations, that is, the timetable. Hence, the energy-efficient train timetabling problem may be extended to include other objectives. For example, Yang et al. [23] consider both energy and train travel time as optimization objectives. Some recent studies also account for regenerative braking. Yang et al. [28] develop a scheduling approach to coordinate arrivals and departures of trains within the same electricity supply zones, thereby effectively recovering the regenerative energy. Generally, these studies focus solely on the timetabling stage, where frequency (in line planning) is fixed and train cost (in-vehicle scheduling) is not considered.

The efficient use of railway rolling stock (vehicle scheduling) is an important objective pursued by a railway agency or company because of intensive capital investment in rolling stock [29]. To this end, Lai et al. [29] develop an optimization

model to improve the efficiency of rolling stock usage considering necessary regulations and practical constraints, where a hybrid heuristic process is designed to improve solution quality and efficiency. Haahr et al. [27] use CPLEX and a column and row generation approach to assign rolling stock units to timetable services in passenger railways, prepare daily schedules, and check their real-time applicability by testing different disruption scenarios. These vehicle scheduling studies require the trips as input data. Based on Schöbel [8], in a metro system, every trip describes the operation of a train between the start and end time of the line at its first and its last stations (given from the timetable). The objective function then aims at minimizing the number of trains and the costs for their movements.

Particularly, Schmid and Ehmke [30] and Schöbel [8] demonstrate that the integration of timetabling and vehicle scheduling is more beneficial than the sequential planning process. However, the studies on the integration of timetabling and vehicle scheduling usually adopt relatively general and generic models, and the TTP model is often simplified [31]. For example, Cadarso and Marin [32] focus on shunting operations and their timetable is calculated by readjusting frequencies, whereas specific details such as the dwell times and segment travel times are not considered. This motivates us to consider train cost in the timetabling stage, while focusing on the specified TTP formulation.

In addition, the conflict-free timetable is also pursued by the operators [33, 34]. It is worth mentioning that the discrete time-space graph proposed in Caprara et al. [6] is a directed multigraph in which nodes correspond to arrivals/departures at a certain station and at a given time stamp. This graph is widely used to formulate the TTP and derive different integer programming models that correspond to specific objectives. For example, Caprara et al. [6] use the graph to derive an integer linear programming (ILP) model with Lagrangian relaxation, which is embedded within a heuristic algorithm. Cacchiani et al. [12] use the graph to formulate an ILP with linear programming (LP) relaxation. Cacchiani et al. [35] use the time-space graph-based LP relaxation of an ILP to derive a dual bound in the TTP for a set of stations in an urban area interconnected by tracks, thus aiming to resolve the conflicts and evaluate the capacity saturation.

2.2. Passenger-Oriented Objectives. For a metro system, passenger-oriented timetables that consider reliability and reduction of passenger time are most desirable [11]. In some studies, the passenger demand is considered stable [36]. Assuming that passengers prefer easily memorable timetables, such timetables are usually cyclic. They are designed on the basis of a period event scheduling problem (PESP), aiming to minimize passenger travel and wait time [15, 37] and to maximize network stability [38]. However, as mentioned above, metro passengers arrive randomly and do not remember the timetable. Besides, cyclic timetables are less flexible than noncyclic ones in accounting for the passenger demand [2].

Recent studies consider time-varying demands due to the growing concerns for service level and congestion at stations. Such timetables are noncyclic, with objectives of improving

passenger satisfaction [3, 7, 10]. Particularly, Barrena et al. [7] propose two nonlinear programming (NP) formulations to generalize noncyclic train timetables on a single line, which are solved by a fast adaptive large neighborhood search meta-heuristic. The objective is to minimize passenger wait times at stations. Niu et al. [11] construct mathematically rigorous and algorithmically tractable nonlinear mixed integer programming (MIP) models for both real-time scheduling and medium-term planning applications, to jointly synchronize effective passenger loading time windows and train arrival and departure times at each station. Their work aims to minimize the total waiting times of passengers at stations. Robenek et al. [5] define a timetable as a set of departure times of every train from its origin station on every line and consider four attributes in passenger satisfaction: the in-vehicle-time, the waiting time at transfers, the number of transfers, and the schedule passenger delay. They describe the hybrid timetables through additional constraints that are imposed on the original passenger centric train timetabling problem [2], so that the passengers would obtain the same level of service as a cyclic timetable, with more flexibility. A specifically defined simulated annealing heuristic is proposed to solve the problem. Luan et al. [16] integrate the TTP and preventive maintenance time slots (PMTSs) planning problem on a general railway network. They propose an MILP model to jointly optimize train routes, orders and passing times at each station, and PMTSs. The objective is to minimize the sum of the absolute arrival time deviations of real trains at destinations between the ideal and actual timetables. A Lagrangian-based label correcting algorithm is designed for solving the time-dependent least cost path problem.

In addition, the delay management (DM) problem determines whether trains should wait for a delayed feeder train or should depart on time, while considering passenger-oriented objectives [13]. Some recent studies integrate the macroscopic DM with the microscopic TTP [39]. Particularly, Dollevoet et al. [14] formulate an integer programming (IP) model and propose an iterative optimization approach to solve such a bilevel problem that the macroscopic level is the delay management and the microscopic level is the train timetabling. The optimization approach repeats a process that uses DM to find a solution and uses TTP to validate it, until a feasible solution is found. It should be noted that the graph-based models are adopted in their work, which simplify the formulations: the TTP is formulated with the alternative graph [40], and the DM is formulated with an event-activity network [15].

2.3. Integrated Objectives. Since the operator- and passenger-oriented objectives are partly conflicting, focusing on only one side would yield undesirable solutions; for example, the best possible service for passengers may also be the most costly alternative for the operator [2]. Thus, many TTP studies integrate these objectives jointly [22, 41].

Wang et al. [3] aim to minimize passenger travel time and energy in their nonsmooth nonconvex programming (NSNCP) model. They propose an event-driven model that involves arrival and departure events and passenger

arrival rates change events and use nonlinear programming approaches and evolutionary algorithms to solve the problem.

Qi et al. [42] focus on a line planning problem with budget constraints and evaluate the station layout with train timetable indicators. The objectives include constructing cost for additional tracks, total travel time, and network capacity. A mixed integer linear programming (MILP) model is formulated to address the TTP, and then a bilevel programming (BP) model is designed to address the integrated problem. Both CPLEX and a local search-based heuristic are developed to solve the models.

Robenek et al. [2] model the PCTTP as an MILP model with objectives of maximizing the operator's profit while maintaining passenger satisfaction. The model uses the output of the line planning problem as the input of the train timetabling problem, to address the gap found between them. CPLEX is used to construct a Pareto Frontier of the contradictory objectives.

Li et al. [43] combine dynamic train regulation and passenger flow control to minimize the timetable and the headway deviations for metro lines, thereby reducing the operator profit loss and the passenger delay. The model predictive control and a quadratic programming algorithm are proposed to solve the problem, and stability analysis is conducted to verify the system convergence.

However, the above studies do not consider the impacts of frequency. Schöbel [44] reviews some approaches to model and solve the line planning problem, which demonstrate the impacts of frequency to both operator- and passenger-oriented objectives. Particularly, Samanta and Jha [45] consider different objectives in the line planning stage which include minimizing user cost, operator cost, and location cost and maximizing the ridership or the service coverage of the line. The objectives are separately formulated with NP models, and a GA is developed to optimize the objective functions subject to constraints on the number and spacing of stations. Fu et al. [46] propose a BP model to determine line frequencies and the stopping patterns for a mix of fast and slow train lines, with the objectives of minimizing passenger travel time and the number of transfers and maximizing train capacity occupancy. The heuristics solve the problem by combining additional conflicting demand-supply factors. Cascetta and Coppola [47] specify both the frequency-based assignment models and the timetable-based models in the demand forecasting, to analyze their differences in modal split estimates and flows on individual trains. The results show that when passenger demand is time-varying or the timetable is noncyclic, the difference is significant. Thus, the integration of frequency and timetabling is worth investigating.

2.4. Research Gap. Table 1 highlights some publications related to the TTP. All the graph-based models use integer programs, including the model in this paper. Although different objectives are considered, very few studies integrate the objectives in all three stages. Particularly, Wang et al. [3] only consider the objectives in one stage, that is, timetabling stage; Robenek et al. [2] integrate passenger satisfaction and

operating profit (including train cost) but do not consider frequency and energy; Schöbel [8] takes into account objectives in all stages with an iterating framework; however, crucial factors such as passenger wait time, dwell time, and train capacity are missing. Kroon et al. [15] integrate the objectives in the TTP and vehicle scheduling but omit energy.

Apart from the studies focusing on the delay management, most TTP-related studies optimize arrival and departure times of trains at each station; that is, the dwell times are optimized in the solution approach. However, the dwell times are heavily affected by the number of boarding and alighting passengers and their boarding and alighting rates, associated with congestion at stations. Thus, they should be considered as parameters that are dependent on other decision variables, that is, the departure times from the first station [2, 5] and the segment travel times. In this paper, we account for these characteristics and introduce specific equations to determine the dwell times in different scenarios. Such a process for determining dwell times has not been found in the literature.

Finally, heuristics and exact approaches are designed to find solutions based on the properties of the models. Particularly, rolling horizon [3] and iterative optimization [8, 14] are used to solve problems that have similar properties to our proposed model. However, the boundary constraints of the rolling horizon limit the solution flexibility of our model; the iterative optimization considers different stages with same importance, while this paper focuses only on the timetabling stage. Besides, the algorithms used in the rolling horizon and iterative optimization are relatively generic [3, 8]. Thus, a more specific algorithm is desirable for better solutions. In this regard, we develop a novel rolling optimization approach to solve the model, which combines the benefits of rolling horizon and iterative optimization.

3. Model Formulation

In this section, we present a mixed integer programming formulation for the TTP. It is a multiobjective optimization problem. The weighted-sum scalarization of different objective functions in terms of generalized monetary cost is used to resolve the tradeoffs among them [8].

3.1. Problem Formulation. Table 2 summarizes the parameters and variables used in this paper. Some are also explained further in the text. The model's simplifying assumptions are listed below and also explained in the problem description.

Assumption 1. Within one period, trains travel through each segment with the same speed profile; that is, the travel times for different trains in a segment are equal.

Assumption 2. We assume that the last train arrives at the first station at the end time of the period, to ensure that no late-arriving passengers will be left behind.

Assumption 3. Passenger arrival rate $\lambda_i(t)$ should not exceed total passenger boarding rate $\zeta_b \cdot n_d$. If $\lambda_i(t) > \zeta_b \cdot n_d$, the queue of waiting passengers increases even if a train is dwelling at

TABLE I: Summary of publications related to the TTP.

Work	Model	Objectives	Decision variable	Solution approach
Caprara et al. [6]	ILP	Maximize profit	Arrival/departure times	Lagrangian-based heuristic
Cacchiani et al. [12]	ILP	Maximize profit	A full timetable of a train	Column generation, decomposition, and heuristic
Dollevoet et al. [13]	IP	Minimize total passenger delay	If a connection is maintained/used	Modified Dijkstra
Niu and Zhou [10]	IP	Minimize passenger wait time	Arrival/departure times	GA
Barrena et al. [7]	LP	Minimize passenger wait time	Arrival/departure times	Branch and bound
Dollevoet et al. [14]	IP	Minimize (1) passenger arrival times, (2) maximize consecutive delay	Connection and arrival/departure times	Iterative optimization approach
Kroon et al. [15]	MIP	Minimize the number of trains, passenger transfer time, and total travel time	Arrival/departure times	CPLEX
Niu et al. [11]	MIP	Minimize passenger wait time	Arrival/departure times	GAMS
Wang et al. [3]	NSNCP	Minimize passenger travel time and energy	Arrival/departure times	Rolling horizon with SQP and GA
Robenek et al. [2]	MILP	Maximize operating profit while maintaining passenger satisfaction	Departure times at first station	CPLEX
Robenek et al. [5]	MILP	Maximize passenger satisfaction	Departure times at first station	Simulated Annealing (SA) heuristic
Schöbel [8]	Eigenmodel	Passenger- and operator-oriented objectives in all three planning stages	Frequency, A/D times, trip assignment	Iterating algorithm
Luan et al. [16]	MILP	Minimize absolute arrival time deviations	A/D times, route selection	Lagrangian relaxation
Yin et al. [4]	MILP	Minimize passenger wait time and energy	A/D times, train control	Lagrangian-based heuristic
This paper	MIP	Minimize passenger wait and travel time, energy, and train cost	Segment travel times and arrival times at the first station	Branch and bound, Rolling optimization

the station. In practice, passenger entrance flow control* is taken to avoid such a case.

*Remark**. At most busy metro stations in China, passenger entrance flow control is adopted during peak hours to ensure operation safety, and thus the passenger arrival rate is a restricted number, which is expected not to exceed passenger boarding rate. Assumption 3 is introduced for consistency with practical experience.

Assumption 4. All passengers in train k who have reached their destination are assumed to exit the train at its arrival time of station i , that is, $t_{k,i}^0$. All waiting passengers at station i able to board train k are assumed to board the train at its departure time, that is, $t_{k,i}^b$. In other words, the number of in-vehicle passengers $P_{k,t}^v$ changes only at the arrival/departure times of train k at stations.

This paper focuses on a bidirectional metro line with a single track, as shown in Figure 1. Properties of such a metro

line are described in [48]. The timetabling problem is studied within a given period. The time-varying passenger demand is given as input, and a series of trains (whose frequency is to be optimized) are scheduled to transport the passengers.

The arrival times at the first station are decision variables that must be optimized. The departure times at the first station are determined by the dwell times. As mentioned earlier, this paper obtains the dwell times based on trains' conditions when they arrive at stations. In this regard, optimizing arrival times at the first station is equivalent to optimizing departure times at the first station, as in [2, 5]. For the second station, the arrival times are determined by the segment travel time in the first segment. Practically, for the metro operators, uniform segment travel times are preferred within one period, because they not only make the system more stable and easier to manage but also simplify synchronized passenger transfers with connecting transit lines. Assumption 1 is introduced to consider this preference. Thus, within one period, the segment travel time in each segment needs to be optimized only once (for all trains within this period). The speed profile

TABLE 2: Parameters and variables.

Name	Description	Units	Purpose
\mathcal{F}	Set of stations	-	Set
\mathcal{K}	Set of trains	-	Set
\mathcal{U}^i	Set of time stamps in which trains can arrive at station i	-	Set
\mathcal{W}^i	Set of time stamps in which trains departure from station i	-	Set
θ	Shunting yard of the metro line	-	Symbol
i, j	Station index	-	Symbol
I	Total number of stations	-	Symbol
k	Train index	-	Symbol
K	Total number of trains within a given period/Frequency	-	Symbol
t	Time index	-	Symbol
u_k	Arrival node of train k	-	Symbol
w_k	Departure node of train k	-	Symbol
$\pi(u_k)$	Time stamp associated with node u_k (or w_k)	-	Symbol
$\Delta(u_k, w_{k'})$	Time duration between two nodes u_k and $w_{k'}$ ($k = k'$ is allowed)	-	Symbol
R^k	Path of train k	-	Symbol
T	Time duration	-	Symbol
$\lfloor a \rfloor$	The largest integer that is smaller than real number a	-	Symbol
$\lceil a \rceil$	The smallest integer that is larger than real number a	-	Symbol
$\vec{\mathbf{b}}$	A vector (represented by any bold alphabet)	-	Symbol
\mathbb{M}	A matrix (represented by any double-struck alphabet)	-	Symbol
$\mathbb{M}_{a:b,x:y}$	A sub-matrix whose elements correspond to those in matrix \mathbb{M} , from rows a to b , and from columns from x to y	-	Symbol
$\ \mathbb{M}\ $	Sum of all (numerical) elements in matrix \mathbb{M}	-	Symbol
c_e	Average energy cost	¥/Kwh	Parameter
c_o	Average train operating cost	¥/(train-km)	Parameter
c_{pt}	Average passenger travel time cost	¥/(passenger-hour)	Parameter
c_{pw}	Average passenger wait time cost	¥/(passenger-hour)	Parameter
c_{tra}	Average capital cost of one train	¥/(train-hour)	Parameter
C_{total}	Total cost of the whole system	¥	Parameter
C_{total}^e	Total cost of energy consumption	¥	Parameter
C_{total}^p	Total cost of passenger time	¥	Parameter
C_{total}^{tra}	Total cost of train (rolling stock)	¥	Parameter
\bar{d}_i	Maximum allowable dwell time at station i	Second	Parameter
\underline{d}_i	Minimum allowable dwell time at station i	Second	Parameter
e_i	Energy consumption per unit mass for one train in segment i	Kwh/Kg	Parameter
f_i	Tractive force per unit mass	m/s ²	Parameter
f_r	Running resistance per unit mass	m/s ²	Parameter
F_s	Fleet size (the number of trains)	Trains	Parameter
h_{min}	Minimum headway	Second	Parameter
M_k	Net weight of a train	Kg/train	Parameter
M_p	Average weight of a passenger	Kg/passenger	Parameter
n_d	Number of doors per train	-	Parameter
N_k	Capacity of train k	Passengers	Parameter

TABLE 2: Continued.

Name	Description	Units	Purpose
$P_{k,i}^a$	Number of passengers alighting train k at station i	Passengers	Parameter
$P_{k,i}^b$	Number of passengers boarding train k at station i	Passengers	Parameter
$P_{k,(i,j)}^b$	Number of passengers boarding train k at station i with destination j	Passengers	Parameter
$P_i(t_1, t_2)$	Number of passengers arriving at station i during $[t_1, t_2]$	Passengers	Parameter
$P_{(i,j)}(t_1, t_2)$	Number of passengers arriving at station i with destination j during $[t_1, t_2]$	Passengers	Parameter
$P_{i,t}^w$	Number of passengers waiting at station i at time t	Passengers	Parameter
$P_{k,t}^v$	Number of passengers in train k at time t	Passengers	Parameter
$s_i(i)$	Length of segment i (between station i and $i + 1$)	Meter	Parameter
$s_t^{\max}(i)$	Maximum segment travel time for segment i	Second	Parameter
$s_t^{\min}(i)$	Minimum segment travel time for segment i	Second	Parameter
$t_{k,i}^0$	Arrival time of train k at station i	Second	Parameter
$t_{k,i}^a$	Alighting end time for train k at station i	Second	Parameter
$t_{k,i}^b$	Boarding end time for train k at station i /Departure time	Second	Parameter
t_{end}^p	End time of passenger demand for a period	Second	Parameter
t_{end}^t	End time of a complete timetable for a period	Second	Parameter
T_{round}^k	Round trip time (total travel time from the departure from the shunting yard to the arrival at the shunting yard) of train k	Second	Parameter
T_{pt}	Total passenger travel time	Second	Parameter
T_{pw}	Total passenger wait time	Second	Parameter
T_{turn}	Turnaround time	Second	Parameter
$v_k(t)$	Speed of train k at time t	m/s	Parameter
$x_k(t)$	Position of train k at time t	Meter	Parameter
$\tau_{i,j}(t)$	Passenger arrival rate at station i with destination j at time t	Passengers/second	Parameter
$\lambda_i(t)$	Passenger arrival rate at station i at time t	Passengers/second	Parameter
ζ_a	Passenger alighting rate	Passengers/second	Parameter
ζ_b	Passenger boarding rate	Passengers/second	Parameter
χ_x	Parameter for energy approximation	-	Parameter
χ_y	Parameter for energy approximation	-	Parameter
$s_t(i)$	Segment travel time of segment i	Second	Decision
$t_{k,1}^0$	Arrival time of train k at the first station	Second	Decision

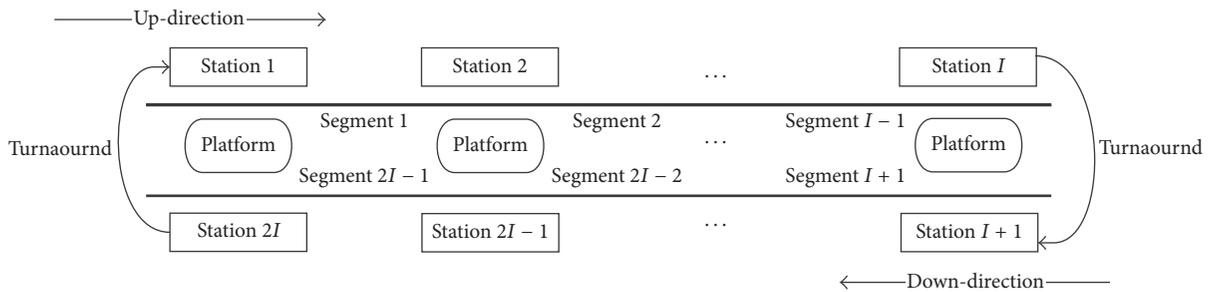


FIGURE 1: A bidirectional metro line with 2I stations.

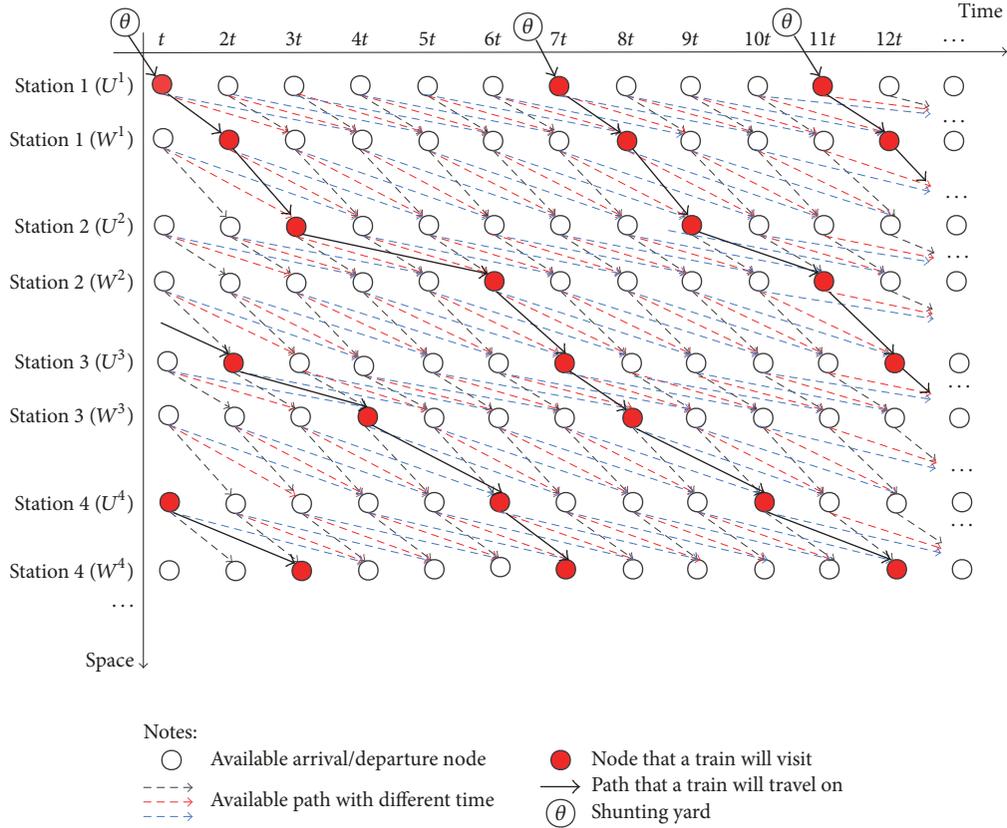


FIGURE 2: Representation of discrete time-space graph.

optimization is outside the scope of this paper. The departure times at the second station are also determined by the dwell time determining process.

Analogously, for subsequent train arrivals and departures, the arrival times are determined by the previous segment travel time, and the departure times are determined by the dwell times. Thus, within a period, the arrival times at the first station and the segment travel times determine the train timetable.

To represent the TTP, a directed multigraph called discrete time-space graph (DTSG) is proposed [6]. As shown in Figure 2, θ represents the shunting yard of the metro line and \mathcal{U}^i (\mathcal{W}^i) corresponds to the set of time stamps in which trains can arrive at (depart from) station i . The nodes in \mathcal{U}^i (\mathcal{W}^i) are called arrival (departure) nodes. Unlike the original graph that is designed for a railway network [6, 12], this paper considers a single-track metro line, and thus the constraints are different and implicitly imposed in the definition of the structure of the DTSG.

The continuous time is discretized into evenly spaced time intervals (1 second in this paper), and the period extends from 0 to t_{end}^t . Note that the values of t_{end}^t and t_{end}^p are different. t_{end}^p is the last time stamp of the passenger demand. The period mentioned above refers to the one extending from 0 to t_{end}^p . Assumption 2 imposes that the last train departs from the first station at time t_{end}^p . However, if the DTSG only considers the period from 0 to t_{end}^p , the arrival/departures of

the last train at subsequent stations are not considered, and the passenger-oriented objectives are hard to formulate. Thus, the DTSG considers the period from 0 to t_{end}^t . The relation between t_{end}^t and t_{end}^p is set as:

$$t_{\text{end}}^t = t_{\text{end}}^p + \sum_{i=1}^{2I-1} s_t(i) + \sum_{i=1}^{2I} \bar{d}_i, \quad (1)$$

where $s_t(i)$ is the segment travel time of segment i and \bar{d}_i is the maximum dwell time of station i . This setting ensures that even the slowest train can be completely included in the DTSG.

The construction of the DTSG is noncyclic. The objective functions of passenger time, energy, and train cost are associated with the chosen arcs and are described below.

3.2. Passenger Time

3.2.1. Passenger Arriving, Alighting, and Boarding

(1) *Time-Varying Demand Information.* The passenger demand at each station can be expressed by $\tau_{i,j}(t)$, which is defined as

$$\begin{aligned} \tau_{i,j}(t) &\geq 0, & \text{If } 1 \leq i < j \leq I \text{ or } I \leq i < j \leq 2I; \\ \tau_{i,j}(t) &= 0, & \text{Otherwise;} \\ & & \forall i, j \in \mathcal{S}. \end{aligned} \quad (2)$$

Passenger arrival rates at stations are $\lambda_i(t) = \sum_{j=i+1}^{2I} \tau_{i,j}(t)$, $\forall i, j \in \mathcal{J}$. In this paper, demands are regarded as deterministic values, which are given by $\tau_{i,j}(t)$. Thus, the number of arriving passengers at station i with destination j during time interval $[t_1, t_2]$ is

$$P_{(i,j)}(t_1, t_2) = \sum_{t=t_1}^{t_2} \tau_{i,j}(t), \quad \forall i, j \in \mathcal{J}. \quad (3)$$

The number of arriving passengers at station i within time interval $[t_1, t_2]$ is the sum of arriving passengers with different destinations:

$$\begin{aligned} P_i(t_1, t_2) &= \sum_{j=i+1}^{2I} P_{(i,j)}(t_1, t_2) = \sum_{j=i+1}^{2I} \sum_{t=t_1}^{t_2} \tau_{i,j}(t) \\ &= \sum_{t=t_1}^{t_2} \lambda_i(t), \quad \forall i, j \in \mathcal{J}. \end{aligned} \quad (4)$$

(2) *Determination of Dwell Arcs.* Letting n_d be the number of doors per train, the average number of alighting (boarding) passengers per door when train k arrives at (departs from) station i is $P_{k,i}^a/n_d$ ($P_{k,i}^b/n_d$). In addition, we denote the average number of boarding passengers per second (boarding rate) as ζ_b and alighting rate as ζ_a , which vary with passenger density in trains. For simplicity, we consider that boarding passengers wait until all alighting passengers get off the train; that is, the boarding process starts at time $t_{k,i}^a$.

The dwell time is the sum of alighting and boarding times. The calculation of alighting time is solely based on the number of alighting passengers:

$$t_{k,i}^a = t_{k,i}^0 + \frac{P_{k,i}^a}{(n_d \cdot \zeta_a)}, \quad \forall k \in \mathcal{K}, i \in \mathcal{J}. \quad (5)$$

The calculation of boarding time is slightly more complex. Figure 3 shows the relations of train remaining capacity, waiting passengers, boarding rate, and max/min dwell times. The horizontal axis represents time, and the vertical axis represents the number of passengers. These four factors determine the boarding time for train k at station i , and in Figure 3 they are represented by (1) remaining capacity of train k at station i : $N_k - P_{k,t}^v$, a horizontal line; (2) number of waiting passengers at station i at time t , a curve $P_{i,t}^w$; (3) maximum allowable number of boarding passengers, a diagonal function $f_k(t)$ with the slope of $n_d \cdot \zeta_b$; (4) minimum and maximum dwell times at station i , that is, $t_{k,i}^0 + \underline{d}_i$ and $t_{k,i}^0 + \bar{d}_i$. In Figure 3, the diagonal $P = f_k(t)$ intersects the horizontal line $P = N_k - P_{k,t}^v$ at time $t_1^{k,i}$ and intersects the curve $P = P_{i,t}^w$ at time $t_2^{k,i}$. The curve $P = P_{i,t}^w$ intersects the horizontal line $P = N_k - P_{k,t}^v$ at time $t_3^{k,i}$. Note that the time is discrete, and Assumption 3 ensures that the slope of $P = f_k(t)$ always exceeds the slope of $P = P_{i,t}^w$. Thus, we have

$$t_1^{k,i} = \max \left\{ t \mid (t - t_{k,i}^a) \cdot n_d \cdot \zeta_b \leq N_k - P_{k,t}^v \right\}, \quad (6)$$

$\forall k \in \mathcal{K}, i \in \mathcal{J}$,

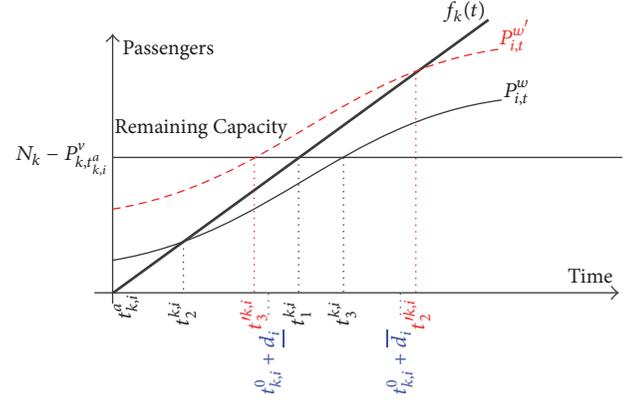


FIGURE 3: Relations of remaining capacity, waiting passengers, boarding rate, and max/min dwell times.

$$t_2^{k,i} = \max \left\{ t \mid (t - t_{k,i}^a) \cdot n_d \cdot \zeta_b \leq P_{i,t}^w \right\}, \quad (7)$$

$\forall k \in \mathcal{K}, i \in \mathcal{J}$,

$$t_3^{k,i} = \max \left\{ t \mid P_{i,t}^w \leq N_k - P_{k,t}^v \right\}, \quad \forall k \in \mathcal{K}, i \in \mathcal{J}. \quad (8)$$

Practically, in an oversaturated scenario, a train departs from a station as soon as the number of in-train passengers reaches its capacity. Based on this strategy, (6) shows that the boarding passengers who arrive after $t_1^{k,i}$ exceed train capacity, and they will be left behind; (7) shows that before $t_2^{k,i}$ there are passengers waiting at station, and after $t_2^{k,i}$ all passengers board the train; (8) points that the train is full at $t_3^{k,i}$, and passengers arriving after $t_3^{k,i}$ should wait for the next train.

The departure time of train k at station i ($t_{k,i}^b$) is determined based on the relations among $t_1^{k,i}$, $t_2^{k,i}$, $t_3^{k,i}$, $t_{k,i}^0 + \underline{d}_i$, and $t_{k,i}^0 + \bar{d}_i$. Based on Assumption 3, we consider two cases: (1) $t_1^{k,i} > t_2^{k,i}$ and (2) $t_1^{k,i} \leq t_2^{k,i}$. Note that \underline{d}_i and \bar{d}_i should be considered in both cases. In case (1), we can further derive that $t_3^{k,i} > t_1^{k,i} > t_2^{k,i}$ (which is the solid curve in Figure 3), where the number of boarding passengers per second is determined by passenger arrival rate. Thus, the departure time is determined as $t_{k,i}^b = \min\{t_{k,i}^0 + \bar{d}_i, \max\{t_2^{k,i}, t_{k,i}^0 + \underline{d}_i, t_3^{k,i}\}\}$. In case (2), we can further derive that $t_3^{k,i} \leq t_1^{k,i} \leq t_2^{k,i}$ (which is the dashed curve in Figure 3), where the number of boarding passengers per second is determined by the actual passenger boarding rate. Thus, the departure time is determined as $t_{k,i}^b = \min\{t_{k,i}^0 + \bar{d}_i, t_1^{k,i}\}$. The departure time of train k at station i is expressed as

$$t_{k,i}^b = \begin{cases} \min \{t_{k,i}^0 + \bar{d}_i, \max \{t_2^{k,i}, t_{k,i}^0 + \underline{d}_i, t_3^{k,i}\}\}, & t_1^{k,i} > t_2^{k,i} \\ \min \{t_{k,i}^0 + \bar{d}_i, t_1^{k,i}\}, & t_1^{k,i} \leq t_2^{k,i}. \end{cases} \quad (9)$$

Finally, we consider the relations between alighting/boarding rates and congestion at stations. Here the congestion is described by the number of in-vehicle passengers.

Practically, the greater the passenger density in a train is, the slower the passengers can get off/on the train. Here, we design a monotonically decreasing function $\rho(P_{k,t}^v) \leq 1$ with respect to the number of in-vehicle passengers, to approximate actual alighting/boarding rates according to the passenger density in trains. For simplicity, in this paper, the value of $\rho(P_{k,t}^v)$ during the dwell process is fixed and is determined by the number of in-vehicle passengers at its arrival time. That is, ζ_a and ζ_b remain constant during the dwell process:

$$\begin{aligned}\zeta_a &= \rho\left(P_{k,t_{k,i}^0}\right) \cdot \zeta_a^{\text{ideal}} \\ \zeta_b &= \rho\left(P_{k,t_{k,i}^0}\right) \cdot \zeta_b^{\text{ideal}},\end{aligned}\quad (10)$$

$$\forall k \in \mathcal{K}, i \in \mathcal{S},$$

where ζ_a^{ideal} and ζ_b^{ideal} are the ideal alighting and boarding rates in the uncongested condition.

(3) *Passenger Alighting and Boarding.* Once the dwell arcs are determined, the alighting and boarding times for train k at station i are determined. Based on the results of dwell arcs, three cases should be considered: (1) $t_1^{k,i} < t_2^{k,i}$, (2) $t_1^{k,i} \geq t_2^{k,i}$ and $t_2^{k,i} \leq t_{k,i}^0 + d_i$; (3) $t_1^{k,i} \geq t_2^{k,i}$ and $t_2^{k,i} > t_{k,i}^0 + d_i$. In case (1), waiting passengers board the train at rate ζ_b . In case (2), actual passenger boarding rate is equal to the passenger arrival rate. In case (3), the actual boarding rate is ζ_b . Thus, the number of boarding passengers for train k at station i is

$$\begin{aligned}P_{k,i}^b &= \begin{cases} P_{i,t_{k,i}^b}^w; & \text{if } t_1^{k,i} \geq t_2^{k,i}, t_2^{k,i} \leq t_{k,i}^0 + d_i; \\ (t_{k,i}^b - t_{k,i}^a) \cdot \zeta_b \cdot n_d; & \text{otherwise;} \end{cases} \quad (11) \\ &\forall k \in \mathcal{K}, i \in \mathcal{S}.\end{aligned}$$

Wang et al. [3] point out that, in a busy metro line, passengers with different destinations are well mixed at each station. Thus, the ratio between the numbers of boarding passengers with destination j and of total boarding passengers is equivalent to the ratio between $\tau_{i,j}(t)$ and $\lambda_i(t)$:

$$\begin{aligned}P_{k,(i,j)}^b &= \frac{\sum_{t=t_{k-1,i}^b}^{t_{k,i}^b} \tau_{i,j}(t)}{\sum_{t=t_{k-1,i}^b}^{t_{k,i}^b} \lambda_i(t)} \cdot P_{k,i}^b; \\ &\forall k \in \mathcal{K}, j \in \mathcal{S} \setminus 1, i = 1, 2, \dots, j-1.\end{aligned}\quad (12)$$

The number of alighting passengers from train k at station j consists of all previously boarding passengers in train k with destination j :

$$P_{k,j}^a = \sum_{i=1}^{j-1} P_{k,(i,j)}^b; \quad \forall k \in \mathcal{K}, j = 2, \dots, I, I+2, \dots, 2I. \quad (13)$$

There are no alighting passengers at the first station in each direction (station 1 and $I+1$).

3.2.2. Passenger Wait and Travel Time

(1) *Passenger Wait Time.* Total passenger wait time at a station can be calculated by summing up the products of the cumulative number of passengers and their time. Let k_{last} be the index of last train that has departed from station i before time t , that is, $t_{k_{\text{last}},i}^b \leq t < t_{k_{\text{last}+1},i}^b$. The cumulative number of passengers at station i at time t is the difference between cumulative numbers of arrival passengers and boarding passengers within the period $[0, t]$:

$$P_{i,t}^w = P_i(0, t) - \sum_{k=1}^{k_{\text{last}}} P_{k,i}^b; \quad \forall i \in \mathcal{S}. \quad (14)$$

Since the continuous time is discretized into 1 second intervals, the passenger wait time between times t and $t+1$ is $P_{i,t}^w \cdot (t+1-t) = P_{i,t}^w$. Thus, the total passenger wait time is

$$T_{pw} = \sum_{i \in \mathcal{S}} \sum_{t=0}^{t_{\text{end}}} P_{i,t}^w. \quad (15)$$

(2) *Passenger Travel Time.* The number of in-train passengers in train k at time t is the difference between cumulative numbers of boarding passengers and alighting passengers. Let i_{last} be the index of the last station that train k has visited before time t . We consider two cases: (1) train k is dwelling at station i_{last} , that is, $t_{k,i_{\text{last}}}^0 \leq t < t_{k,i_{\text{last}}}^b$; (2) train k is in the segment between station i_{last} and $i_{\text{last}+1}$, that is, $t_{k,i_{\text{last}}}^b \leq t < t_{k,i_{\text{last}+1}}^0$. Based on Assumption 4, the boarding passengers at station i_{last} are not considered in case (1) but are included in case (2). Thus, the number of in-vehicle passengers in train k at time t is

$$\begin{aligned}P_{k,t}^v &= \begin{cases} \sum_{i=1}^{i_{\text{last}}-1} P_{k,i}^b - \sum_{i=1}^{i_{\text{last}}} P_{k,i}^a; & t_{k,i_{\text{last}}}^0 \leq t < t_{k,i_{\text{last}}}^b; \\ \sum_{i=1}^{i_{\text{last}}} (P_{k,i}^b - P_{k,i}^a); & t_{k,i_{\text{last}}}^b \leq t < t_{k,i_{\text{last}+1}}^0; \end{cases} \quad (16) \\ &\forall k \in \mathcal{K}.\end{aligned}$$

Analogously, the total passenger travel time is

$$T_{pt} = \sum_{k \in \mathcal{K}} \sum_{t=0}^{t_{\text{end}}} P_{k,t}^v. \quad (17)$$

Finally, the total cost of passenger time is formulated as

$$C_{\text{total}}^p = \frac{(c_{pw} \cdot T_{pw})}{3600} + \frac{(c_{pt} \cdot T_{pt})}{3600}. \quad (18)$$

The cost unit is yuan (¥).

3.3. *Energy.* In general, for each segment i , with given segment travel time $s_t(i)$ and segment length $s_l(i)$, the speed profile determines energy use [49]. Thus, given a speed profile

generating method, the energy per unit mass of different trains in segment i is the same. We denote the energy per unit mass for one train in segment i as e_i . Its calculation usually considers the following train dynamic equations [48]:

$$\begin{aligned} \frac{dv_k(t)}{dt} &= f_t - f_r, \\ \frac{dx_k(t)}{dt} &= v_k(t); \quad \forall t \in [t_{k,i}^b, t_{k,i+1}^0], i, i+1 \in \mathcal{I}, \\ f_r &= \alpha \cdot v_k^2(t) + \beta \cdot v_k(t) + \gamma, \\ v_k(t_{k,i}^0) &= v_k(t_{k,i}^b) = 0. \end{aligned} \quad (19)$$

The first two equations are train motion equations, where $v_k(t)$ and $x_k(t)$, respectively, represent traveling speed and distance. The third one is the Davis resistance function where Davis parameters are α , β , and γ . The last one requires that every train stops at each station. An optimal train control strategy can be obtained based on (19), in which energy per unit mass is

$$e_i = \int_{t_{k,i}^b}^{t_{k,i+1}^0} f_t \cdot v_k(t) dt; \quad \forall i, i+1 \in \mathcal{I}, k \in \mathcal{K}. \quad (20)$$

However, in (20), the relation between energy and decision variable $s_t(i)$ is unclear. It is also difficult to integrate the optimal train control with other objective functions [18]. Since the speed profile optimization is not considered, for simplicity, we use the method in [50] to generate practical speed profiles. Based on these speed profiles, we then adopt a linear piecewise curve to approximate the energy per unit mass of train k in segment i [48]:

$$e_i = (\chi_x \cdot s_t(i) + \chi_y) \cdot s_l(i); \quad \forall i \in \mathcal{I}. \quad (21)$$

We introduce here an estimation procedure for the values of χ_x and χ_y , which is based on mean squared error (MSE). First, we set several feasible values (e.g., $z = 1, 2, \dots, 100$) of $s_l(z)$ and $s_t(z)$. Then, we obtain optimal speed profiles based on (19) and (20), thereby obtaining the optimal energy with respect to $s_l(z)$ and $s_t(z)$. Finally, we define the optimal energy per unit mass as e_z , and the estimated energy per unit mass obtained with (21) as \tilde{e}_z . The MSE is formulated as

$$\text{MSE}(\chi_x, \chi_y) = \frac{1}{100} \sum_{z=1}^{100} (\tilde{e}_z - e_z)^2. \quad (22)$$

The optimal values of χ_x and χ_y can be obtained when $\text{MSE}(\chi_x, \chi_y)$ reaches its minimum and are found as follows:

$$\begin{aligned} \frac{\partial \text{MSE}(\chi_x, \chi_y)}{\partial \chi_x} &= 0, \\ \frac{\partial \text{MSE}(\chi_x, \chi_y)}{\partial \chi_y} &= 0. \end{aligned} \quad (23)$$

With χ_x^* and χ_y^* obtained, the total energy can be calculated as

$$\begin{aligned} E_{\text{total}} &= \sum_{i=1}^{2I-1} \sum_{k=1}^K [(\chi_x^* \cdot s_t(i) + \chi_y^*) \cdot s_l(i)] \\ &\quad \cdot (M_k + M_p \cdot P_{k,t_{k,i}^b}^v). \end{aligned} \quad (24)$$

Thus, the total cost of energy consumption is

$$\begin{aligned} C_{\text{total}}^e &= c_e \cdot E_{\text{total}} = c_e \\ &\quad \cdot \sum_{i=1}^{2I-1} \sum_{k=1}^K [(\chi_x^* \cdot s_t(i) + \chi_y^*) \cdot s_l(i)] \\ &\quad \cdot (M_k + M_p \cdot P_{k,t_{k,i}^b}^v). \end{aligned} \quad (25)$$

Note that when all tracks are straight and level, χ_x and χ_y suffice for estimating energy. However, if grades and curves are considered, we may have $e_i = (\chi_x^i \cdot s_t(i) + \chi_y^i) \cdot s_l(i)$, $\forall i \in \mathcal{I}$, where χ_x^i and χ_y^i are specified parameters to describe the characteristics in segment i . The parameter estimation procedure is the same.

3.4. Train Cost. In this paper, the train cost is estimated with empirical equations with respect to the interest rate, train price, and average round trip time, where long-term requirements (e.g., the consideration of the highest peak during a year and demand evolution over the planning horizon) are not considered.

Train cost consists of train capital cost and operating cost (excluding energy). Train capital cost is related to the price and economic life of a train, as well as interest rate. Let C_{annual} be the annual cost per train, C_{train} be the current price per train, r be the interest rate, and T_{ls} be the lifespan of a train. The relations among these four factors can be expressed as

$$C_{\text{annual}} = \frac{C_{\text{train}} \cdot r \cdot (1+r)^{T_{\text{ls}}}}{(1+r)^{T_{\text{ls}}} - 1}. \quad (26)$$

The capital cost per train hour c_{tra} can be then calculated by dividing the annual cost C_{annual} by the product of average working hours per day (D_{work}) and working days per year (Y_{work})

$$c_{\text{tra}} = \frac{C_{\text{annual}}}{D_{\text{work}} \cdot Y_{\text{work}}}. \quad (27)$$

In addition, a train can be reassigned after it finishes a round trip and returns to the shunting yard, and the round trip time T_{round}^k can be obtained as

$$T_{\text{round}}^k = t_{k,2I}^b - t_{k,1}^0 + T_{\text{turn}}, \quad \forall k \in \mathcal{K}. \quad (28)$$

Thus, the actual period for train running is $\sum_{k=1}^K T_{\text{round}}^k$ ($\sum_{k=1}^K T_{\text{round}}^k \leq t_{\text{end}}^t$). The estimated fleet size F_s within period $[0, \sum_{k=1}^K T_{\text{round}}^k]$ depends on the round trip time T_{round}^k and the departure headway from the shunting

yard (which is equivalent to the arrival time of train k at the first station $t_{k,1}^0$):

$$F_s = \frac{(\sum_{k=1}^K T_{\text{round}}^k) / K}{[\sum_{k=2}^K (t_{k,1}^0 - t_{k-1,1}^0)] / (K-1)}, \quad (29)$$

where the term $(\sum_{k=1}^K T_{\text{round}}^k) / K$ represents the average round trip time and the term $[\sum_{k=2}^K (t_{k,1}^0 - t_{k-1,1}^0)] / (K-1)$ represents the average headway. The total cost of trains is formulated as

$$C_{\text{total}}^{\text{tra}} = c_{\text{tra}} \cdot F_s \cdot \left(\frac{\sum_{k=1}^K T_{\text{round}}^k}{3600} \right) + c_o \cdot K \cdot \frac{\sum_{i=1}^{2I-1} s_l(i)}{1000}. \quad (30)$$

The first term $c_{\text{tra}} \cdot F_s \cdot (\sum_{k=1}^K T_{\text{round}}^k / 3600)$ represents the capital cost of F_s trains for the actual running period $[0, \sum_{k=1}^K T_{\text{round}}^k]$. The second term $c_o \cdot K \cdot \sum_{i=1}^{2I-1} s_l(i)$ represents the operating cost with frequency K (trains) on a line with the length of the sum of $2I-1$ segments (Km). Both decision variables (segment travel times and arrival times at the first station) affect the train cost.

3.5. Mixed Integer Programming Formulation. The optimization model aims to minimize the weighted-sum of different cost functions, so that the conflicting objectives are generalized in terms of monetary cost. In this paper, the weights of different functions are reflected by their average costs, that is, c_e , c_o , c_{pt} , c_{pw} , and c_{tra} . For example, an operator-oriented timetable (neglecting passenger satisfaction) can be obtained by setting $c_{\text{pt}} = 0$ and $c_{\text{pw}} = 0$; a passenger-oriented timetable can be obtained by setting $c_e = c_o = c_{\text{tra}} = 0$. The weights for integrated objectives can be set similarly, with larger cost coefficients for larger weights. The model is formulated as follows:

Minimize

$$C_{\text{total}} = C_{\text{total}}^p + C_{\text{total}}^e + C_{\text{total}}^{\text{tra}}. \quad (31)$$

Subject to

$$t_{K,1}^0 = t_{\text{end}}^p, \quad (32)$$

$$s_t(I) = T_{\text{turn}}, \quad (33)$$

$$s_t^{\min}(i) \leq s_t(i) \leq s_t^{\max}(i); \quad \forall i \in \mathcal{I}, \quad (34)$$

$$t_{k,i}^b - t_{k-1,i}^b \geq h_{\min}; \quad \forall k = 2, \dots, K, i \in \mathcal{I}, \quad (35)$$

$$P_{k,t}^v \leq N_k; \quad \forall k \in \mathcal{K}, t \in [0, t_{\text{end}}^t]. \quad (36)$$

This model is a nonsmooth, nonconvex programming model, where the nonsmoothness is caused by the limited train capacity, and the nonconvexity is caused by the determination of dwell times [3]. Constraint (32) is equivalent to Assumption 2. Constraint (33) considers the turnaround

from station I to $I+1$ as traveling in segment I , with fixed travel time T_{turn} . Constraint (34) ensures that the segment travel time should be optimized within a reasonable range $[s_t^{\min}(i), s_t^{\max}(i)]$. Constraint (35) imposes the minimum headway constraint for both departures and arrivals (based on Assumption 1). Constraint (36) requires that the number of in-train passengers must not exceed train capacity.

In Caprara et al. [6], the time-space graph is used to derive an integer linear programming model, thereby allowing a considerably faster solution. Analogously, this paper transforms the nonsmooth, nonconvex programming model into a mixed integer programming (MIP) model based on the DTSG. First, we construct two $(2I \times t_{\text{end}}^t)$ matrixes \mathbb{U} and \mathbb{W} , respectively, representing the selection of arrival and departure nodes in the DTSG:

$$\mathbb{U}_{i,t} = \begin{cases} 1; & \text{if } t = \pi(u_k), u_k \in \mathcal{U}^i, k \in \mathcal{K}; \\ 0; & \text{otherwise;} \end{cases} \quad (37)$$

$$\mathbb{W}_{i,t} = \begin{cases} 1; & \text{if } t = \pi(w_k), w_k \in \mathcal{W}^i, k \in \mathcal{K}; \\ 0; & \text{otherwise;} \end{cases}$$

$$\forall i \in \mathcal{I}, t \in [0, t_{\text{end}}^t],$$

where $\mathbb{U}_{i,t} = 1$ ($\mathbb{W}_{i,t} = 1$) if a train arrives at (departs from) station i at time t . As mentioned earlier, the value of $\mathbb{U}_{i,t}$ is determined by $\pi(u_k) = t_{k,1}^0$, for $\forall u_k \in \mathcal{U}^1, k \in \mathcal{K}$. The value of $\mathbb{W}_{i,t}$ is determined by $\mathbb{U}_{i,t}$ and the dwell arcs, that is, $\Delta(u_k, w_k) = t_{k,i}^b - t_{k,i}^0$, for $\forall u_k \in \mathcal{U}^i, w_k \in \mathcal{W}^i, i \in \mathcal{I}, k \in \mathcal{K}$. The value of $\mathbb{U}_{i+1,t}$ is determined by $\mathbb{W}_{i,t}$ and the travel time in segment i , that is, $\Delta(w_k, u_k) = s_t(i)$, for $\forall w_k \in \mathcal{W}^i, u_k \in \mathcal{U}^{i+1}$, and $i \in \mathcal{I} \setminus \{2I\}, k \in \mathcal{K}$. Thus, the decision variables determine \mathbb{U} and \mathbb{W} .

Based on the structures of \mathbb{U} and \mathbb{W} , as well as Assumption 4, we introduce two $(2I \times t_{\text{end}}^t)$ matrixes $\mathbb{P}_{\mathbb{A}}$ and $\mathbb{P}_{\mathbb{B}}$, respectively, representing passenger alighting and boarding information (number and time):

$$\mathbb{P}_{\mathbb{A},i,t} = \begin{cases} P_{k,i}^a; & t = \pi(u_k), u_k \in \mathcal{U}^i, k \in \mathcal{K}; \\ 0; & \text{otherwise;} \end{cases} \quad (38)$$

$$\mathbb{P}_{\mathbb{B},i,t} = \begin{cases} P_{k,i}^b; & t = \pi(w_k), w_k \in \mathcal{W}^i, k \in \mathcal{K}; \\ 0; & \text{otherwise;} \end{cases}$$

$$\forall i \in \mathcal{I}, t \in [0, t_{\text{end}}^t].$$

The cumulative numbers of waiting passengers \mathbb{Q} ($2I \times t_{\text{end}}^t$) and in-vehicle passengers \mathbb{V} ($K \times t_{\text{end}}^t$) can be constructed based on $\mathbb{P}_{\mathbb{A}}$ and $\mathbb{P}_{\mathbb{B}}$:

$$\mathbb{Q}_{i,t} = Q_{i,t} - \|\mathbb{P}_{\mathbb{B},i,t}\|; \quad \forall i \in \mathcal{I}, t \in [0, t_{\text{end}}^t], \quad (39)$$

$$\mathbb{V}_{k,t} = \begin{cases} \mathbb{V}_{k,t-1} + P_{k,i}^b; & \text{if } t = \pi(w_k), w_k \in \mathcal{W}^i, i \in \mathcal{I}; \\ \mathbb{V}_{k,t-1} - P_{k,i}^a; & \text{if } t = \pi(u_k), u_k \in \mathcal{U}^i, i \in \mathcal{I}; \\ \mathbb{V}_{k,t-1}; & \text{otherwise;} \end{cases} \quad (40)$$

$$\forall k \in \mathcal{K}, t \in [1, t_{\text{end}}^t].$$

In (39), $Q_{i,t} = Q_{i,0} + \sum_{q=0}^t \lambda_i(q)$ represents the total waiting passengers at station i , where $Q_{i,0}$ is the number of waiting passengers at the start of this period. According to the definition of cumulative passengers, we have $Q_{i,t} = P_{i,t}^w$. Thus, the total passenger wait time can be expressed by \mathbb{Q} : $T_{\text{pw}} = \sum_{i \in \mathcal{I}} \sum_{t=0}^{t_{\text{end}}^i} P_{i,t}^w = \sum_{i \in \mathcal{I}} \sum_{t=0}^{t_{\text{end}}^i} Q_{i,t} = \|\mathbb{Q}\|$ (since $Q_{i,t} \geq 0$). In (40), the initial number of in-vehicle passengers is zero, that is, $\mathbb{V}_{k,0} = 0$. The number of in-vehicle passengers changes only at arrival/departure times, which is consistent with Assumption 4. Analogously, the total passenger travel time can be expressed as $T_{\text{pt}} = \|\mathbb{V}\|$. Thus, the total cost for passenger time can be rewritten as

$$C_{\text{total}}^p = \frac{1}{3600} (c_{\text{pw}} \cdot \|\mathbb{Q}\| + c_{\text{pt}} \cdot \|\mathbb{V}\|). \quad (41)$$

For energy cost, the energy per unit mass in segment i can be rewritten as

$$e_i = (\chi_x \cdot \Delta(w_k, u_k) + \chi_y) \cdot s_i(i), \quad (42)$$

$$w_k \in \mathcal{W}^i, u_k \in \mathcal{U}^{i+1}; \forall i \in \mathcal{I} \setminus \{2I\}.$$

The total energy can be rewritten as

$$E_{\text{total}} = \sum_{i=1}^{2I-1} \sum_{k=1}^K \sum_{t=0}^{t_{\text{end}}^i} \phi_t \cdot [(\chi_x \cdot \Delta(w_k, u_k) + \chi_y) \cdot s_i(i)] \quad (43)$$

$$\cdot (M_k + M_p \cdot \mathbb{V}_{k,t}), \quad w_k \in \mathcal{W}^i, u_k \in \mathcal{U}^{i+1},$$

where $\phi_t = 1$ if $t = \pi(w_k)$; otherwise $\phi_t = 0$. The total energy cost can be rewritten correspondingly.

For train cost, the round trip time T_{round}^k can be rewritten as

$$T_{\text{round}}^k = \Delta(u_k, w_k) + T_{\text{turn}}, \quad (44)$$

$$u_k \in \mathcal{U}^1, w_k \in \mathcal{W}^{2I}; \forall k \in \mathcal{K}.$$

The departure headway from the shunting yard can be rewritten as

$$t_{k,1}^0 - t_{k-1,1}^0 = \Delta(u_{k-1}, u_k), \quad (45)$$

$$u_{k-1}, u_k \in \mathcal{U}^1; \forall k \in \mathcal{K} \setminus \{1\}.$$

The total cost of trains can be rewritten correspondingly.

Finally, the last departing train constraint (32) is equivalent to

$$\pi(u_K) = t_{\text{end}}^t, \quad u_K \in \mathcal{U}^1. \quad (46)$$

The turnaround constraint (33) is equivalent to

$$\Delta(w_k, u_k) = T_{\text{turn}}, \quad w_k \in \mathcal{W}^I, u_k \in \mathcal{U}^{I+1}; \forall k \in \mathcal{K}. \quad (47)$$

The segment travel time window constraint (34) is equivalent to

$$s_t^{\min}(i) \leq \Delta(w_k, u_k) \leq s_t^{\max}(i), \quad (48)$$

$$w_k \in \mathcal{W}^{i-1}, u_k \in \mathcal{U}^i; \forall k \in \mathcal{K}, i \in \mathcal{I} \setminus \{1\}.$$

The minimum headway constraint (35) is equivalent to

$$\Delta(w_{k-1}, w_k) \geq h_{\min}, \quad (49)$$

$$w_{k-1}, w_k \in \mathcal{W}^i; k \in \mathcal{K} \setminus \{1\}, i \in \mathcal{I}.$$

The train capacity constraint (36) is equivalent to

$$\mathbb{V}_{k,t} \leq N_k, \quad \forall k \in \mathcal{K}, t \in [0, t_{\text{end}}^t]. \quad (50)$$

The selection of arcs determines the arrival nodes u_k and departure nodes w_k , thereby determining every element in the optimization model. Thus, the model is transformed into an MIP. The next section discusses how its solution can be found.

4. Solution Approach

In this section, we decompose the TTP into two subproblems. The first subproblem optimizes the segment travel times (OST), aiming at minimizing the costs for energy and passenger segment travel times. A branch-and-bound algorithm and a frequency determining algorithm are introduced to solve the OST. The second subproblem optimizes departure headways from the shunting yard (OH), aiming to minimize passenger wait time, dwell time and train cost. A novel rolling optimization algorithm is designed to solve the OH.

4.1. Decomposition. The rationale of the decomposition is explained as follows. The objectives in the optimization model include passenger wait time (at their original stations), passenger travel time (segment travel times and station dwell times), energy, and train cost. Based on (43), the total energy is related to the energy per unit mass and the number of passengers in each segment. The energy per unit mass is solely determined by segment travel times $s_i(i)$. Since passengers board at their origins and alight at their destinations regardless of which train they take, the sum of in-vehicle passengers in one segment i is determined by passenger demand. In this regard, the total energy can be expressed as

$$E_{\text{total}} = M_k \cdot K \cdot \sum_{i=1}^{2I-1} e_i + M_p \quad (51)$$

$$\cdot \sum_{i=1}^{2I-1} \sum_{j=i+1}^{2I} \sum_{r=1}^i \sum_{t=0}^{t_{\text{end}}^p} \tau_{r,j}(t) \cdot e_i.$$

Therefore, the total energy cost depends only on segment travel times $s_i(i)$. The passenger segment travel times also depend only on segment travel times.

On the other hand, passenger wait time, dwell time, and train cost depend on the train arrival and departure times at stations. Based on Assumption 1, train arrival/departure times at stations are affected by the departure headways from the shunting yard. Thus, it is reasonable to use segment travel times as input and consider the costs for passenger wait time, dwell time, and trains as functions of the departure headways.

In general, the OST subproblem is formulated as follows:
Minimize

$$C_{\text{total}}^{\text{OST}} = \sum_{i=1}^{2I-1} \left\{ c_e \cdot M_k \cdot K \cdot e_i + (c_e \cdot M_p \cdot e_i + c_{\text{pt}} \cdot s_i(i)) \cdot \left[\sum_{j=i+1}^{2I} \sum_{r=1}^i \sum_{t=0}^{t_{\text{end}}^p} \tau_{r,j}(t) \right] \right\}. \quad (52)$$

Subject to constraints (47) and (48).

It should be noted that since segment travel times are given inputs, minimizing the cost of passenger dwell time is equivalent to minimizing the cost of total travel time. Since we already obtain the in-train passenger matrix \mathbb{V} , there is no need to further derive dwell times at stations. Thus, the objectives in the OH subproblem include the total passenger time cost (both \mathbb{Q} and \mathbb{V}) instead of costs of passenger wait and dwell times. The OH subproblem is formulated as follows (with segment travel times as input):

Minimize

$$C_{\text{total}}^{\text{OH}} = C_{\text{total}}^p + C_{\text{total}}^{\text{tra}}. \quad (53)$$

Subject to constraints (46), (49) and (50).

4.2. Solution Approach for OST. The solution approach for OST consists of branch-and-bound (B&B) and frequency determining algorithms. On one hand, the B&B algorithm needs frequency as input. On the other hand, the frequency determining algorithm needs segment travel times obtained by the B&B to evaluate the performance. Thus, we iteratively evaluate the performance with different frequencies (each frequency is used as input for B&B to find a solution of segment travel times) and choose the one with the best performance. The segment travel times are then determined correspondingly.

4.2.1. Branch-and-Bound (B&B) Algorithm. The OST subproblem is an integer linear programming problem with $2I - 1$ decision variables ($s_i(i)$, for $i = 1, 2, \dots, 2I - 1$). Here we use a generic B&B, which is widely used to solve linear programming problems [7]. Generally, a B&B needs an input model formulated as follows:

$$\min \quad \overrightarrow{\mathbf{c}} \cdot \overrightarrow{\mathbf{x}}^T, \quad (54)$$

$$\text{s.t.} \quad \mathbb{A} \cdot \overrightarrow{\mathbf{x}}^T \leq \overrightarrow{\mathbf{b}} \quad (55)$$

$$\mathbb{A} \text{eq} \cdot \overrightarrow{\mathbf{x}}^T = \overrightarrow{\mathbf{b}} \text{eq},$$

where $\overrightarrow{\mathbf{c}}$ is the coefficient vector and $\overrightarrow{\mathbf{x}}^T$ is the decision variable vector. Both $\overrightarrow{\mathbf{c}}$ and $\overrightarrow{\mathbf{x}}^T$ have $2I - 1$ elements. \mathbb{A} and $\mathbb{A} \text{eq}$ are coefficient matrixes, and $\overrightarrow{\mathbf{b}}$ and $\overrightarrow{\mathbf{b}} \text{eq}$ are constant vectors representing the bounds. Here \mathbb{A} is a $(4I - 2) \times (2I - 1)$ matrix, and $\mathbb{A} \text{eq}$ is a $1 \times (2I - 1)$ matrix. Correspondingly, $\overrightarrow{\mathbf{b}}$ has $2I - 1$ elements, and $\overrightarrow{\mathbf{b}} \text{eq}$ has one element. We define a vector with $2I - 1$ elements as

$$\overrightarrow{\mathbf{e}}_j = (0, 0, \dots, 1, \dots, 0, 0), \quad j = 1, \dots, 2I - 1. \quad (56)$$

The decision variable vector can be represented as $\overrightarrow{\mathbf{x}}_i = s_i(i)$ for $i = 1, 2, \dots, 2I - 1$, where $\overrightarrow{\mathbf{x}}_i$ is the i -th element in the vector $\overrightarrow{\mathbf{x}}$. We then define $\Upsilon_i = \sum_{j=i+1}^{2I} \sum_{r=1}^i \sum_{t=0}^{t_{\text{end}}^p} \tau_{r,j}(t)$ for $i = 1, 2, \dots, 2I - 1$ as the total number of passengers that travel through segment i . Based on (52), $\overrightarrow{\mathbf{c}}_i$ can be represented as

$$\begin{aligned} \overrightarrow{\mathbf{c}}_i &= c_e \cdot M_k \cdot K \cdot \chi_x \cdot s_i(i) + \Upsilon_i \\ &\cdot (c_e \cdot M_p \cdot \chi_x \cdot s_i(i) + c_{\text{pt}}); \end{aligned} \quad (57)$$

$$i = 1, 2, \dots, 2I - 1.$$

Constraints (55) are represented as follows. Based on constraint (47), we have $\mathbb{A} \text{eq} = \overrightarrow{\mathbf{e}}_1$ and $\overrightarrow{\mathbf{b}} \text{eq} = T_{\text{turn}}$. Based on constraint (48), we have $\mathbb{A}_{i,:} = -\overrightarrow{\mathbf{e}}_i$, for $i = 1, 2, \dots, 2I - 1$; $\mathbb{A}_{i,:} = \overrightarrow{\mathbf{e}}_i$, for $i = 2I, 2I + 1, \dots, 4I - 2$, where $\mathbb{A}_{i,:}$ represents the i -th row of matrix \mathbb{A} . Analogously, $\overrightarrow{\mathbf{b}}_i = -s_i^{\min}(i)$, for $i = 1, 2, \dots, 2I - 1$; $\overrightarrow{\mathbf{b}}_i = s_i^{\max}(i)$, for $i = 2I, 2I + 1, \dots, 4I - 2$. Since we use the standard B&B algorithm [40], the detailed algorithm description is not shown here.

4.2.2. Frequency Determining Algorithm. Here we consider cyclic timetables (train departure headways from the shunting yard are constant, e.g., 200 s) to discuss the frequency, because they have following properties: (1) The cyclic timetable with minimum cost (i.e., the best cyclic timetable) can be considered as an upper bound of the noncyclic timetable. (2) If passenger demand is stable, the best cyclic timetable is a good approximation to an optimal noncyclic one. (3) The cyclic timetables are easy to obtain.

First, the range of frequency should be estimated. We define the peak passenger volume in one segment as q_{peak} , which is the maximum value among the numbers of passengers traveling through the segment, that is, $q_{\text{peak}} = \max(\Upsilon_i \mid i = 1, 2, \dots, 2I - 1)$. The minimum frequency K_{\min} is determined by the train capacity N_k and q_{peak} : $K_{\min} = \lceil q_{\text{peak}}/N_k \rceil$. The maximum frequency K_{\max} is determined by the minimum headway: $K_{\max} = \lfloor t_{\text{end}}^p/h_{\min} \rfloor + 1$.

Second, for each frequency $K \in [K_{\min}, K_{\max}]$, we use K as input of the B&B, to obtain the segment travel times. Since the frequency for a cyclic timetable determines the departure headways from the shunting yard ($h = t_{\text{end}}^p/(K - 1)$, where h is the departure headway), the cyclic timetable that corresponds to frequency K can be obtained. Then, the total cost of the timetable can be calculated with (52).

Finally, we obtain different cyclic timetables with different costs, each corresponding to a frequency $K \in [K_{\min}, K_{\max}]$. We select the timetable with minimum cost as the best cyclic timetable, and its corresponding segment travel times as the OST solution within this period. The detailed algorithm is as shown in Algorithm 1.

4.3. Solution Approach for OH. Here we design a rolling optimization (RO) algorithm that optimizes the arrival time at the first station for one train at a time. The main difference between the RO and the rolling horizon (RH) approach is in the RH, the bound conditions (i.e., the start and end times of each horizon) require some of the original variables to have fixed, known values [3], while in the RO, each decision variable may still be optimized. Thus, the rolling optimization is more flexible.

4.3.1. The Rolling Optimization (RO) Algorithm. Based on (53), the objectives include the total costs of passenger time and trains. To obtain these costs, we need the segment travel times as input. Here, instead of considering the whole period, we consider a much shorter period $t_p \leq h_{\min} + \sum_{i=1}^{2I-1} s_i(i) + \sum_{i=1}^{2I} \bar{a}_i$, e.g., 600 s, for solving the OST. For example, for each train $k \geq 2$, its previous train arrives at the first station at time $t_{k-1,1}^0$ (the first train arrives at time 0), and we consider the period $[t_{k-1,1}^0, t_{k-1,1}^0 + t_p]$ to solve the OST (the constraint of t_p ensures that $t_{k-1,1}^0 + t_p \leq t_{\text{end}}^t$). In this way, the key ideas of “iterative optimization” (the frequencies, segment travel times, and headways are interdependent) and “rolling horizon” (the smaller period studied in each step) are adopted, and future passenger demand is considered.

Then, we discuss the costs for one train $k \geq 2$ if its arrival time at the first station is given. Since each train is scheduled in chronological order, the waiting passenger matrix \mathbb{Q} and the in-train passenger matrix \mathbb{V} can be constructed chronologically. For each train $k \geq 2$, the departure time of its preceding train at the first station is $t_{k-1,1}^b$, and train k departs from the first station at time $t_{k,1}^b$. We define the passenger wait time with respect to train k as $T_{\text{pw}}(k)$, which is formulated as

$$T_{\text{pw}}(k) = \sum_{i=1}^{2I} \sum_{t=t_{k-1,1}^b}^{t_{k,1}^b} \mathbb{Q}_{i,t}; \quad \forall k \in \mathcal{K}. \quad (58)$$

Analogously, we define the passenger travel time for train k as $T_{\text{pt}}(k)$, and we obtain

$$T_{\text{pt}}(k) = \sum_{t=t_{k,1}^b}^{t_{k,2I}^b} \mathbb{V}_{k,t}; \quad \forall k \in \mathcal{K}. \quad (59)$$

The estimated fleet size of train k (F_s^k) depends on its round trip time (T_{round}^k) and the departure headway from the shunting yard ($t_{k,1}^0 - t_{k-1,1}^0$):

$$F_s^k = \frac{T_{\text{round}}^k}{t_{k,1}^0 - t_{k-1,1}^0}; \quad \forall k \in \mathcal{K}. \quad (60)$$

Based on (18), (30), (53), (58), (59), and (60), we define the total costs in the OH for train k as $C_{\text{total}}^{\text{OH}}(k)$, which can be expressed as

$$C_{\text{total}}^{\text{OH}}(k) = \frac{c_{\text{pw}} \cdot T_{\text{pw}}(k) + c_{\text{pt}} \cdot T_{\text{pt}}(k) + c_{\text{tra}} \cdot F_s^k \cdot T_{\text{round}}^k}{3600} + c_o \quad (61)$$

$$\cdot \frac{\sum_{i=1}^{2I-1} s_l(i)}{1000}; \quad \forall k \in \mathcal{K}.$$

The value of $C_{\text{total}}^{\text{OH}}(k)$ is a function of the arrival time for train k at the first station. Since the arrival time for the previous train is given, we consider the value of $C_{\text{total}}^{\text{OH}}(k)$ as a function of the headway h between train $k-1$ and train k ($h = t_{k,1}^0 - t_{k-1,1}^0$), for consistency with other TTP studies.

Next we discuss how the optimal value for h can be obtained. Based on (58), (59), and (60), both the passenger time and train cost depend on passenger demand. Since the passenger demand is time-varying, the relation between $C_{\text{total}}^{\text{OH}}(k)$ and h cannot be described with a fixed function.

The expected objective function of OH is shown in Figure 4. Its value first decreases (the headway starts from h_{\min}) to a minimum value and then increases monotonically. We specify that this function reaches its minimum value at headway h^* . Since the DTSG divided the continuous time into seconds, we iteratively calculate the objective function of OH with respect to h (starting from $h = h_{\min}$, and $h = h + 1$ in each following step), until the optimal headway h^* is found, or the total cost increases as h increases (in case that h^* cannot be reached). It should be noted that sometimes a headway is not feasible due to the model constraints (e.g., a short headway might violate the minimum headway constraint). In such cases, we define the value of $C_{\text{total}}^{\text{OH}}(k)$ to be $+\infty$ if it corresponds to an infeasible headway.

Finally, for each train $k \geq 2$, we can obtain its optimal headway h^* (between train $k-1$ and train k). By rolling optimization of departure headways for trains, the timetable can be obtained. Let $\bar{\mathbf{H}}$ be the vector of optimal headways for all trains $k \geq 2$. The detailed rolling optimization (RO) algorithm is described in Algorithm 2. The arrival and departure times of the first train should be obtained beforehand, and $t_{1,1}^0 = 0$.

4.3.2. Other Approaches. As pointed out by Wang et al. [3], other approaches such as a pattern search [51] or a genetic algorithm [36] can be applied to solve the TTP. Here we briefly describe the generic structures of these two approaches.

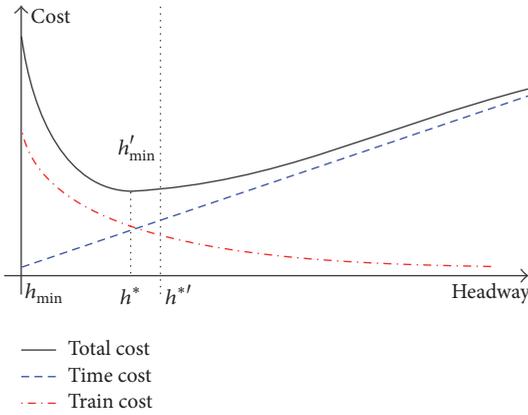
A genetic algorithm (GA) is an iterative heuristic that seeks an optimal individual among a population of solutions in each generation, and uses selection, crossover, mutation and possibly other operations to obtain new generations. Such process is repeated until the termination condition is satisfied. Since the size of the chromosome should be preset, the frequency is considered here, a given and unchangeable input. Then, the segment travel times can be obtained with

- Step 1:** Estimate the frequency range, i.e., $[K_{\min}, K_{\max}]$. Set the initial frequency of the cyclic timetable as $K = K_{\min}$, the best frequency found during the process as $K_{\text{best}} = 0$, and the minimum total cost found during the process as $TC_{\min} = +\infty$. Go to Step 2.
- Step 2:** For frequency K , obtain the segment travel times through the B&B algorithm, and obtain the uniform departure headways from the shunting yard $\mathbf{h} = \mathbf{t}_{\text{end}}^p / (K - 1)$. Go to Step 3.
- Step 3:** Calculate total cost TC_K corresponding to the frequency K . If $TC_K < TC_{\min}$, then set $TC_{\min} = TC_K$, and $K_{\text{best}} = K$. Go to Step 4.
- Step 4:** If $K = K_{\max}$, then go to Step 5. Otherwise, set $K = K + 1$, and return to Step 2.
- Step 5:** Output K_{best} and TC_{\min} . The segment travel times within this period are the solution of the B&B algorithm that corresponds to frequency K_{best} .

ALGORITHM 1: Frequency determining algorithm.

- Step 1:** Set $\vec{H} = (0, 0, \dots, 0, 0)$. Set the period t_p for OST solution approach. Set $k = 2$, $\mathbf{h} = \mathbf{h}_{\min}$, and go to Step 2.
- Step 2:** Obtain the segment travel times for train k by OST solution approach. Obtain the arrival and departure times at stations for train k . Go to Step 3.
- Step 3:** Calculate the value of $C_{\text{total}}^{\text{OH}}(\mathbf{k})$ with respect to \mathbf{h} , and calculate the value of $C_{\text{total}}^{\text{OH}}(\mathbf{k})$ with respect to $\mathbf{h}' = \mathbf{h} + 1$. Go to Step 4.
- Step 4:** If $C_{\text{total}}^{\text{OH}}(\mathbf{k}) > C_{\text{total}}^{\text{OH}}(\mathbf{k})$, then go to Step 5; Otherwise, set $\mathbf{h} = \mathbf{h}'$, and return to Step 2.
- Step 5:** Set the optimal headway for train k as $\mathbf{h}^* = \mathbf{h}$, and set $\vec{H}(\mathbf{k} - 1) = \mathbf{h}^*$. Go to Step 6.
- Step 6:** If $\|\vec{H}\| \geq t_{\text{end}}^p$, then the rolling optimization is finished, and thus the timetable for the period $[0, t_{\text{end}}^p]$ can be obtained. Otherwise, set $k = k + 1$, $\mathbf{h} = \mathbf{h}_{\min}$, and return to Step 2.

ALGORITHM 2: Rolling optimization algorithm.

FIGURE 4: Total cost for train k with respect to headway.

the given frequency. An individual consists of headways for all trains ($K - 1$ headways), and its objective function is calculated based on (31). The initial individuals are generated as the cyclic headways. Random values are then added or subtracted to the headways, considering constraints (46) and (49). The details of standard GA [3, 36] are not shown here.

Compared with the proposed solution approach (RO) in this paper, generic GA (GGA) has following properties: (1) In terms of input, both GGA and RO require passenger demand as input. However, GGA requires fixed frequency as input, while RO determines optimal frequency during the process. (2) In terms of optimization process, GGA optimizes

the solution (of all trains) in a random direction, while RO optimizes one train at a time. (3) In terms of computation complexity, GGA calculates the objective function of the whole system for each chromosome, while RO calculates the objective of one train at each step. These differences result in different computation performances, which will be further demonstrated in the numerical experiments.

For the numerical experiments, we set the parameter values in the generic GA as follows. The population size is 20, the maximum iteration number is 1000, the crossover rate is 0.7, and the mutation rate is 0.1. The algorithm terminates if the maximum number of iterations is reached or the best fitness value in this generation is the same as in the previous 100 generations ahead.

The pattern search (PS) algorithm is proposed to solve unconstrained optimization problems [52], which varies one variable (headway) at a time by steps of the same magnitude. When no such increase or decrease in any one variable further decreases the objective function, it decreases the step size and repeats the process until the steps are deemed sufficiently small. Here the unconstrained objective is constructed with the Augmented Lagrangian (AL) method, which adds the constraints as penalty terms to the objective function. The penalty term is the product of a Lagrangian multiplier and a constraint equation (e.g., from $g(x) \geq 0$ to $g(x) - y^2 = 0$, where y is a new variable). Let the number of constraints be m . Set the multiplier that corresponds to the j -th constraint as ω_j , and set a large enough parameter σ . By minimizing the objective function with respect to y (so that the intermediate

TABLE 3: Parameter values in the experiments.

Parameter	Value
M_k (Kg)	1.99×10^5
M_p (Kg)	75
N_k (Passenger)	1290
c_e (¥/Kwh)	0.79
c_o (¥/(train·Km))	20
c_{pt} (¥/(passenger-hour))	10
c_{pw} (¥/(passenger-hour))	20
c_{tra} (¥/(train-hour))	800
h_{min} (s)	100
f_r (m/s)	$1.36 \times 10^{-4} \cdot v_k^2(t) + 1.45 \times 10^{-2} \cdot v_k(t) + 0.08$
χ_x	-1.944×10^{-4}
χ_y	0.4176

variable y does not influence the objective function), we can obtain the typical form of the penalty term as

$$\frac{1}{2\sigma} \sum_{j=1}^m \left\{ \left[\max(0, \omega_j - \sigma \cdot g_j(x)) \right]^2 - \omega_j^2 \right\}. \quad (62)$$

The parameter value ω_j is optimized in each iterative step, until the termination condition is satisfied. In each step of AL, we use the PS to find a minimum value of the unconstrained objective function. In this way, we integrate AL and PS and define such approach as AL + PS. The same drawback exists that the frequency is preset and considered as input. The initial solution is the cyclic headways. For the numerical experiments, we set the values of parameters in the AL + PS as follows. The initial step size for PS is 32 s. The speeding factor for PS is 1. The shrinking factor for PS is 0.5. The allowance error for both PS and AL is 1 s (since the continuous time is discretized into seconds). The augmented factor for AL is 1.2. The large enough parameter σ in AL is 200.

5. Numerical Experiments

In this section, we present several numerical experiments to test the effectiveness and efficiency of the proposed solution approach. All experiments are performed with Matlab on a PC with 2 GHz Intel Core7 and 8 GB memory. Parameters in the examples are shown in Table 3. We adopt here the same energy calculation method as in [48]. To save space, we do not show the detailed calculations and directly provide the estimated values of parameters χ_x and χ_y in Table 3.

5.1. A Small Case Study. Here we consider a short metro line, as shown in Figure 5. The segment lengths are 1800 m, 1600 m, and 2000 m. For simplicity, the minimum and maximum dwell times for each station are all set as 30 s and 90 s, respectively.

The baseline passenger demand is shown in Table 4. The “Boarding (Alighting) Sum” represents the number of

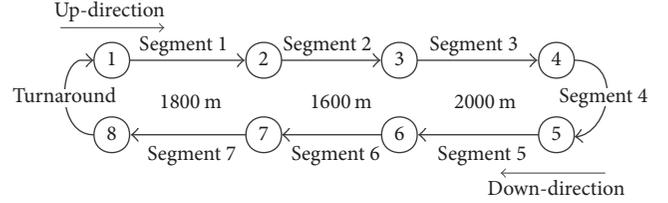


FIGURE 5: Representation of a short metro line.

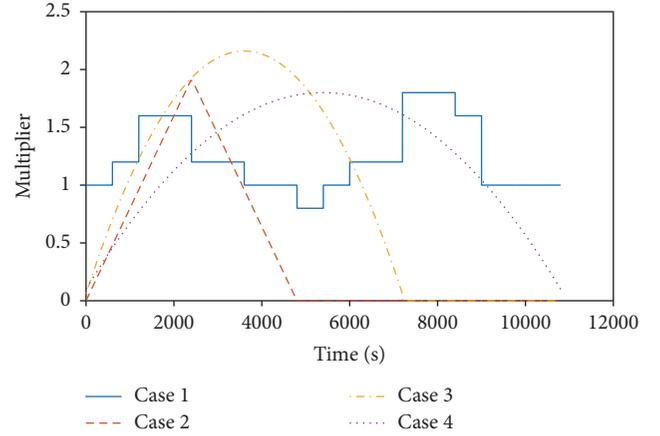


FIGURE 6: Passenger demand distributions in different cases.

total boarding (alighting) passengers per second at a station. Four cases are considered here. As shown in Figure 6, the curves represent demand variance over time, and “multiplier” represents the ratio between actual and baseline passenger demand. For example, in case 1, during period [1200, 2400] (s), the multiplier is 1.6, and passenger demand from station 2 to station 3 at $t \in [1200, 2400]$ is 0.96 (0.6×1.6) passengers/second. The four cases have different passenger demand distributions.

The computational results of different cases are shown in Table 5 and Figure 7. Specifically, in Table 5, the noncyclic result obtained by RO is denoted as $N_{K=K_N}$, which schedules K_N trains within a given time duration. The cyclic result obtained by Algorithm 1 is denoted as $U_{K=K_U}$, which schedules K_U trains within a given time duration. Particularly, the best cyclic result is denoted as $U_{K=K_U}^{best}$. The GGA result is denoted as $GA_{K=K_N}$, which uses frequency K_N from RO result as input. The objectives of different results are calculated, including passenger wait time (T_{pw}), passenger travel time (T_{pt}), total energy (E_{total}), fleet size (F_s), and total cost (C_{total}). Figure 7 specifically shows the numerical results in passenger satisfaction.

Table 5 and Figure 7 demonstrate that, under time-varying passenger demand, noncyclic timetables obtained by RO outperform the other timetables, while GGA results outperform cyclic ones obtained by Algorithm 1. This result verifies that the noncyclic timetable improves passenger satisfaction. Particularly, the passenger wait time (T_{pw}), passenger travel time (T_{pt}), and total cost (C_{total}) have their lowest values in the RO results, which demonstrates that the RO

TABLE 4: Baseline passenger demands for a short metro line (passengers per second).

Station index	1	2	3	4	5	6	7	8	Boarding sum
1	0	0.3	0.3	0.45	0	0	0	0	1.05
2	0	0	0.6	0.3	0	0	0	0	0.9
3	0	0	0	0.3	0	0	0	0	0.3
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0.45	0.75	0.3	1.5
6	0	0	0	0	0	0	0.6	0.3	0.9
7	0	0	0	0	0	0	0	0.45	0.45
8	0	0	0	0	0	0	0	0	0
Alighting sum	0	0.3	0.9	1.05	0	0.45	1.35	1.05	-

TABLE 5: Comparison between cyclic and non-cyclic timetables.

Time duration (s)	Timetable type		T_{pw} (10^6)	T_{pt} (10^6)	E_{total} (10^3)	F_s	C_{total} (10^4)	Gap (%)
0-10800	Non-cyclic	$N_{K=60}$	7.2330	14.384	18.069	7	12.320	-
	Non-cyclic	$GA_{K=60}$	7.2786	14.402	18.069	7	12.329	0.07
	Cyclic	$U_{K=53}^{best}$	8.3360	14.523	16.148	6	12.496	1.41
	Cyclic	$U_{K=60}$	7.3561	14.406	18.069	7	$+\infty$	-
0-4800	Non-cyclic	$N_{K=24}$	2.3171	4.9470	7.1411	6	4.3586	-
	Non-cyclic	$GA_{K=24}$	2.5489	4.9758	7.1411	6	4.4954	3.04
	Cyclic	$U_{K=24}^{best}$	2.6102	4.9988	7.1411	6	4.5407	4.01
	Cyclic	$U_{K=23}$	2.7619	5.0347	6.8667	6	4.5681	4.59
0-7200	Non-cyclic	$N_{K=45}$	4.3948	11.292	13.615	8	8.8022	-
	Non-cyclic	$GA_{K=45}$	4.5726	11.328	13.615	8	8.9245	1.37
	Cyclic	$U_{K=48}^{best}$	4.3975	11.319	14.438	8	9.0243	2.46
	Cyclic	$U_{K=45}$	4.7268	11.417	13.615	8	9.0345	2.57
0-10800	Non-cyclic	$N_{K=60}$	6.0280	14.301	18.069	7	11.621	-
	Non-cyclic	$GA_{K=60}$	6.3113	14.560	18.069	7	11.854	1.97
	Cyclic	$U_{K=60}^{best}$	6.4638	14.330	18.069	7	11.875	2.14
	Cyclic	$U_{K=59}$	6.5981	14.381	17.795	7	11.899	2.34
	Cyclic	$U_{K=61}$	6.3572	14.327	18.344	7	11.884	2.21

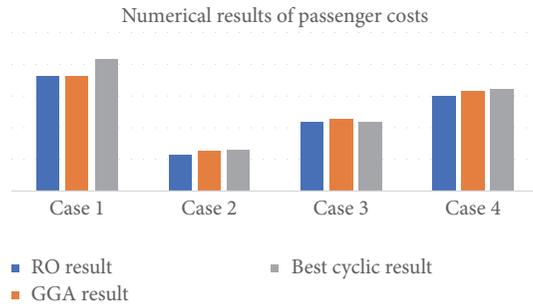


FIGURE 7: Passenger costs for different results in different cases.

method is effective in solving the proposed train timetabling problem. In case 1, the result of $U_{K=60}$ is infeasible because it violates the minimum headway constraint at intermediate stations (the dwell times are different from $U_{K=60}$ to $N_{K=60}$), and thus we have $C_{total} = +\infty$. It is interesting that $N_{K=60}$ and $U_{K=60}$ require the same energy and fleet size. This result

verifies the reasonableness of model decomposition in this paper from two perspectives: (1) When all passengers reach their destinations, energy is solely determined by segment travel times (they are the same in $N_{K=60}$ and $U_{K=60}$). (2) With given segment travel times, the estimated fleet size is solely determined by headways. Similar results can be observed in other cases.

The gaps in Table 5 shows that case 2 has the largest gap, case 3 and 4 have similar gaps, and case 1 has the smallest gap. Note that, in case 2, passenger demand changes drastically; in cases 3 and 4, demand patterns are quite similar; in case 1, within certain subperiods, passenger demand is the same, which is the steadiest among all cases. Therefore, we can conclude that as the passenger demand varies more significantly over time, the gap between the cyclic and noncyclic timetables increases.

Figure 8 illustrates the obtained train timetables in different cases, where the departure headways vary in accordance with the fluctuating passenger demand. For example, in cases 2, 3, and 4, the headways are small in the peak period and

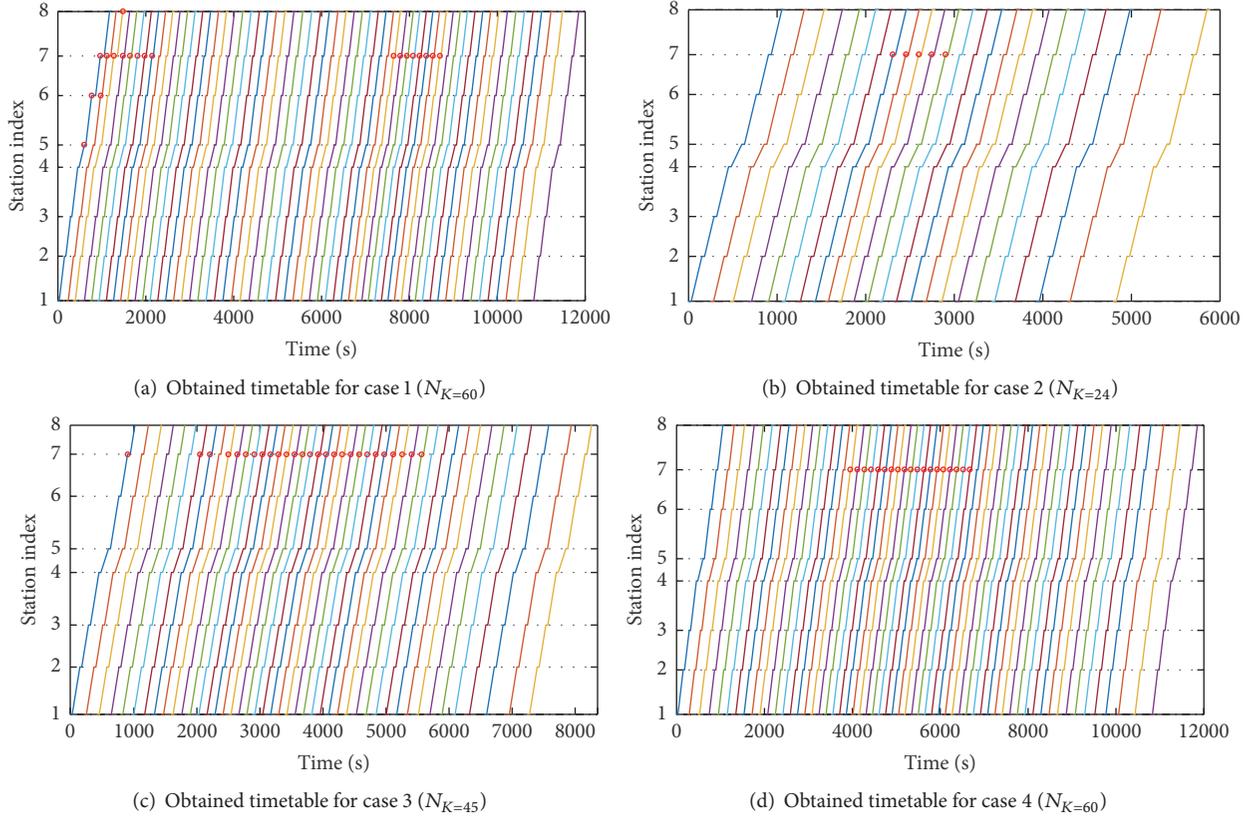


FIGURE 8: Obtained timetables for different cases.

are larger before and after the peak period. In case 1, the small headways occur during subperiods [1200, 2400] (s) and [7200, 8400] (s), where the multipliers of passenger demand are largest (1.6 and 1.8). In Figure 8, the circles represent the dwell times that exceed the minimum dwell time (i.e., 30 s). We observe that the excess dwell times (circles) mostly occur at station 7. This is consistent with the OD pairs we present in Table 4. Specifically, the difference between the cumulative boarding sum and the cumulative alighting sum reaches its maximum value at station 7 ($1.5 + 0.9 - 0.45 = 1.95$). In this regard, the alighting and boarding rates are smallest at station 7. In addition, the total number of alighting and boarding passengers is largest (i.e., 1.8 passengers per second) at station 7, and thus the excess dwell times here are reasonable.

Moreover, despite the symmetry of passenger demand distributions, the headways at the ends of periods are larger than at their starts. This can be explained by that passenger demand increases at the start and decreases to zero at the end, which leads to different cost changes. Specifically, the passenger time cost increases more significantly when demand increases, while the train cost does not change with demand. Thus, the headway tends to be smaller at the start of periods than at the ends. In fact, it is often seen in actual operations that the headways between the last trains at the end of day exceed those between the first trains in the morning.

Finally, Table 6 shows the computational performances for different methods, including RO method, GGA, and

TABLE 6: Computational performances for different methods.

	RO method	Generic GA	AL + PS
(Case 1) Time (s)	225.5	8299.4	10800
Total cost (¥)	12.32×10^4	12.33×10^4	12.42×10^4
(Case 2) Time (s)	86.1	4658.2	10800
Total cost (¥)	4.3586×10^4	4.4954×10^4	4.4872×10^4
(Case 3) Time (s)	120.6	5009.3	10800
Total cost (¥)	8.8022×10^4	8.9245×10^4	8.9567×10^4
(Case 4) Time (s)	240.717	8229.1	10800
Total cost (¥)	11.621×10^4	11.854×10^4	11.774×10^4

AL + PS method. As described above, GGA and AL + PS require a preset fixed frequency, and thus we use the same frequency from RO result K_N as input. We terminate the algorithm when its computation time reaches 3 hours and output the best solution obtained by then.

Table 6 demonstrates that RO is very efficient, obtaining a timetable very fast. In comparison, GGA takes more than 1.5 hours to obtain a timetable, and AL + PS takes more than 3 hours in all cases. In each case, the timetable obtained with the RO method has the lowest total cost, which further verifies its effectiveness. The GGA timetables perform better than AL + PS in cases 1 and 3 but are outperformed in cases 2 and 4.

TABLE 7: Real-world infrastructure data of Beijing Metro Line 4.

Station	Minimum dwell time (s)	Maximum dwell time (s)	Segment	Minimum travel time (s)	Maximum travel time (s)	Segment length (m)
1/48 AHQB	30	90	1	124	170	1363
2/47 BGM	30	90	2	114	156	1251
3/46 XY	30	90	3	152	209	1672
4/45 YMY	30	90	4	118	162	1295
5/44 BDDM	30	90	5	81	111	887
6/43 ZGC	30	90	6	82	113	900
7/42 HDHZ	30	90	7	97	133	1063
8/41 RMDX	30	90	8	96	131	1051
9/40 WGC	30	90	9	151	207	1658
10/39 NL	30	90	10	138	190	1517
11/38 BJZ	30	90	11	131	180	1441
12/37 XZM	30	90	12	93	128	1025
13/36 XJK	30	90	13	100	138	1100
14/35 PAL	30	90	14	100	138	1100
15/34 XS	30	90	15	79	109	869
16/33 LJHT	30	90	16	92	126	1011
17/32 XD	30	90	17	74	102	815
18/31 XWM	30	90	18	105	144	1152
19/30 CSK	30	90	19	109	150	1200
20/29 TRT	30	90	20	149	205	1643
21/28 BJSRS	30	90	21	135	185	1480
22/27 MJP	30	90	22	75	103	827
23/26 JMX	30	90	23	90	124	989
24/25 GYXQ	30	90	-	-	-	-

Combining all the computational results in this case study, we can conclude some disadvantages of GGA and AL + PS compared to RO. (1) Although the timetables obtained by them are noncyclic, train departure headways in their results differ greatly from one another without clear correspondence to the passenger demand. In comparison, RO timetables are more regular (headways change gradually) and capture well the passenger demand patterns. (2) Due to the computation complexity, the computation time for running GGA and AL + PS once is far greater than RO. (3) They both require a preset frequency as input. If we do not know the best frequency beforehand, which is true in most real-world scenarios, we must run these methods with different frequencies and select the best solution, thereby further increasing substantially computation time.

In general, the RO method is superior to GGA and AL + PS in obtaining a noncyclic timetable with time-varying passenger demand.

5.2. Case Study of Beijing Metro Line 4. This section aims to test the applicability and effectiveness of RO in a real-world case study based on Beijing's Metro Line 4 (BML4). The infrastructure information and some operational parameters of BML4 are shown in Table 7. The turnaround time is 120 s. To satisfy the large passenger demand, we set train capacity

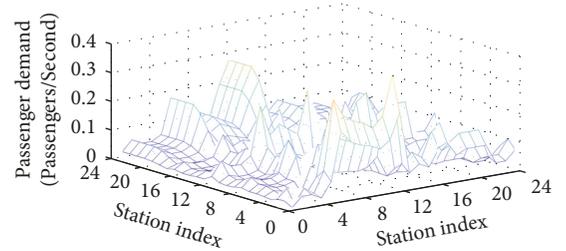


FIGURE 9: Baseline passenger demand for BML4.

as $N_k = 2160$ passengers, which is a maximum overloaded capacity. If not mentioned, the other parameters are the same as in Table 3. The baseline passenger demand data are based on the research conducted by Duan et al. [53], which are presented in Figure 9, where $z(x, y), \forall x < y$ represents morning peak demand and $z(x, y), \forall x > y$ represents evening peak demand.

Here we consider three types of scenarios whose passenger demands are large and unsteady, which are difficult to manage: (a) period of 3 hours (i.e., 10800 s) including morning peak; (b) period of 3 hours including evening peak; and (c) a shopping festival hosted in the shopping centers next to the XD station (whose index is 8/41). Representations of

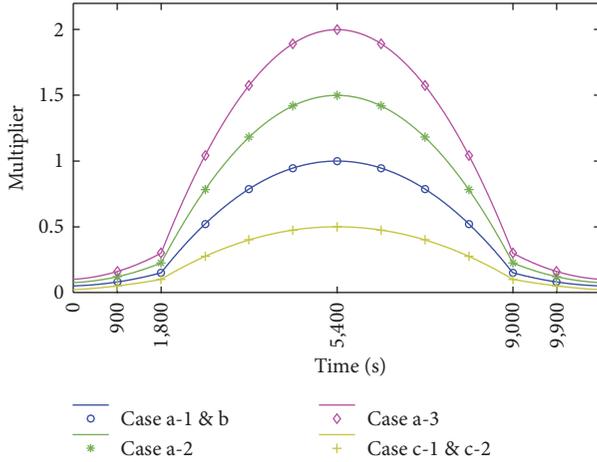


FIGURE 10: Passenger demand patterns in different cases.

passenger demands are the same as in the small case study (as shown in Figure 10): the actual passenger demand is the product of baseline passenger demand (Figure 9) and the multipliers (Figure 10). To test RO with more instances, we design three subcases of scenario (a) with different peak multipliers, and two subcases of scenario (c): without the special event (c-1) or with it (c-2). In case (c-2), the passenger demand at station XD is tripled.

It should be noted that the real operation data of BML4 is dynamic and confidential. Besides, the operator today prefers to schedule trains uniformly. Therefore, in this experiment, we use cyclic timetables with best performance to approximate the real-world timetables. The notations are the same as in the small case study.

Table 8 and Figure 11 show the computational results in different cases. Note that, for a major metro line such as BML4, the computation times of GGA and AL + PS are extremely large, and in many cases they do not yield a satisfactory solution within a reasonable time. Thus, we do not present their results in this experiment. The results of $U_{K=46}$ in case (a-1), $U_{K=54}$ in case (a-2), and $U_{K=54}$ in case (a-3) are infeasible, and thus their total costs are $+\infty$. Figure 11 shows that considerable costs are reduced by the noncyclic timetables obtained by RO. Interestingly, in the cases where cyclic timetables are feasible, that is, cases (a-1), (b), (c-1), and (c-2), the frequency of the noncyclic timetable is always larger than the best cyclic ones. This is because the passenger demand in the real-world fluctuates so drastically that cyclic timetables cannot capture its dynamic patterns, and the minimum headway constraint makes it hard to obtain feasible cyclic timetables. In this regard, adding more trains in a cyclic timetable may be either uneconomic (i.e., the increased train operating cost exceeds the reduced passenger time value) or infeasible.

In comparison, the noncyclic timetables obtained by RO respond well to demand changes. For example, in case (a-1), the result of $N_{K=46}$ substantially reduces passenger wait time compared to $U_{K=44}^{\text{best}}$ and $U_{K=46}$. The adaptation for time-varying demand is further demonstrated in cases (c-1) and (c-2). In case (c-1), the demand distribution pattern is

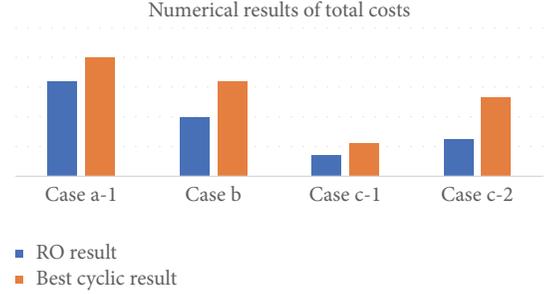


FIGURE 11: Total costs for different results in different cases.

similar to cases (a-1) and (b), and the cost reduction of the noncyclic timetable is closer to the other cases (36.08%). In case (c-2), the special event leads to a more heterogeneous demand distribution, and the cost reduction reaches 53.35%. Besides, the computation times (shown below each case) are acceptable for such a major and busy metro line.

However, in cases (a-2) and (a-3), the passenger wait times for noncyclic timetables are larger than for cyclic ones. This occurs because the RO method considers the objective function that corresponds to an infeasible headway as $+\infty$. With a large demand, the first departure headway that is feasible may greatly exceed h^* (as can be seen in Figure 4), whereas the cyclic timetable does not consider the feasibility and divides headways uniformly. In other words, the increase of departure headways (leading to larger T_{pw}) is due to safety issues, which is consistent with the actual operational requirements that put safety ahead of costs. In addition, with parameters given in this experiment, the largest frequency for BML4 is $K_N = 54$, as can be concluded by comparing cases (a-2) and (a-3). Therefore, real-time dispatching and passenger flow control methods should be applied with passenger demand at or above the levels in case (a-2).

Figures 12, 13, and 14 illustrate the noncyclic timetables obtained in case (a-1), case (b), and case (c-2), respectively. The circles here represent the occurrences of maximum dwell times. These figures illustrate that the headways vary in accordance with demand patterns, which is consistent with the results in the small case study. Particularly in Figure 14, the dense headways in the middle are set to transport many passengers due to the special event. This shows the flexibility of noncyclic timetables in special scenarios.

In addition, from the distribution of the circles, we can see that case (a-1) is the busiest scenario among these figures. The many circles in case (a-1) help explain the results in cases (a-2) and (a-3) (which have larger demands than case (a-1)). With so many trains taking maximum dwell times to load passengers, real-time dispatching and passenger flow control methods are necessary. In case (b), the maximum dwell times mostly occur at stations 7, 11, and 12. Specifically, the neighborhood of station 7 (HDHZ) includes famous schools, fashion shopping streets, and dense office buildings. The neighborhood of station 11 (BJZ) includes a major bus hub, the Beijing Zoo, and one of the largest wholesale markets. The neighborhood of station 12 (XZM), which has transfers with two other busy metro lines, includes a railway station, large

TABLE 8: Computational results for BML4.

Scenario			T_{pw} (10^8)	T_{pt} (10^8)	E_{total} (10^4)	F_s	C_{total} (10^6)	Gap (%)
Case (a-1) 3869 s	Non-cyclic	$N_{K=46}$	3.8665	3.0555	7.4898	31	3.1788	-
	Cyclic	$U_{K=44}^{best}$	5.3237	3.0943	7.2385	30	3.9925	20.38
	Cyclic	$U_{K=46}$	4.9805	3.0985	7.4898	31	$+\infty$	-
Case (a-2) 1544 s	Non-cyclic	$N_{K=54}$	11.426	4.3898	9.2418	37	7.7869	-
	Cyclic	$U_{K=54}$	10.653	4.4060	9.2418	37	$+\infty$	-
Case (a-3) 1133 s	Non-cyclic	$N_{K=54}$	19.568	4.7953	9.4566	38	12.426	-
	Cyclic	$U_{K=54}$	18.891	4.816	9.4958	38	$+\infty$	-
Case (b) 3932 s	Non-cyclic	$N_{K=38}$	2.2416	2.1501	5.9624	25	1.9902	-
	Cyclic	$U_{K=30}^{best}$	4.4170	2.1632	4.9379	21	3.1748	37.31
Case (c-1) 6848 s	Non-cyclic	$N_{K=30}$	0.5482	1.0256	4.3421	19	0.7004	-
	Cyclic	$U_{K=22}^{best}$	1.2744	1.0946	3.3695	14	1.0957	36.08
Case (c-2) 4068 s	Non-cyclic	$N_{K=41}$	1.1415	1.5061	6.0564	26	1.2342	-
	Cyclic	$U_{K=25}^{best}$	3.8100	1.5533	3.9911	17	2.6459	53.53

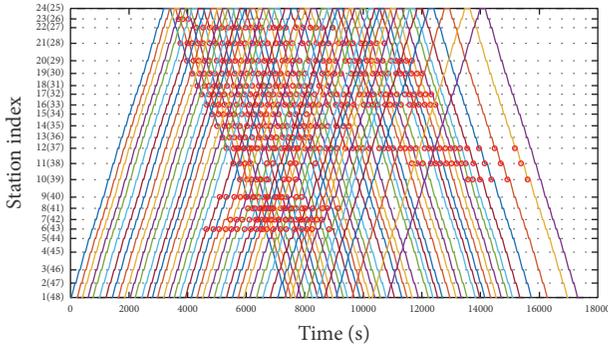


FIGURE 12: Timetable for morning peak (case a-1).

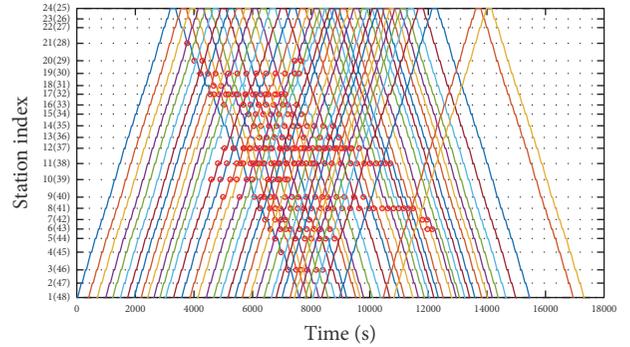


FIGURE 13: Timetable for evening peak (case b).

shopping malls, and office buildings. They are well known for the large numbers of passengers during peak hours. Thus, we find that the model is reasonable and consistent with real-world scenarios. Analogously, in case (c-2), the circles are mostly seen at stations near the station XD, where the special event is held.

It should be noted that the last two trains have an unusually small headway in case (b). The reason is that we assume that the last train must arrive at the first station at the end of the period. It is expected that if there is still demand after 10800 s, the headways will be optimized by our approaches.

5.3. Additional Experiments. In this section, we implement three sets of additional experiments to investigate the relations between the values of costs and the objectives. The first experiment tests the sensitivity of passenger time and peak frequency to unit passenger time cost, based on the demand in case (b) of BML4 study. In practice, the unit time value of passenger travel time is often estimated at half of passenger wait time. Thus, we set $c_{pt} = 0.5 \cdot c_{pw}$ and obtain the results that correspond to different values of c_{pw} , as shown in Figure 15. Results show that as time value increases, total passenger wait time tends to decrease, while the frequency within the peak hour tends to increase. They demonstrate

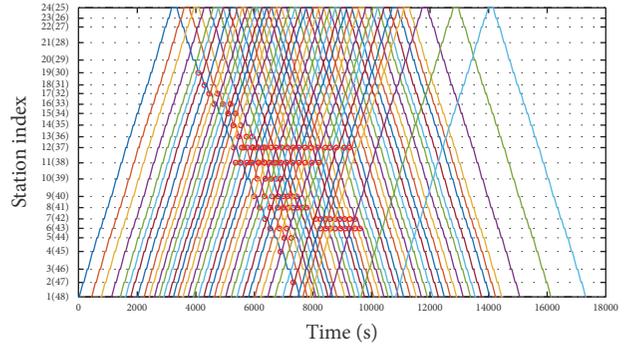


FIGURE 14: Timetable for a special event (case c-2).

the tradeoffs between passenger-oriented objectives (service level) and operator-oriented objectives (train cost). It may be expected that if the average value of time continues growing, the system manager should add more trains to improve passenger satisfaction.

In addition, neither passenger wait time nor peak hour frequency changes are strictly monotonic. This occurs because the minimum headway constraints would somehow limit the flexibility of the RO method, especially based on a major metro line (as described above), where very different

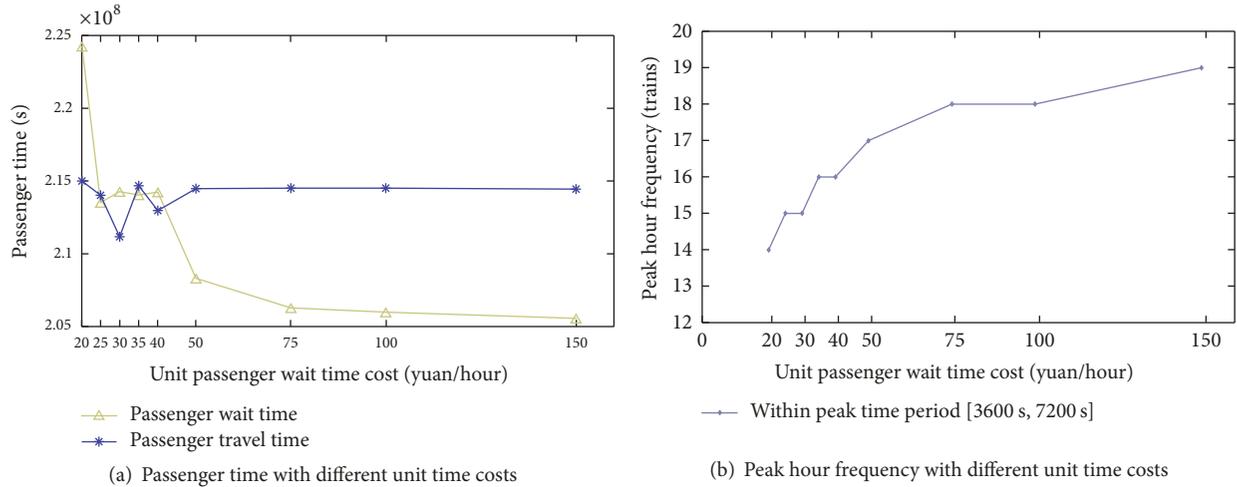


FIGURE 15: Sensitivity to unit passenger time cost.

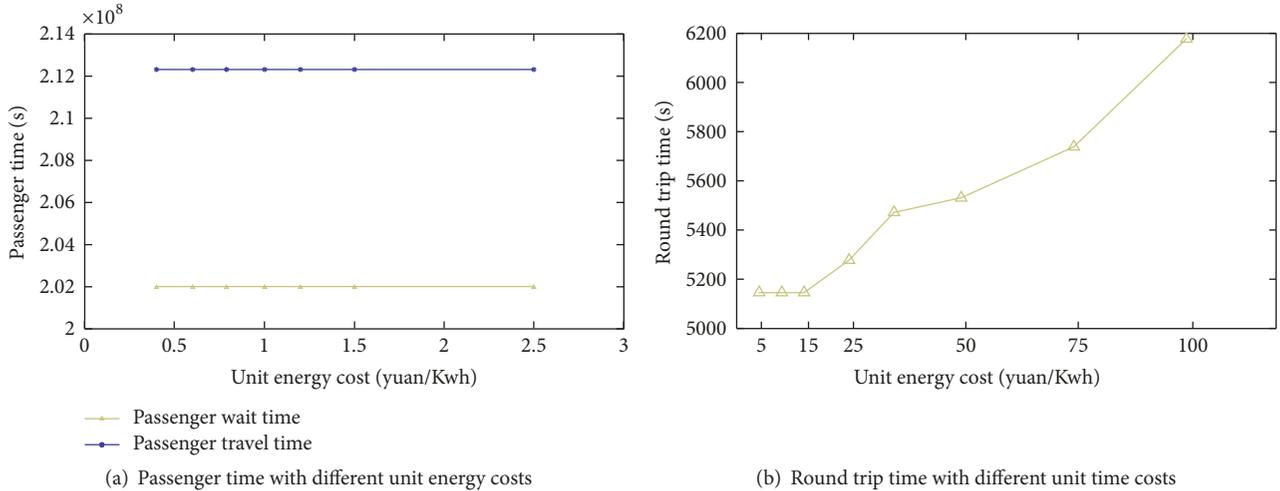


FIGURE 16: Sensitivity to unit energy cost.

distributions of passenger demand lead to varying dwell times. Thus, in the real-world operations, safety issues may limit improvements in service level. The timetables respond well to different values of time.

Besides, we find that total passenger travel time barely changes. In fact, all the segment travel times reach minimum values. The reason will be specifically discussed in the second experiment.

The second experiment tests how unit energy cost affects passenger time and train round trip time, based on the passenger demand in case (c-1) of BML4. Figure 16(a) demonstrates that when unit energy cost changes within a practical range, neither passenger wait time nor travel time is affected. The minimum segment travel times are obtained in all cases. This can be seen in daily operations that, with large enough passenger demand, the metro operators would run trains at the fastest speed, that is, with minimum segment travel times.

To extend the discussion, we run more experiments to show the relations between larger unit energy costs and round trip time and obtain the results shown in Figure 16(b). Although such large unit energy costs are unrealistic, we can see that the round trip time increases; that is, the average speed decreases, as energy cost increases. In other words, as the ratio between unit energy cost and passenger demand increases, round trip time increases. This result is consistent with the practical operation strategy that trains should run faster with more passengers and slower when they carry fewer passengers (where the unit energy cost is fixed, and thus the ratio depends on the passenger demand).

The third experiment aims to test the sensitivity of the passenger wait time and the total cost to train capital cost, based on the passenger demand in case (c-1). Figure 17 illustrates that, with the increased train capital cost, the passenger wait time and total cost tend to increase. The tradeoffs between passenger wait time and train cost, that is,

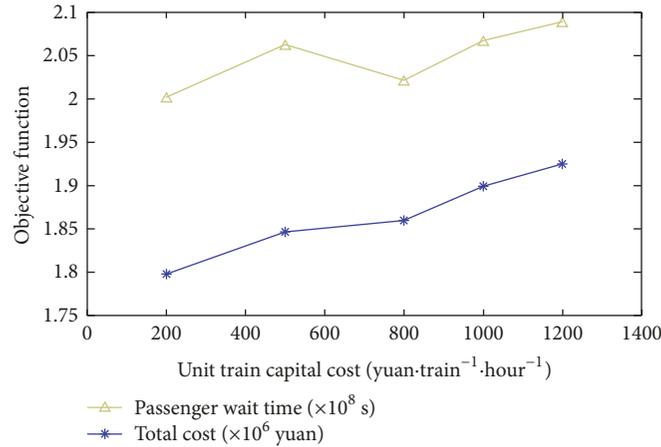


FIGURE 17: Sensitivity to train capital cost.

the passenger-oriented and operator-oriented objectives, is again demonstrated. The fluctuation of passenger wait time is caused by the minimum headway constraints, as explained for the first experiment. In addition, the monotonically increasing total cost demonstrates that technologies which increase the train's economic life help reduce total cost in a metro system.

Generally, we obtain reasonable and practical results with the proposed solution approach. These results are consistent with practical experiences and can be used as decision-making support in daily operations. Since similar results are obtained with different passenger demand patterns and on different metro lines, the predictive ability of the model is demonstrated. The effectiveness of the proposed approach is further verified.

6. Conclusions

This paper investigates the train timetabling problem (TTP) with time-varying passenger demand and accomplishes the following:

(1) It accounts for both the passenger-oriented and operator-oriented objectives. Factors in the line planning stage, the timetabling stage, and the vehicle scheduling stage are integrated in our model and solution approach. The noncyclic timetable obtained in this paper outperforms the cyclic timetable, the noncyclic timetables obtained by the generic GA, and a classic direct optimization approach.

(2) The dwell time determining process is specifically designed to consider both congestion at stations and passenger demand. Besides, limited train capacity and varying passenger alighting/boarding rates are accounted for. Thus, the noncyclic timetable obtained in this paper is more practical in depicting passenger loading evolution. For this reason, some timetables with otherwise good properties would violate the minimum headway constraint at intermediate stations and thus be infeasible. Passenger flow control methods and real-time dispatching can be employed to make these timetables applicable.

(3) Based on a general scheduling scheme, the TTP is decomposed into the OST and OH subproblems. The solution

approach includes a branch-and-bound (B&B) algorithm, a frequency determining algorithm, and a novel rolling optimization (RO) method. Then, computational results based on a short metro line and the Beijing Metro Line 4 are obtained. They verify the effectiveness and efficiency of the proposed solution approach and demonstrate that the noncyclic timetables obtained in this paper respond well to passenger demand characteristics. Since the results are based on different cases with different passenger distributions, the predictive ability of the proposed solution approach can also be demonstrated.

(4) Additional experiments are presented whose results demonstrate the following points: (a) As passenger time values increase, passenger wait time may be further reduced by scheduling more trains or by shifting departure times for riders. (b) Average train speed decreases as passenger demand increases, due to longer dwell times at stations. This dwell time effect outweighs the speed increase justified by more riders per train. (c) Longer train economic life not only reduces total system cost, but also indirectly improves the service level.

This paper has some limitations. Specifically, the linear piecewise approximation simplifies the energy calculation. Assumption 1 is consistent with a common strategy in practice but does not account for the relevant effects of driver behaviors. Assumption 4 may be extended to a more accurate model which distributes the passenger boarding and alighting over the station dwell times. A more systematic comparison with real data of BML4 in the case study should be conducted to fully validate RO method.

These drawbacks can be addressed by a more complex model which considers driver behaviors (by estimating the magnitude of their relevance), line layouts, and physics of real train run process, as well as simultaneously optimizing departure headways and segment travel times under continuous time. Future studies can investigate the improvement of our solution approach, to reduce the computation time needed for such a complex model. Besides, it would also be interesting to adjust the model's parameters with respect to different real-world operation data, to obtain more realistic timetables.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

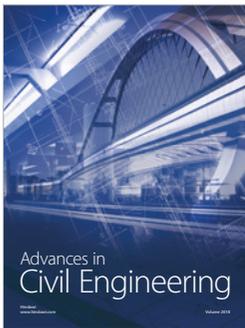
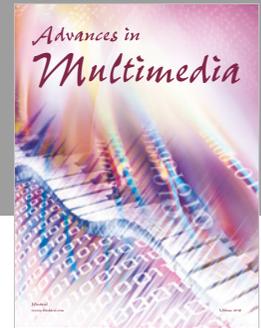
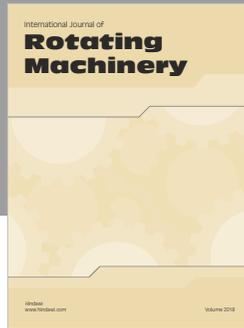
Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant no. U1434209), National Key Research and Development Program of China (2017YFB1201105), and Research Foundation of State Key Laboratory of Railway Traffic Control and Safety, Beijing Jiaotong University (Grant no. RCS2018ZZ003).

References

- [1] S. Burggraeve, S. H. Bull, P. Vansteenwegen, and R. M. Lusby, "Integrating robust timetabling in line plan optimization for railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 134–160, 2017.
- [2] T. Robenek, Y. Maknoon, S. S. Azadeh, J. Chen, and M. Bierlaire, "Passenger centric train timetabling problem," *Transportation Research Part B: Methodological*, vol. 89, pp. 107–126, 2016.
- [3] Y. Wang, T. Tang, B. Ning, T. J. J. van den Boom, and B. De Schutter, "Passenger-demands-oriented train scheduling for an urban rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 1–23, 2015.
- [4] J. Yin, L. Yang, T. Tang, Z. Gao, and B. Ran, "Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches," *Transportation Research Part B: Methodological*, vol. 97, pp. 182–213, 2017.
- [5] T. Robenek, S. Sharif Azadeh, Y. Maknoon, and M. Bierlaire, "Hybrid cyclicity: Combining the benefits of cyclic and non-cyclic timetables," *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 228–253, 2017.
- [6] A. Caprara, M. Fischetti, and P. Toth, "Modeling and solving the train timetabling problem," *Operations Research*, vol. 50, no. 5, pp. 851–861, 2002.
- [7] E. Barrena, D. Canca, L. C. Coelho, and G. Laporte, "Single-line rail rapid transit timetabling under dynamic passenger demand," *Transportation Research Part B: Methodological*, vol. 70, pp. 134–150, 2014.
- [8] A. Schöbel, "An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 348–365, 2017.
- [9] A. Berbey, R. Galan, J. D. Sanz Bobi, and R. Caballero, "A fuzzy logic approach to modelling the passengers' flow and dwelling time," in *Proceedings of the 18th International Conference on Urban Transport and the Environment, UT12*, pp. 359–369, Spain, May 2012.
- [10] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.
- [11] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.
- [12] V. Cacchiani, A. Caprara, and P. Toth, "A column generation approach to train timetabling on a corridor," *4OR*, vol. 6, no. 2, pp. 125–142, 2008.
- [13] T. Dollevoet, D. Huisman, M. Schmidt, and A. Schöbel, "Delay management with rerouting of passengers," *Transportation Science*, vol. 46, no. 1, pp. 74–89, 2012.
- [14] T. Dollevoet, F. Corman, A. D'Ariano, and D. Huisman, "An iterative optimization framework for delay management and train scheduling," *Flexible Services and Manufacturing Journal*, vol. 26, no. 4, pp. 490–515, 2014.
- [15] L. G. Kroon, L. W. P. Peeters, J. C. Wagenaar, and R. A. Zuidwijk, "Flexible connections in PESP models for cyclic passenger railway timetabling," *Transportation Science*, vol. 48, no. 1, pp. 136–154, 2014.
- [16] X. Luan, J. Miao, L. Meng, F. Corman, and G. Lodewijks, "Integrated optimization on train scheduling and preventive maintenance time slots planning," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 329–359, 2017.
- [17] S. Su, X. Li, T. Tang, and Z. Gao, "A subway train timetable optimization approach based on energy-efficient operation strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 883–893, 2013.
- [18] X. Li and K. Lo Hong, "An energy-efficient scheduling and speed control approach for metro rail operations," *Transportation Research Part B: Methodological*, vol. 64, pp. 73–89, 2014.
- [19] A. Higgins, E. Kozan, and L. Ferreira, "Optimal scheduling of trains on a single line track," *Transportation Research Part B: Methodological*, vol. 30, no. 2, pp. 147–161, 1996.
- [20] J. Wang, Y. Yu, R. Kang, and J. Wang, "A novel space-time-speed method for increasing the passing capacity with safety guaranteed of railway station," *Journal of Advanced Transportation*, vol. 2017, Article ID 6381718, 2017.
- [21] X. Xu, K. Li, and L. Yang, "Scheduling heterogeneous train traffic on double tracks with efficient dispatching rules," *Transportation Research Part B: Methodological*, vol. 78, pp. 364–384, 2015.
- [22] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: branch-and-bound algorithms with enhanced lower bounds," *Transportation Research Part B: Methodological*, vol. 41, no. 3, pp. 320–341, 2007.
- [23] L. Yang, K. Li, Z. Gao, and X. Li, "Optimizing trains movement on a railway network," *Omega*, vol. 40, no. 5, pp. 619–633, 2012.
- [24] S. Yang, K. Yang, Z. Gao, L. Yang, and J. Shi, "Last-Train Timetabling under Transfer Demand Uncertainty: Mean-Variance Model and Heuristic Solution," *Journal of Advanced Transportation*, vol. 2017, pp. 1–13, 2017.
- [25] L. Yang, Y. Zhang, S. Li, and Y. Gao, "A two-stage stochastic optimization model for the transfer activity choice in metro networks," *Transportation Research Part B: Methodological*, vol. 83, pp. 271–297, 2016.
- [26] G. Malavasi, P. Palleschi, and S. Ricci, "Driving and operation strategies for traction-energy saving in mass rapid transit systems," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 225, no. 5, pp. 475–482, 2011.
- [27] J. T. Haahr, J. C. Wagenaar, L. P. Veelenturf, and L. G. Kroon, "A comparison of two exact methods for passenger railway rolling stock (re)scheduling," *Transportation Research Part E: Logistics and Transportation Review*, vol. 91, pp. 15–32, 2016.
- [28] X. Yang, A. Chen, X. Li, B. Ning, and T. Tang, "An energy-efficient scheduling approach to improve the utilization of

- regenerative energy for metro systems,” *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 13–29, 2015.
- [29] Y.-C. Lai, D.-C. Fan, and K.-L. Huang, “Optimizing rolling stock assignment and maintenance plan for passenger railway operations,” *Computers & Industrial Engineering*, vol. 85, pp. 284–295, 2015.
- [30] V. Schmid and J. F. Ehmke, “Integrated timetabling and vehicle scheduling with balanced departure times,” *OR Spectrum*, vol. 37, no. 4, pp. 903–928, 2015.
- [31] V. Guihaire and J. K. Hao, “Transit network design and scheduling: a global review,” *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1251–1273, 2008.
- [32] L. Cadarso and A. Marin, “Integration of timetable planning and rolling stock in rapid transit networks,” *Annals of Operations Research*, vol. 199, pp. 113–135, 2012.
- [33] J. M. Mera, E. Carabaño, M. Soler, and E. Castellote, “Increasing metro line capacity by optimisation of track circuit in a speed code Automatic Train Protection system,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 1, pp. 165–180, 2016.
- [34] S. Chiusolo, A. Dicembre, S. Ricci, and F. Sorace, “Automation of high density metro lines: Rome line a case study,” in *Proceedings of the 13th International Conference on Automated People Movers and Transit Systems 2011: From People Movers to Fully Automated Urban Mass Transit*, pp. 491–501, France, May 2011.
- [35] V. Cacchiani, F. Furini, and M. P. Kidd, “Approaches to a real-world Train Timetabling Problem in a railway node,” *OMEGA - The International Journal of Management Science*, vol. 58, pp. 97–110, 2016.
- [36] X. Xu, K. Li, and X. Li, “A multi-objective subway timetable optimization approach with minimum passenger time and energy consumption,” *Journal of Advanced Transportation*, vol. 50, no. 1, pp. 69–95, 2016.
- [37] L. G. Kroon and L. W. P. Peeters, “A variable trip time model for cyclic railway timetabling,” *Transportation Science*, vol. 37, no. 2, pp. 198–212, 2003.
- [38] D. Sparing and R. M. P. Goverde, “A cycle time optimization model for generating stable periodic railway timetables,” *Transportation Research Part B: Methodological*, vol. 98, pp. 198–223, 2017.
- [39] F. Corman, A. D’Ariano, A. D. Marra, D. Pacciarelli, and M. Samà, “Integrating train scheduling and delay management in real-time railway traffic control,” *Transportation Research Part E: Logistics & Transportation Review*, 2016.
- [40] A. D’Ariano, D. Pacciarelli, and M. Pranzo, “A branch and bound algorithm for scheduling trains in a railway network,” *European Journal of Operational Research*, vol. 183, no. 2, pp. 643–657, 2007.
- [41] K. Ghoseiri, F. Szidarovszky, and M. J. Asgharpour, “A multi-objective train scheduling model and solution,” *Transportation Research Part B: Methodological*, vol. 38, no. 10, pp. 927–952, 2004.
- [42] J. Qi, L. Yang, Y. Gao, S. Li, and Z. Gao, “Integrated multi-track station layout design and train scheduling models on railway corridors,” *Transportation Research Part C: Emerging Technologies*, vol. 69, pp. 91–119, 2016.
- [43] S. Li, M. M. Dessouky, L. Yang, and Z. Gao, “Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines,” *Transportation Research Part B: Methodological*, vol. 99, pp. 113–137, 2017.
- [44] A. Schobel, “Line planning in public transportation: models and methods,” *OR Spectrum*, vol. 34, no. 3, pp. 491–510, 2012.
- [45] S. Samanta and M. K. Jha, “Modeling a rail transit alignment considering different objectives,” *Transportation Research Part A: Policy and Practice*, vol. 45, no. 1, pp. 31–45, 2011.
- [46] H. L. Fu, L. Nie, L. Y. Meng, B. R. Sperry, and Z. H. He, “A hierarchical line planning approach for a large-scale high speed rail network: the China case,” *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 61–83, 2015.
- [47] E. Cascetta and P. Coppola, “Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services,” *Transportation Research Part A: Policy and Practice*, vol. 84, pp. 93–108, 2016.
- [48] J. Yin, T. Tang, L. Yang, Z. Gao, and B. Ran, “Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: an approximate dynamic programming approach,” *Transportation Research Part B: Methodological*, vol. 91, pp. 178–210, 2016.
- [49] A. P. Cudala, A. Fernández, C. Sicre, and M. Domínguez, “Fuzzy optimal schedule of high speed train operation to minimize energy consumption with uncertain delays and drivers behavioral response,” *Engineering Applications of Artificial Intelligence*, vol. 25, no. 8, pp. 1548–1557, 2012.
- [50] H. Huang, K. Li, and Y. Wang, “A simulation method for analyzing and evaluating rail system performance based on speed profile,” *Journal of Systems Science and Systems Engineering*.
- [51] R. Hooke and R. Jeeves, “Direct search solution of numerical and statistical problems,” *Journal of the ACM*, vol. 8, pp. 212–229, 1961.
- [52] W. C. Davidon, “Variable metric method for minimization,” *SIAM Journal on Optimization*, vol. 1, no. 1, pp. 1–17, 1991.
- [53] W. Duan, Y. Chen, and J. Lai, *Passenger flow characteristics in Beijing Subway Line 4. Urban Rapid Rail Transit*, vol. 26, 43–46, 26, 2013.



Hindawi

Submit your manuscripts at
www.hindawi.com

