Demand-oriented train services optimization for a congested urban rail line: Integrating short turning and heterogeneous headways

Sijie Li[a,b], Ruihua Xu[b], Ke Han[c]

[a]*School of Urban Rail Transportation, Shanghai University of Engineering Science, Shanghai 201620, People's Republic of China*

[b]*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, People's Republic of China*

[c]*Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

Demand-oriented train services optimization for a congested urban rail line: Integrating short turning and heterogeneous headways

This paper focuses on the demand-oriented passenger train scheduling problem for a congested urban rail line, simultaneously considering uneven spatial and temporal demand distributions. To account for the left-behind phenomenon and passenger selection behavior, a passenger-train interaction framework is developed to dynamically assign passengers to capacitated trains. A mixed integer nonlinear programming model that combines heterogeneous headways and short turning as an integrated strategy (HH-ST) is proposed with the aim of jointly minimizing the passenger waiting time and operational costs, and balancing train loads. A two-stage genetic algorithm based on an integer coding approach is proposed to solve this problem efficiently. To verify the effectiveness of the proposed method, the HH-ST strategy is compared with alternative strategies, namely short-turning alone, heterogeneous headways alone and regular schedule (with no strategy), through a real-world case study of Shanghai Metro Line 9. The results show that the HH-ST strategy provides a better trade-off between users' cost and operators' cost than other strategies, thus achieving a better match between transport capacity and passenger demand.

Keywords: Train scheduling; dynamic passenger demand; heterogeneous headways; short turning; urban rail transit

1. Introduction

In order to alleviate the current situation of urban road congestion, public transport has drawn more and more attention in recent years. Due to the characteristics of large carrying capacity, low carbon emissions, fast speed and high reliability, urban rail transit plays an important role in public transport system of big cities (e.g., New York, Tokyo, Beijing and Shanghai). With the gradual expansion of the urban rail transit network, the corresponding travel demand continues to increase, presenting significant peaks in both temporal and spatial dimensions. This results in more conspicuous contradictions between supply and demand, especially during rush hours. Heavily crowded trains make passengers feel uncomfortable, and extreme densities of waiting passengers on the platforms pose potential security risks. In view of this, transport capacity supply should be reallocated in an efficient and reasonable manner to meet increasing demand. Therefore, an important issue emerges regarding how to schedule trains based on the

temporal-spatial heterogeneous passenger demand to jointly benefit the operators and the passengers, in the operation and management of urban rail transit.

To optimize temporal distribution of train services, the train timetabling problem has been studied in two primary forms over the past few decades. One kind of studies is focused on constructing the periodic timetables with fixed headway throughout the planning horizon, typically based on the periodic event scheduling problem (PESP) (Liebchen 2008; Serafini and Ukovich 1989). Periodic timetables have the advantages of being easily memorized by passengers, and are able to reduce the passenger waiting time in the case of constant demand, which are often used in practice for large railway systems (Kroon et al. 2009; Sparing and Goverde 2017). However, passengers in the urban rail system tend to ignore the periodic timetables if the fixed headways are short, leading to stochastic passenger demands. In addition, travel demands are time-varying due to the tidal effect of urban travel. In this case, fixed-headway timetables may result in longer passenger waiting times in oversaturated periods because passengers might fail to board the coming train and have to wait for the next one, or it might lead to low train capacity utilization in unsaturated periods due to the lack of passengers. Thus, understanding the temporal demand variation and adjusting service headway dynamically become crucial.

The availability of detailed origin-destination (OD) trip records obtained from automated fare collection (AFC) systems allows researchers to design demand-oriented, non-periodic timetables to improve the service level of rail transit systems. Sun et al. (2014) formulated three optimization models to design demand-sensitive timetables by demonstrating train operation using equivalent time interval, and concluded that dynamical timetabling with capacity constraints is advantageous over fixed, non-demand-responsive timetables. Barrena et al. (2014a, 2014b) constructed a timetable according to the dynamic demand patterns with the aim of minimizing passenger average waiting time, and presented a branch-and-cut algorithm and adaptive large neighborhood search metaheuristic to solve small and large instances, respectively. Addressing the penalty for left-behind passengers, Niu and Zhou (2013) focused on optimizing a passenger train timetable for an urban rail corridor under heavily congested conditions and limited train fleet availabilities. In a subsequent study, Niu, Zhou and Gao (2015) developed a unified quadratic integer programming model to jointly synchronize effective passenger loading time windows and train timetable with given and predetermined train skip-stop patterns. Different from these researches where objectives that are only related to the user cost, Yin et al. (2017) proposed an integrated approach for the train scheduling problem in order

to minimize both the energy consumption and passenger waiting time. In addition, taking passenger transfers into account in the process of train scheduling, Wang et al. (2015) and Niu, Tian and Zhou (2015) have attempted to set up a demand-oriented timetable for multiple train services in a connected transit network.

Regarding the spatial imbalances of passenger demand, some studies focus on designing efficient operation strategies in the process of train scheduling to improve the match between demand and supply over the space dimension (Chang, Yeh and Shen 2000; Ulusoy, Chien and Wei 2010; Cortés, Jara-Díaz and Tirachini 2011). For urban rail transit trunk lines passing through city center or radial lines connecting distant residential area with city center, providing the same services along the entire line may be inefficient, as central zones and suburb zones face different levels of demand. In this case, short-turning is a convenient operational tactic that a portion of trains are selected to serve short cycles on those sections with high demand. Short turning can make full use of the transport capacity to meet the passenger demand in different sections, helping to reduce operator costs.

At early stages, most of the literature about short-turning is analyzed in a bus context. Furth (1987) and Ceder (1989) are among the early investigators who optimized short turn operations for buses. Furth (1987) found the schedule offset between the full-length and short-turn patterns that will balance loads and minimize fleet size and wait time, by recognizing the importance of coordinating these patterns. Ceder (1989) presented a set of procedures to automatically identify feasible short-turn points based on passenger load profile data, and derive the minimum fleet size required by a given bus timetable. Delle Site and Filippi (1998) provided a more comprehensive optimization framework for bus operations including short-turning and variable vehicle size, and developed a net benefit maximization model under constant as well as elastic demands. Tirachini, Cortés and Jara-Díaz (2011) developed a short turning model that considers both operators' and users' costs. This is the first work that finds analytical expressions for optimal frequencies both inside and outside the short cycle. By further observing the unbalances within and between directions of demands, Cortés, Jara-Díaz and Tirachini (2011) combined short turning and deadheading in an integrated strategy to deal with mixed load patterns.

For a rail transit system, Ghoneim and Wirasinghe (1986) optimized the train service by adopting a zone-stop schedule during the peak periods, and a dynamic programming technique is employed to determine the number of zones, stations that belong to each zone and headway between trains to the same zone. Sun et al. (2016) extended existing studies mainly

by relaxing the assumption that the full length route must be operated, and designed flexible models to accommodate more complex combinations of train routes and calculate potential waiting costs if no direct services are available for some origin-destination pairs. Zhu, Mao and Zhou (2017) considered explicit capacity constraints, and developed an optimization model to find an efficient short-turning strategy for a bi-directional urban rail line aiming at reducing total passenger waiting time. Moreover, the short turning strategy is also used to control the operations of a rail line in real time (Canca et al. 2012; Canca et al. 2016). For example, Canca et al. (2016) applied short turning strategy to manage passenger overloads after an episode of demand increase. By means of inserting special short-turning services in an efficient manner, they optimized the location of turn-back points and service offset with the objective of diminishing the passenger waiting time while preserving certain level of quality of service.

Table 1. Summary of relevant publications

Author (Year)	Objective	Demand	Operation strategies	Decision variables	Model	Left behind
Sun et al. (2014)	Waiting time	Time-varying		Train departure times	Mixed integer linear	√
Barrena et al. (2014a)	Waiting time	Time-varying		Train departure times	Nonlinear	
Barrena et al. (2014b)	Waiting time	Time-varying		Train departure times	Linear	
Niu and Zhou (2013)	Waiting time	Time-varying		Train departure and arrival times	Integer nonlinear	√
Niu, Zhou and Gao (2015)	Waiting time	Time-varying	SS	Train departure and arrival times	Mixed integer nonlinear	
Yin et al. (2017)	Waiting time and energy consumption	Time-varying		Train departure and arrival times, speed profile	Mixed integer linear	
Delle Site and Filippi (1998)	User and operator costs	Constant	ST	Turn-back points, offset, frequencies, vehicle size	Numerical method	
Tirachini, Cortés and Jara-Díaz (2011)	User and operator costs	Constant	ST	Turn-back points, frequencies, vehicle size	Numerical method	

Cortés, Jara-Díaz and Tirachini (2011)	User and operator costs	Constant	ST&DE	Turn-back points, frequencies, vehicle size	Numerical method	
Ghoneim and Wirasinghe (1986)	User and operator costs	Constant	ST	Turn-back points, headway, fleet size	Linear	
Ulusoy, Chien and Wei (2010)	User and operator costs	Constant	ST& EX	Frequencies	Mixed integer nonlinear	
Sun et al. (2016)	User and operator costs	Constant	ST	Turn-back points, frequencies	Mixed integer nonlinear	
Zhu, Mao and Zhou (2017)	Waiting time	Constant	ST	Turn-back points, offset, frequencies	Mixed integer nonlinear	√
Canca et al. (2016)	Waiting time	Constant	ST	Turn-back points, offset	Mixed integer nonlinear	
This paper	User and operator costs Train loads balance	Time-varying	ST	Train departure and arrival times, Turn-back points, frequencies	Mixed integer nonlinear	√

Key: ST= Short-turn; SS= Skip-stop; DE= Deadheading; EX= Express

For the ease of comparison, we list the detailed characteristics of some closely related studies in Table 1, which are divided into train timetabling problem and train operation strategies optimization problem. To the best of our knowledge, few studies that simultaneously consider the temporal and spatial heterogeneity of passenger demand in the design of demand-oriented rail transit train services. On the one hand, as we can see from the first part of Table 1, most studies addressing the time-varying OD passenger demands in an integrated design of train schedules always use the same train service along the entire route and over the planning period, they did not consider short-turning strategy as a decision variable (Niu, Zhou and Gao (2015) only took into account predetermined skip-stop patterns). Moreover, most literature on non-periodic timetabling only consider from the passenger perspective, the objective of timetable design is mainly to reduce passenger waiting time. On the other hand, according to the second part of Table 1, the researches on the spatial imbalance of passenger demand mainly optimize the short-turning strategies at a tactical level, their decision variables include the location of turn-back points, service offset and the associated frequencies within and outside the short cycle, in which the temporal dynamic characteristics of passenger demand are

simplified. Furthermore, most studies ignore the left behind phenomenon, despite this often happens during peak hours in actual operation.

To address these issues, this paper proposes a unified train scheduling framework that synchronously considers both spatial and temporal imbalances of passenger demand, integrating heterogeneous headways and short-turning strategy for a heavily congested urban rail line. Specifically, this study presents the following three contributions.

1) With given time-varying OD passenger demand data, a passengers-trains interaction model is developed to dynamically assign all passengers to each capacitated train. Under oversaturated conditions, the fail-to-board phenomenon can be captured, where some passengers cannot board a full train and must wait for the next one. Besides, the passengers' selection behavior over different route pattern is also taken into account.

2) A mixed integer nonlinear programming model is formulated to minimize both passenger cost and operator cost, and an additional objective of equilibrating vehicle occupancy levels over all the segments is considered to preserve the supply-demand match, which makes the model more practical. To solve the model more efficiently, a two-stage heuristic algorithm is proposed to obtain a good solution in an acceptable computational time.

3）For practical purposes, we compare the proposed integrated strategy with separate heterogeneous headways strategy, separate short-turning strategy and regular schedule (no strategy). These benchmark models are also optimized based on the same inputs, which allows us to better characterize the benefits of the integrated strategy on a fair basis.

The remainder of this paper is structured as follows. A detailed problem statement and assumptions are presented in Section 2. Section 3 presents the calculation methods of the interaction process between passengers and trains. A mathematical model is developed to generate the short-turning strategy and train arrival and departure times in Section 4. Then, a two-stage genetic algorithm based on an integer coding approach is designed to solve this model in Section 5. In Section 6, the performance of the proposed model is evaluated via a case study of Shanghai Metro Line 9. The last section summarizes the paper and discusses future work.

2. Problem statement

This paper considers an urban rail transit line with short turning strategy and heterogeneous headways, as shown in Figure 1. The stations are numbered as $1,2,...,N$, sequentially. The segment between station i and station $i+1$ is denoted as section $(i,i+1)$. There are two

different operation directions, assuming that the passenger demand in two directions are independent to each other, this study only considers the up direction (from station 1 to N).

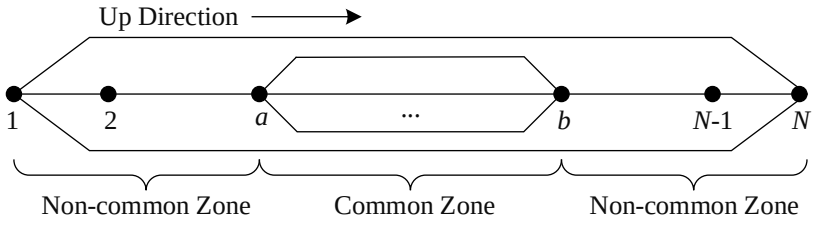


Figure 1. Urban rail line with short turning operation mode

The train operation mode includes two train route patterns. The full-length route pattern covers all stations between station 1 and station N while the short-turn route pattern runs between station a and station b, where a and b are turn back stations. The overlapping sections between the two routes are called common zone, and the other sections are non-common zone. Correspondingly, trains are also divided into two categories, full-length trains and short-turn trains, which run in accordance with each route pattern, respectively.

Depending on the station layout and track setting, limited number of stations can be qualified for trains' turn back activities for an urban rail transit line. A train route, which is formed by selecting two turn back station as a pair of terminals, such as "$a-b$" in Figure 1, is indexed by r for brevity. The set of feasible train routes can be easily predetermined by the service provider.

We take rush hours as the planning period, denoted as $[0,T]$. In order to describe the time-varying characteristic of passenger flow, this paper divides the planning period into equal tiny time intervals, which can be 30 seconds, 1 min, etc., depending on the level of modeling accuracy (Sun et al. 2014; Niu and Zhou 2013). The number of passengers from each origin and to each destination in each time interval can be obtained from the automatic fare collection systems (AFC).

In summary, this paper aims to design a demand-adapted train timetable for a congested urban rail line with given time-varying origin-destination passenger demand data and candidate route set, by determining the departure times of each train, the best short-turn route pattern (specifying a pair of turn back stations) and the ratio of the number of short-turn trains and the number of full-length trains.

2.1. Assumptions

Some assumptions used throughout the paper are as follows.

A1. [Passenger choice behavior] Passengers will always choose the direct trains to reach their destinations. That is, passengers whose origin and destination are within the common zone can choose either full-length or short-turn route, whereas other passengers whose origin and destination are connected without transfer by the full-length route only is entirely assigned to the full-length route (Delle Site and Filippi 1998) .

A2. [Capacity constraint] Considering the train capacity constraint, when the number of in-vehicle passengers reaches the maximum train carrying capacity, the remaining waiting passengers on the platform will not be able to board this train and need to wait for the next one. All trains have the same capacity, and a certain degree of overload is allowed in rush hours, in other words, the actual carrying capacity may exceed the rated carrying capacity by a small proportion.

A3. [Route pattern] In order to facilitate the operation of urban rail transit, only two route patterns are allowed during peak hours, one full-length route and one short-turn route, since more routes mean greater complexities in operational organization. Assuming that the number of short-turn trains is proportional to the number of full-length trains, that is, there are a fixed number of short-turn trains between two adjacent full-length trains or a fixed number of full-length trains between two adjacent short-turn trains. The first train during planning period is assumed be a full-length train.

A4. [Train operation characteristic] All trains are assumed to follow the same speed between two consecutive stations and the same dwell time at each station. This assumption can be easily guaranteed in urban rail transit.

2.2. Notations

In order to describe the modeling process, the definitions of indices, parameters and variables used in the formulation are listed as follows.

Indices:

i, j, a, b	index of stations, $i, j \in \{1, 2, ...N\}$
$(i, i+1)$	index of sections, $(i, i+1) \in \{(1,2),(2,3),...,(N-1, N)\}$
k	index of trains, $k \in \{1, 2, ..., K\}$
t	index of time intervals, $t \in \{1, 2, ..., T\}$
r	index of routes, $r \in \{1, 2, ..., R\}$, $r = 1$ represents the full-length route, the others are short-turn route patterns

Input Parameters:

P_t^{ij} Number of passengers who arrive at station i towards station j at time t

$t_{\mathrm{ru}}^{(i,i+1)}$ Section running time from station i to station $i+1$

t_{dw}^{i} Dwell time at station i

C Vehicle rated carrying capacity

n Number of train marshalling

$\eta_{\max}$ Maximal load rate

l_r Length of route r

t_r Train travel time of route r

$I_{\min}$ Predetermined minimum headway between two consecutive trains at the same station

$I_{\max}$ Predetermined maximum headway between two consecutive trains at the same station

$\bar{\alpha}_r^i$ 1 if route r covers station i; and 0 otherwise

1 Intermediate variables:

I_k^{ij} Number of passengers who arrive at station i towards station j between the departures of train k and its preceding train

R_k^{ij} Number of passengers left behind by train k at station i towards station j

W_k^{ij} Number of waiting passengers at station i with destination station j between the departures of train k and its preceding train

WP_k^{ij} Number of potential passengers who want to board train k at station i towards station j

B_k^{ij} Number of passengers who actually board train k at station i towards station j

B_k^i Total number of passengers who actually board train k at station i

O_k^i Number of passengers alighting from train k at station i

$Q_k^{(i,i+1)}$ Number of passengers remaining on train k at section $(i,i+1)$

$E_k^{(i,i+1)}$ Load rate of train k at section $(i,i+1)$

f_r Number of departure trains of route pattern r during planning period

2 To facilitate modelling, some auxiliary binary decision variables relating to the state of trains

3 are also introduced:

φ_k A binary variable, which is 1 if train k follows full-length route; 0, train k follows short-turn route.

α_k^i A binary variable, which is 1 if train k stops at station i; 0, otherwise.

$\beta_k^{(i,i+1)}$ A binary variable, which is 1 if train k covers section $(i,i+1)$; 0, otherwise.

γ_k^{ij} A binary variable, which is 1 if train k can cover the trip from station i to station j directly; 0, otherwise.

In this model, we define four sets of decision variables:

D_k^i Departure time of train k at station i

A_k^i Arrival time of train k at station i

m Ratio of the number of short-turn trains to the number of full-length trains

y_r A binary variable, which is 1 if route pattern r is selected; 0, otherwise.

3. Trains-passengers Interaction

The operation status of urban rail transit line is the result of the interaction between train operation and passenger travel, as shown in Figure 2. Changes in the train status and the number of passengers on the platform depends on the process of boarding and alighting. Before passengers get on a train, they will consider whether the coming train can provide direct service and whether the train's remaining capacity is sufficient. Through traversing all the trains and stations, the number of passengers waiting on the platform, number of remaining passengers after each train leaves, and the loading status of each train at each section can be determined. The calculation equations for train running and passenger loading events will be described in detail in this section.

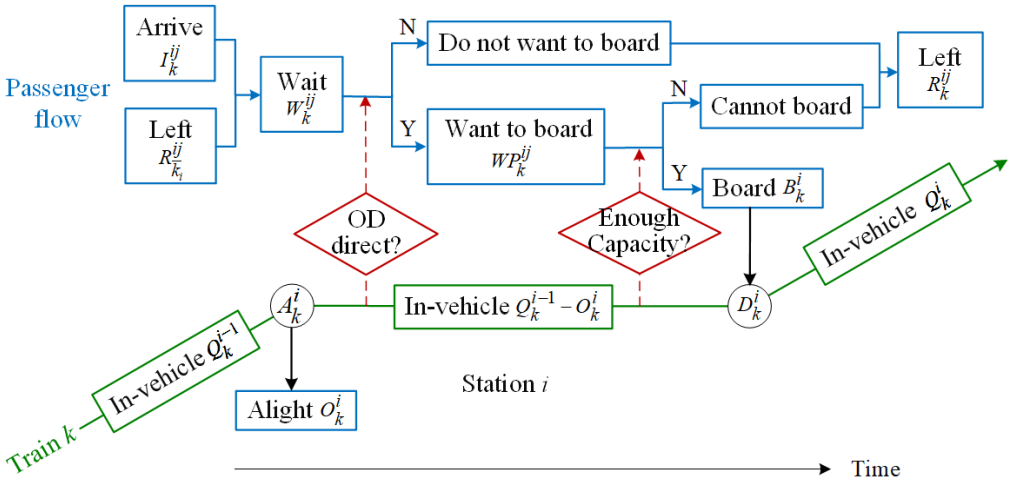


Figure 2. Illustration of interaction between trains and passengers

3.1. Operation of trains

(1) Arrival and departure times

Given section running time $t_{\mathrm{ru}}^{(i,i+1)}$ and dwell time t_{dw}^{i}, the arrival and departure times for each train at every station can be determined as below:

$$D_k^i = D_k^{i-1} + t_{\mathrm{ru}}^{(i-1,i)} + t_{\mathrm{dw}}^{i} \quad \forall k, i > 1 \tag{1}$$

$$A_k^i = D_k^i - t_{\mathrm{dw}}^{i} \quad \forall k, i \tag{2}$$

In order to reduce the number of decision variables, we add dummy running segments to short-turn trains, as shown in Figure 3, extending short-turn train to full length, but actually the short-turn trains do not run during non-common zone. Then we can simplify these variables to the departure times of trains at the start terminal D_k^1.

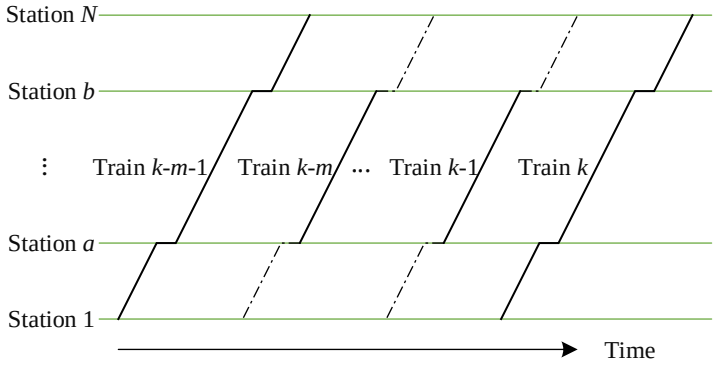


Figure 3. Illustration of train operation

(2) Attributes of trains

We define a binary variable φ_k to indicate the route pattern of train k, which is 1 if train k follows the full-length route; otherwise, it is equal to 0 when train k follows short-turn route. Based on assumption A3, the first train during planning period is a full-length train, and the ratio of the number of short-turn trains to the number of full-length trains is m, then the value of φ_k can be determined by its train index k.

$$\varphi_k = \begin{cases} 1, & \text{if } \mathrm{mod}\left(k,(1+m)\right)=1 \\ 0, & \text{otherwise} \end{cases} \quad (m \geq 1)$$

$$\varphi_k = \begin{cases} 0, & \text{if } \mathrm{mod}\left(k,(1+1/m)\right)=0 \\ 1, & \text{otherwise} \end{cases} \quad (m < 1) \tag{3}$$

We introduce a binary variable α_k^i, which is 1 if train k can serve station i, $\alpha_k^i = 0$ otherwise. Similarly, a binary variable $\beta_k^{(i,i+1)}$ is introduced to describe whether train k serves the section $(i, i+1)$ or not. The value of α_k^i first depends on whether this train follows a full-length route or a short-turn route. If train k follows full-length route, that is $\varphi_k = 1$, it will serves all stations and sections, but if train k follows the short-turn route ($\varphi_k = 0$), it depends

on which short-turn pattern r is adopted. Binary variable $y_r = 1$ denotes route r is selected, the full-length route always exists so $y_1 = 1$. Therefore, α_k^i and $\beta_k^{(i,i+1)}$ can be determined as follows.

$$\alpha_k^i = \varphi_k \cdot y_1 \cdot \bar{\alpha}_1^i + (1-\varphi_k) \cdot \sum_{r=2}^{R} y_r \cdot \bar{\alpha}_r^i \tag{4}$$

$$\beta_k^{(i,i+1)} = \alpha_k^i \cdot \alpha_k^{i+1} \tag{5}$$

Where $\bar{\alpha}_r^i \in \{0,1\}$ denotes whether route r covers station i. Namely, $\bar{\alpha}_r^i = 1$ represents route r covers station i, otherwise $\bar{\alpha}_r^i = 0$. It can be obtained according to the line's topological structure, treated as predetermined input for our proposed model.

(3) Preceding train

In order to calculate the number of passengers arriving within departures of two consecutive trains, we introduce the notation of the nearest preceding train stops at station i before train k, denoted as $\bar{k}_i$.

$$\bar{k}_i = \max\{k' \mid k' < k, \beta_{k'}^{(i,i+1)} = \beta_k^{(i,i+1)} = 1\} \tag{6}$$

Taking Figure 3 as an example, the nearest preceding train stopping at station 1 before train k is train $\bar{k}_1 = k - m - 1$, while the nearest preceding train stopping at station a before train k is train $\bar{k}_a = k - 1$. Then the effective time period for arriving passengers at station i between two consecutive trains can be denoted as $\left(D_{\bar{k}_i}^i, D_k^i\right]$.

3.2. Passenger flow calculation

(1) Arriving passengers

The number of passengers who arrive at station i towards station j between the departures of train k and its nearest preceding train $\bar{k}_i$ can be calculated as below.

$$I_k^{ij} = \sum_{t \in \left(D_{\bar{k}_i}^i, D_k^i\right]} P_t^{ij} \tag{7}$$

(2) Waiting passengers

The number of waiting passengers at station i with destination station j between the departures of train k and its preceding train $\bar{k}_i$ is the sum of arriving passengers and the left-behind passengers by train $\bar{k}_i$.

$$W_k^{ij} = I_k^{ij} + R_{\bar{k}_i}^{ij} \tag{8}$$

According to assumption A1, whether passengers want to board train k depends on whether the coming train can reach their destinations directly. We introduce binary parameter γ_k^{ij} to indicate whether the OD pair from station i to station j can be served by a train k without transfer, its value can be obtained by $\gamma_k^{ij} = \alpha_k^i \cdot \alpha_k^j$. In other words, $\gamma_k^{ij} = 1$ represent train k stops at both station i and station j, otherwise $\gamma_k^{ij} = 0$. Then the number of potential passengers who want to board train k at station i towards station j can be calculated by:

$$WP_k^{ij} = W_k^{ij} \cdot \gamma_k^{ij} \tag{9}$$

(3) Boarding and alighting passengers

Due to the train capacity constraint, the number of passengers who actually board train k, B_k^i, equals the minimum of the total number of potential passengers who want to board train k at station i and the remaining carrying capacity of train k, can be expressed as follows:

$$B_k^i = \min\left\{\sum_{j=i+1}^{N} WP_k^{ij}, Cn\eta_{\max} - Q_k^{(i-1,i)} + O_k^i\right\} \tag{10}$$

Where $Cn\eta_{\max}$ represents the maximal carrying capacity of train k, which is the product of vehicle rated carrying capacity C, the number of cars in a train n and maximal load rate $\eta_{\max}$.

Meanwhile, B_k^i is also equal to the sum of the number of passengers boarding train k at station i with all destinations.

$$B_k^i = \sum_{j=i+1}^{n} B_k^{ij} \tag{11}$$

Assuming that the number of boarding passengers with destination station j is proportional to the number of potential passengers who want to board, the number of passengers boarding train k at station i towards station j can be calculated by the following equation.

$$B_k^{ij} = B_k^i \cdot \frac{WP_k^{ij}}{\sum_{j=i+1}^{n} WP_k^{ij}} \tag{12}$$

The number of passengers alighting from train k at station i can be calculated as follows:

$$O_k^i = \sum_{j=1}^{i-1} B_k^{ji} \tag{13}$$

(4) Left-behind Passengers

The number of left-behind passengers is related to the attributes and the surplus carrying capacity of the coming train. There are the following two cases:

1) If train k cannot serve the trip from station i to station j directly, namely $\gamma_k^{ij} = 0$.

In this case, the passengers waiting at station i with destination station j do not want to get on train k, they will choose to wait for the next direct train, then the number of left-behind passengers by train k is equal to the number of waiting passengers on the platform, as follows.

$$R_k^{ij} = W_k^{ij} \tag{14}$$

2) If train k can serve the trip from station i to station j without transfer, namely $\gamma_k^{ij} = 1$.

In this case, all passengers want to board train k. However, the passenger flow is very large during rush hours, some of passengers may cannot get on the train caused by train capacity restriction. The number of left-behind passengers is calculated by

$$R_k^{ij} = WP_k^{ij} - B_k^{ij} \tag{15}$$

For a unified description, all the passengers who do not take the first coming train are regarded as left-behind passengers, the number of passengers left behind by train k at station i towards station j can be expressed as follows.

$$R_k^{ij} = \gamma_k^{ij} \cdot (WP_k^{ij} - B_k^{ij}) + (1 - \gamma_k^{ij}) \cdot W_k^{ij} \tag{16}$$

(5) In-vehicle passengers

After the process of passengers boarding and alighting, the number of passengers remaining in train k after the train departs from station i can be calculated as below

$$Q_k^{(i,i+1)} = Q_k^{(i-1,i)} + B_k^i - O_k^i \tag{17}$$

The load rate of train k after the train departs from station i, $E_k^{(i,i+1)}$, can be quantified by the ratio of the number of in-vehicle passengers to the train rated carrying capacity, which not only reflects the utilization level of train capacity, and also characterizes the service level of train operation.

$$E_k^{(i,i+1)} = Q_k^{(i,i+1)} / Cn \tag{18}$$

4. Model formulation

4.1. Objective function

Based on the resulting parameters obtained by the passengers-trains interaction model in Section 3, we now formulate the train schedule optimization model. Taking into account the trade-off between operator and passenger, the objective considered in this paper is to minimize the users' cost and the operators' cost. Moreover, equilibrating train occupancy levels is also included in the objective function.

(1) Users' cost

Because short-turn route design will not affect passengers' access, in-vehicle and station dwell costs (Sun et al. 2016), we only account for waiting costs, which are sensitive to the planned train route patterns and departure times of trains. Assuming that passengers arrive at the start time stamp of the time intervals, the initial waiting time for passengers who arrive at station i between departures of train $\bar{k}_i$ and train k with the destination station j is $\sum_{t\in\left(D^i_{\bar{k}_i},D^i_k\right]} P^{ij}_t\cdot(D^i_k-t)$, while the remaining passengers at station i towards station j after train $\bar{k}_i$ leaves need to wait for additional service headway $R^{ij}_{\bar{k}_i}\cdot\left(D^i_k-D^i_{\bar{k}_i}\right)$ as a result of previous boarding failure. The total waiting cost U_1 is the sum of initial waiting time for arriving passengers and extra waiting time for the left-behind passengers, represented as below.

$$U_1=\lambda\cdot\left[\sum_{k=1}^{K}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sum_{t\in\left(D^i_{\bar{k}_i},D^i_k\right]}\alpha^i_k\cdot P^{ij}_t\cdot(D^i_k-t)+\varepsilon\cdot\sum_{k=2}^{K}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\alpha^i_k\cdot R^{ij}_{\bar{k}_i}\cdot(D^i_k-D^i_{\bar{k}_i})\right] \tag{19}$$

Where λ is time value, and ε is penalty coefficient for the left-behind passengers, since they feel more dissatisfied than those who board the first coming train.

(2) Operators' cost

Operators' cost is divided into train running cost and personnel cost. Train running cost includes fuel consumption, train maintenance and so on, is measured by the total vehicle-kilometer $\sum_r f_r nl_r y_r$. Personnel cost refers to the crew salary, is related to the total train travel time $\sum_r f_r t_r y_r$. So the total operators' cost U_2 is expressed as follows.

$$U_2=\theta_1\cdot\sum_r f_r nl_r y_r+\theta_2\cdot\sum_r f_r t_r y_r \tag{20}$$

Where θ_1 is unit operating cost per vehicle per kilometer, θ_2 is unit operating cost of a train per minute.

(3) Train loads balance

In order to achieve the best match between transport capacity supply and traffic demand in time and space dimensions, equilibrating train occupancy levels over all the segments is also considered as an additional objective. During peak hours, [50%,100%] is the reasonable range of train load rate. When the number of in-vehicle passengers exceeds the train rated carrying capacity $Cn\times100\%$, the train will become overcrowded, reducing the level of service. In addition, if the number of in-vehicle passengers is less than 50% of the train rated carrying capacity, it means that transport capacity is surplus, resulting in loss of enterprise benefits. By limiting the train load rate within the reasonable range, we add two kinds of penalty costs, UP_1 and UP_2, to minimize passenger overload and capacity waste, respectively.

$$UP_1 = \delta_1 \cdot \sum_{k=1}^{K}\sum_{i=1}^{N-1} \max\left\{0, Q_k^{(i,i+1)} - Cn\times100\%\right\} \cdot \beta_k^{(i,i+1)} \tag{21}$$

$$UP_2 = \delta_2 \cdot \sum_{k=1}^{K}\sum_{i=1}^{N-1} \max\left\{0, Cn\times50\% - Q_k^{(i,i+1)}\right\} \cdot \beta_k^{(i,i+1)} \tag{22}$$

Where δ_1 is the unit penalty cost for overload, δ_2 is the unit penalty cost for wasted capacity.

To sum up, the objective function of this model is expressed as the sum of users' cost, operators' cost, overload penalty cost and capacity waste penalty cost.

$$\min\ U = U_1 + U_2 + UP_1 + UP_2 \tag{23}$$

4.2. Constraints

(1) Headway constraint

During the process of operation, the departure times of two adjacent trains at any station must meet the headway requirements. On the one hand, due to the limitations of technical conditions, the headway between two consecutive trains should be greater than the minimum tracking interval $I_{\min}$, which is determined mainly by the train control and signaling system. On the other hand, headway will affect the passengers' waiting time, which should be less than the maximum service interval $I_{\max}$.

$$I_{\min} \le D_k^i - D_{k_i}^i \le I_{\max} \quad \forall i,k \tag{24}$$

(2) Departure time constraint

In order to ensure that the train departures spread over the planning period, the first and last trains are scheduled during the start interval and the end interval of planning horizon, respectively.

$$0 \le D_1^1 \le I_{\max} \tag{25}$$

$$T - I_{\max} \le D_K^1 \le T \tag{26}$$

(3) Route pattern constraint

Considering the feasibility and convenience of transport organization, two types of route patterns are allowed in this model. The corresponding constraints are as follows.

$$y_r \in \{0,1\} \tag{27}$$

$$y_1 = 1 \tag{28}$$

$$\sum_{r=2}^{R} y_r = 1 \tag{29}$$

Equation (28) ensures the existence of full-length route pattern, and Equation (29) ensures there is only one short-turn route pattern.

(4) Number of departure trains constraint

Based on the actual operation requirements of urban rail transit, the ratio of the number of short-turn trains to the number of full-length trains is usually 1:1, 2:1 or 1:2. Accordingly, the number of departure trains for each route pattern can be determined as follows.

$$m \in \{1, 2, 1/2\} \tag{30}$$

$$f_r = m \cdot f_1 \cdot y_r \qquad \forall r > 1 \tag{31}$$

From the above, the proposed optimization model that combine short turning and heterogeneous headways as an integrated strategy (HH-ST) is formulated as follows:

Objective function (23)

Subject to constraints (1)-(18), (24)-(31)

5. Solution approach

This train service optimization problem for a congested urban rail line is a non-linear non-convex mixed integer programming model. For practical applications, it contains a number of decision variables and constraints, and is extremely difficult to solve mathematically. In order to obtain high quality solutions within an acceptable time range, meta-heuristics are commonly

used to solve this kind of problem, such as Genetic Algorithm (GA) (J. Wu et al. 2015; Kang et al. 2015; Shafahi and Khani 2010), Simulated Annealing (SA) algorithm (Kang and Zhu 2015), Tabu Search (TS) algorithm (Parbo, Nielsen and Prato 2014), Particle Swarm (PS) algorithm (Meng, Jia and Qin 2010). Among them, GA is a widely used stochastic optimization procedure, which is applicable to proposed model for the following reasons：① The decision variables in this model are discrete integers, which facilitates the coding process of chromosome. ② In the crossover and mutation operations of GA, the acquired offspring can be easily guaranteed to be effective and available by constraining the value range of the decision variables. ③ Since the objective function calculation first needs to deal with the interaction process between passengers and trains, it is convenient and quick to calculate the fitness value of each generation by GA, that is, the equations (1)-(18) can be transformed into the calculation process of fitness function without directly solving the constraints.

5.1. Algorithm framework

Considering the characteristics of HH-ST model, the variables to be optimized are divided into train departure times and train route related variables, so a two-stage solution approach is proposed to solve this problem. The algorithm flowchart is shown in Figure 4, and the structure of the proposed algorithm is described as follows.

Stage I is used to optimize the route scheme, which is determined by selecting one of the feasible combinations of turn-back stations and the ratio of the number of short-turn trains to the number of full-length trains. Because the candidate route set and the ratio of different route patterns are predetermined and limited, an exhaustive search method is used to find all the combinations of the route schemes, then the real-size problem can be broken into a series of sub-problems, which as an input for the Stage II.

For each specific route scheme, Stage II solves a non-linear nonconvex problem by adopting GA to search for the optimal departure times of all trains for each corresponding sub-problem. During the implementation of GA, invoke the interaction process to assign all passengers to all trains, and the resulting parameters are used to calculate the fitness value.

As a result, the best solution to HH-ST model is given by the minimum value of the objective function over the set of sub-problems generated.

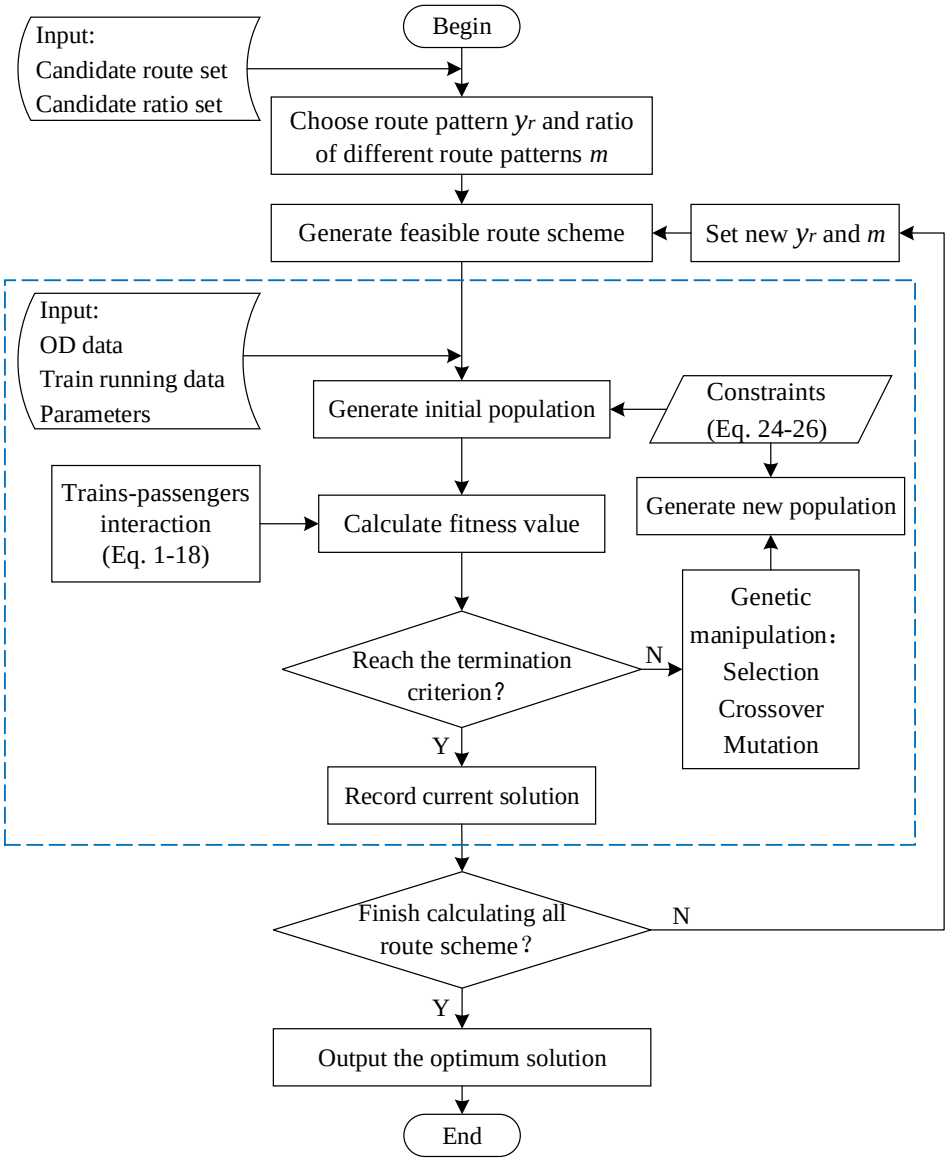


Figure 4. Algorithm flowchart

5.2. Design of GA

In the above algorithm, the specific design of GA will be explained as follows. To deal with constraints conveniently and enhance the efficiency of computing, an improved integer coding approach is introduced. As shown in Figure 5, a chromosome with length $K+1$ is designed for this GA, the value associated with the *k*th gene is denoted as $x_k(k=0,1,\ldots,K)$, x_0 indicates the interval between the first train and the start time of planning period, $x_k(k=1,2,\ldots K-1)$ corresponds to the headway between train k and train $k+1$ at start terminal, and x_K indicates the interval between the last train and the end of study period. For example, the first gene "2" means that the first train departs from the start terminal at 2 minutes, and the second gene "3" means that the headway between the first and second trains is 3 minutes.

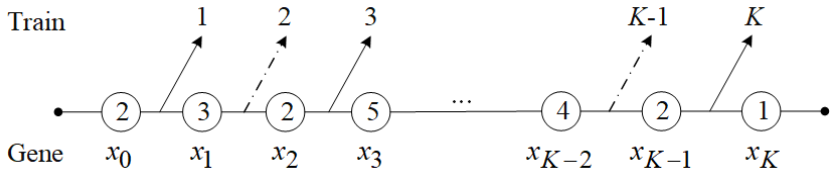


Figure 5. Integer coding illustration of a chromosome

According to the constraints (25) and (26), the first gene x_0 and the last gene x_K are defined as integer variables in the range of $[0, I_{\max}]$, and the values of other genes are restricted to the range of $[I_{\min}, I_{\max}]$ based on constraint (24). For ensuring the departure times of all trains during the study period, the variables should also satisfy the following constraint.

$$\sum_{k=0}^{K} x_k = T \tag{32}$$

By using this integer coding approach, the corresponding departure times D_k^1 of train k at the start terminal can be decoded recursively as follows:

$$D_1^1 = x_0 \tag{33}$$

$$D_k^1 = D_{k-1}^1 + x_{k-1} \quad \forall k > 1 \tag{34}$$

On this basis, the arrival and departure time of each train at each station can be obtained successively according to equations (1)-(2). Correspondingly, all other intermediate variables can be calculated recursively form the first train beginning form the start terminal step-by-step by using equations (7)-(18), then the value of objective function can also be determined.

The fitness value is the criterion used to reflect the merits and demerits of the individual, which is the driving force of the algorithm evolution process. In this paper, the fitness function is represented as below.

$$Fitness = U_{\max} - U \tag{35}$$

Where U is the objective value from Equation (23), $U_{\max}$ is the maximum value of U associated with the current generation.

The genetic manipulation of GA are similar to the other integer coding GA (Y. Wu et al. 2015; Kang et al. 2015; Shafahi and Khani 2010). Selection process adopts both the roulette wheel and elitism regulation. Two-point crossover method is used in crossover operation, the values of the genes between the two selected points are replaced by the values of the same genes in another chromosome. For the mutation operation, a sequence of binary variables is generated randomly to determine the mutation locations, and random integers that satisfy constraints are

generated to replace the original ones, creating a new individual. Finally, the termination criterion of the GA in this research is defined as function tolerance. If the best fitness function does not change after a given number of iterations, then the algorithm terminates.

6. Case study

6.1. Inputs

In this section, a real case of Metro Line 9 in Shanghai is conducted to test the performance of the proposed model and solution. Line 9 is a major trunk line from east to west that connects distant residential area with downtown area, which consists of 26 stations. As shown in Figure 6, the dwell times at every station and the train running times of each section between every two adjacent stations in the considered direction are given. According to the setting conditions of reversing tracks, taking into account the turn-back capacity, route length and other factors, four candidate train routes used in this study are also shown in Figure 6, where each route is represented by a pair of turn-back stations.

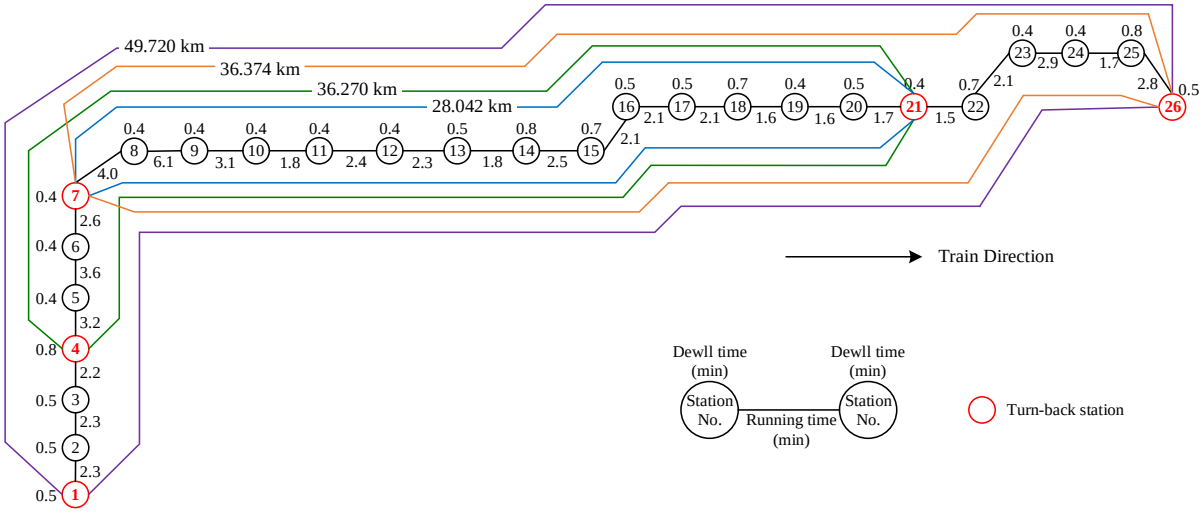


Figure 6. Shanghai Metro Line 9

With a particular focus on a heavily congested urban rail line, the morning rush hours from 7:30am to 8:30am on March 3, 2016 is considered as the planning period for the starting terminal. The input OD passenger demands are recorded as groups at one-minute intervals based on the transaction data collected from Automatic Fare Clearing system. Due to the large amount of data, we present time-dependent demand graphically in the form of section passenger volume per minute, as shown in Figure 7, where the section names are substituted by serial numbers. We can see that the passenger demand curve varies over space and time.

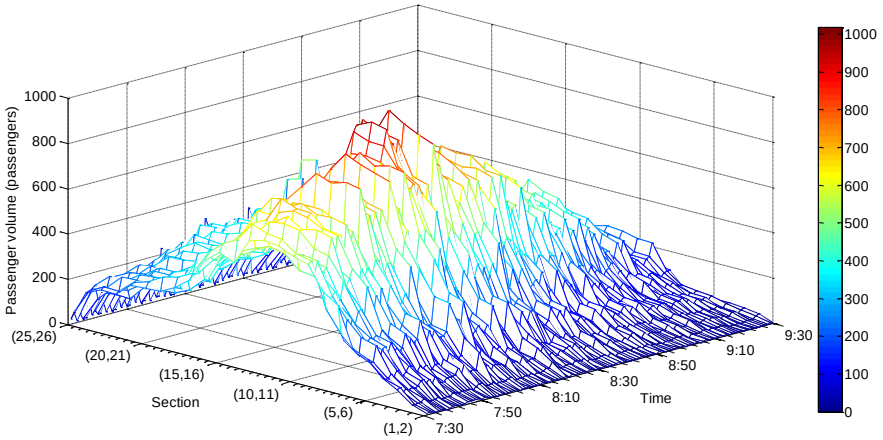


Figure 7. Time-dependent passenger demands

For the large passenger flow direction of Line 9, the number of train services K dispatched during the planning horizon is taken as 20. The number of train marshalling n is 6 vehicles, the vehicle rated carrying capacity C is 310 passengers/vehicle, and the maximal load rate $\eta_{\max}$ is 130%. The minimum and maximum headway are set as 2 minutes and 8 minutes, respectively. The penalization for denied boarding ε is set to 2, that is, the extra waiting time is considered to be 2 times more expensive than initial waiting time. When determining the values of the coefficients in objective function, taking the regular schedule (homogeneous headways and full-length route pattern) as the benchmark, calculate the target values at this time, including initial waiting time, extra waiting time, total vehicle-kilometer, total train travel time, number of overloaded passengers and wasted capacity. Assuming that the importance of each objective are the same in the baseline case, set the value of initial waiting time λ as =1 \$/min, then the values of other coefficients can be obtained: θ_1=15 \$/vehicle/km, θ_2=50 \$/min, δ_1=2 \$/person, δ_2=1 \$/person. In addition, other binary parameters such as $\bar{\alpha}_r^i$, can be determined through the line's topological structure.

6.2. Results and analysis

In order to better evaluate the performance of the integrated strategy combining heterogeneous headways and short turning (HH-ST), we have calculated the train schedules by three additional situations, separate heterogeneous headways strategy (HH), separate short-turning strategy (ST) and regular schedule (No strategy), based on the same input data and a fixed number of train services. All computations are performed on a personal computer with an Intel Core i5-6200U 2.4 GHz and 8 GB RAM, and the mathematical model was programmed and solved with MATLAB R2014a. The computing time required for solving the four models is 8.5 min (HH-

ST), 2.4 min (ST), 6.8 min (HH), 1.6 min (No strategy), respectively. It can be seen that the proposed algorithm can solve each model within an acceptable time range. Due to the long chromosome length in the HH model and HH-ST model, the corresponding calculation time is longer.

6.2.1. Optimal train schedules

To explore the variation of optimal schedules from these strategies, we plot the train service headway profiles in Figure 8, where the first point represents the departure time of the first train in the planning period, and other points means the headways between every two consecutive trains. As we can see, the regular schedule (Figure 8(a)) and the result of ST strategy (Figure 8(c)) have the fixed headway of 3 minutes, while the headways obtained by HH strategy (Figure 8(b)) and HH-ST strategy (Figure 8(d)) vary with time, showing more consistence with time-dependent demand. In Figure 8(c) and 8(d), the hollow point means the offset time between a short-turn service and its preceding full-length service and the solid point means the offset time between a short-turn service and its subsequent full-length service, which indicates that the optimal ratio of two route pattern trains is 1:1. In addition, in HH-ST strategy, we also find that the headway between a short-turn trip and its preceding full-length train is no less than the headway with its subsequent full-length trip. This helps to coordinate the loads for different trains, because the full-length train is already partially loaded when it arrives at the turn-back station, while the short-turn train will start there empty, the more offset time from the preceding full-length train can lead short-turn train captures most of the passengers. This result is consistent with Furth (1987).

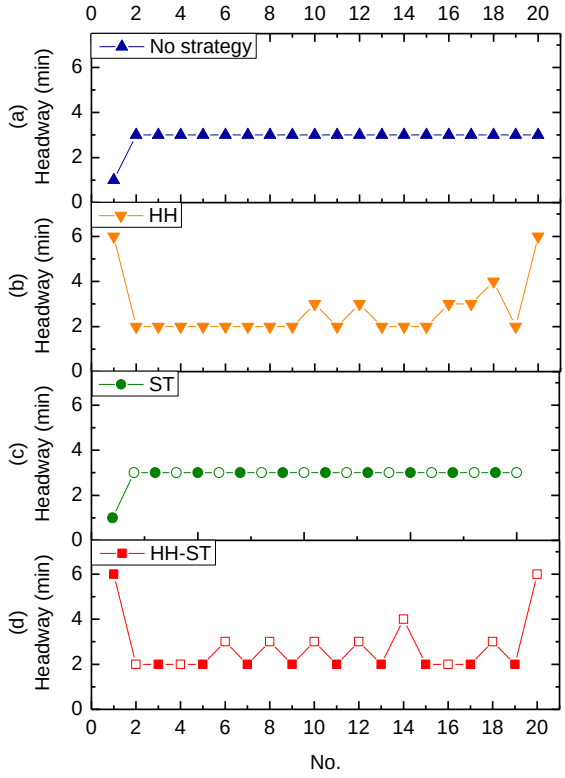


Figure 8. Headway profiles for different strategies

6.2.2. Train Load rates

The corresponding train diagram of the optimized schedules are shown in Figure 9, where each line represents the running trajectory of each train, and the train running trajectories are given different colors to intuitively present the load rates for each train at each section (Li et al. 2017).

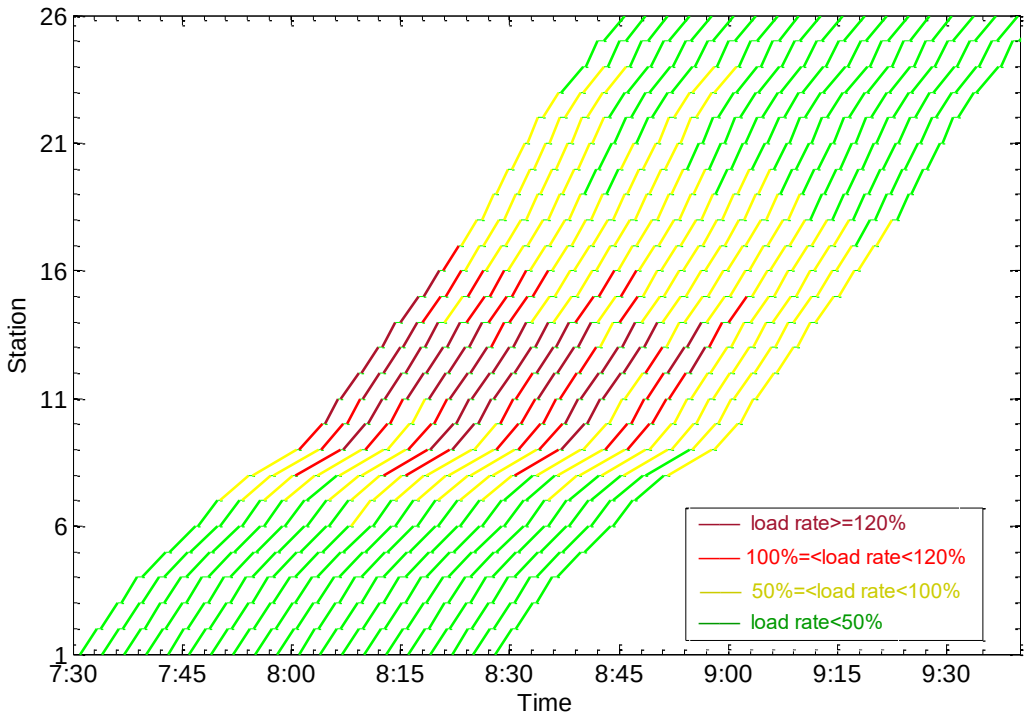
In the regular schedule given in Figure 9(a), the crowded areas are concentrated from Station 8 to Station 17 during 8:00 to 9:00, especially the three sections between Station 11 to Station 14, where colors are almost all crimson, indicating that the trains are extremely congested. However, the train load rates are less than 50% from Station 1 to Station 7 and from Station 24 to Station 26, revealing that the train carrying capacity is partially redundant in suburban areas.

In the schedule optimized by HH strategy given in Figure 9(b), the headways between trains are consistent with the time-dependent demand patterns. Compared to the regular schedule, the number of overloaded segments are significantly reduced. By statistics, there are 44 red segments and 49 crimson segments in Figure 9(a), while the number of segments with load rates greater than 100% and 120% in Figure 9(b) are decreased to 28 and 22, respectively.

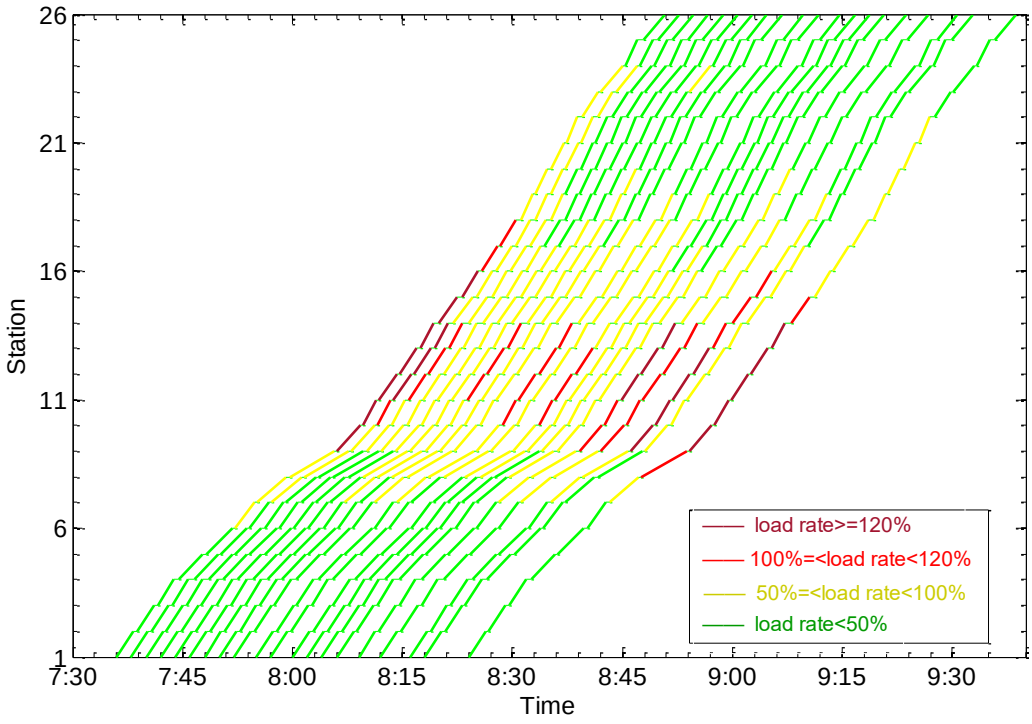
As can be seen in Figure 9(c), the optimal short-turn route pattern is between station 7 and station 26. The spatial distributions and sequence structures of trains are synchronized with the

spatial imbalance of passenger demand distribution. The high-density train distribution inside the short-turn route coincides with the high-demand sections in Figure 7. During the non-common zones, the train service frequency is 1/2 of the regular schedule, helping to reduce the number of segments with low load rate from 241 to 178 in comparison to regular schedule.

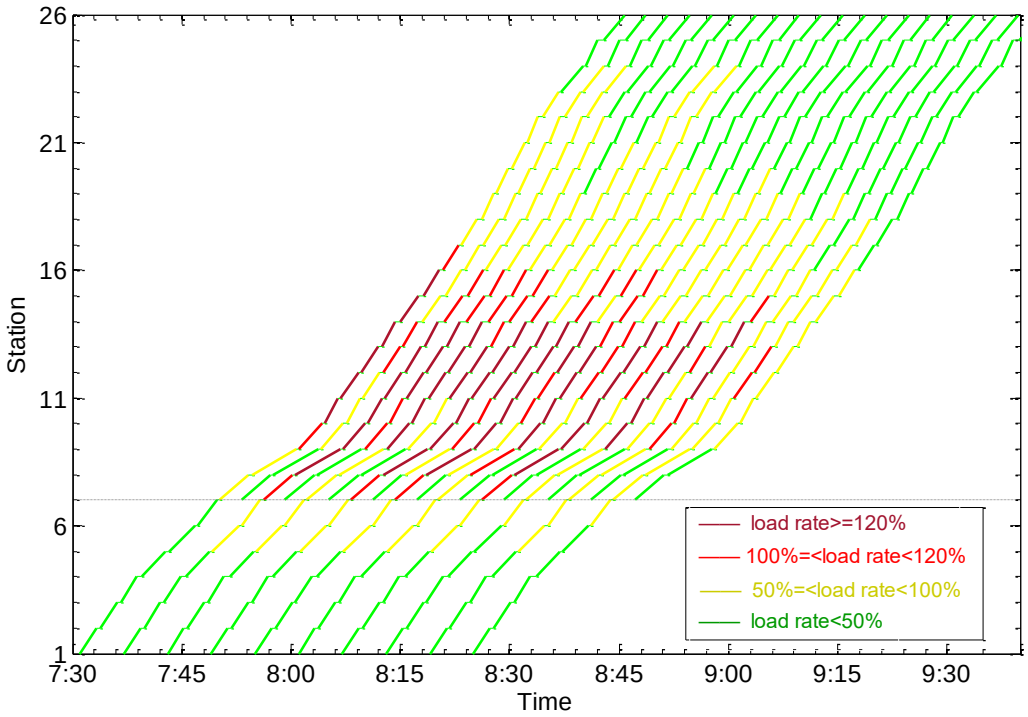
In Figure 9(d), the train schedule obtained by HH-ST shows the best match with passenger demand both in time and space dimensions. This demand-oriented train service plan integrates the advantages of short turning strategy and heterogeneous headways strategy, not only reduce the overloaded segments, but also reduce the segments with excess capacity. After apply the proposed HH-ST model, the train loads become more uniform, contributing to improve the level of service and transport capacity utilization.



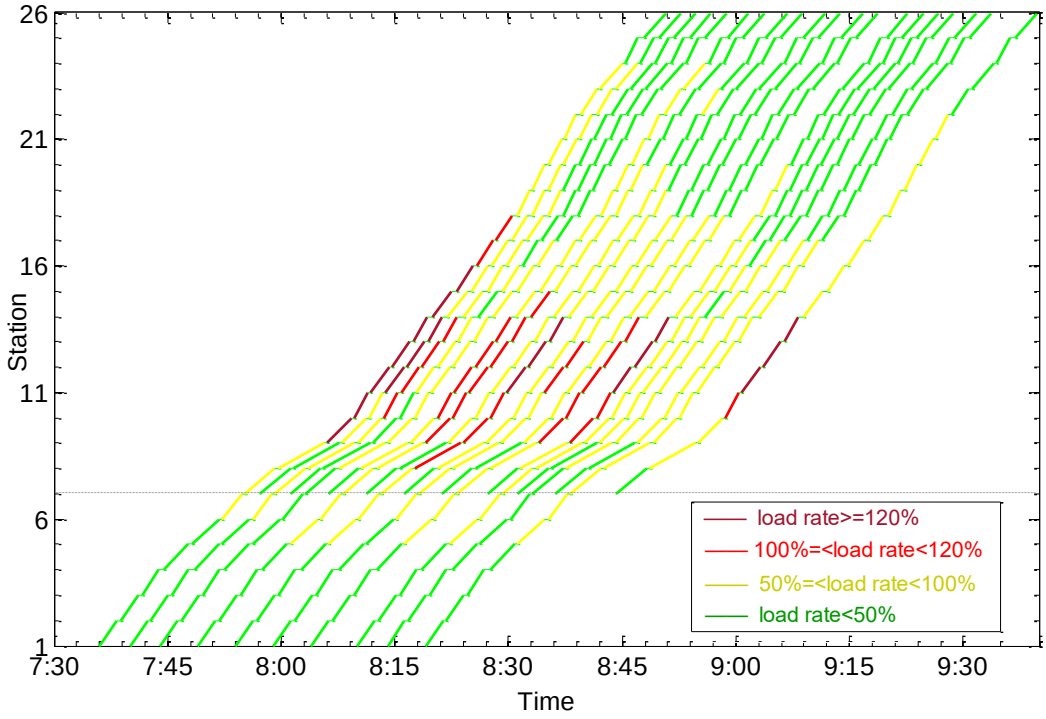
(a) Train diagram with no strategy



(b) Train diagram obtained by HH



(c) Train diagram obtained by ST



(d) Train diagram obtained by HH-ST

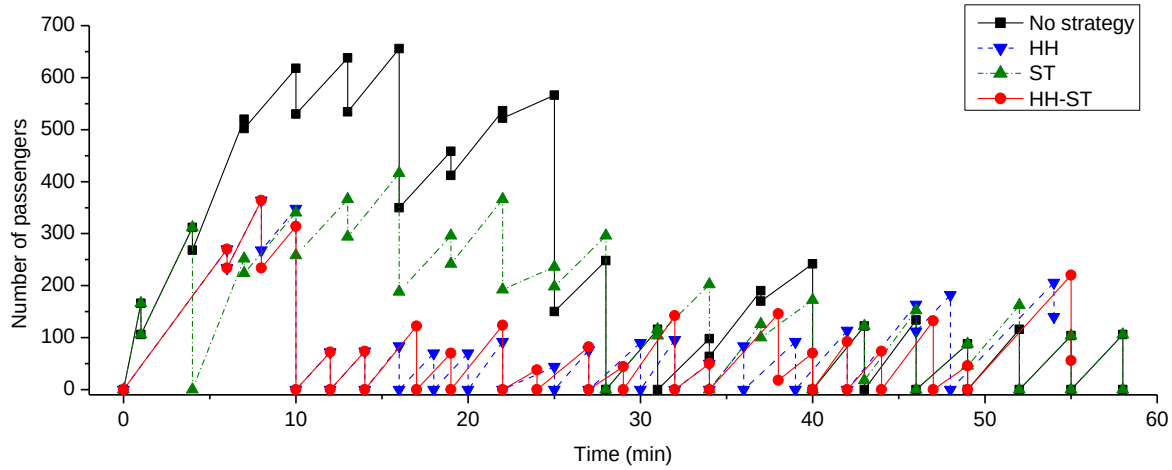
Figure 9. Train diagrams with loading profile for different strategies

6.2.3. Gathering passengers on the platform

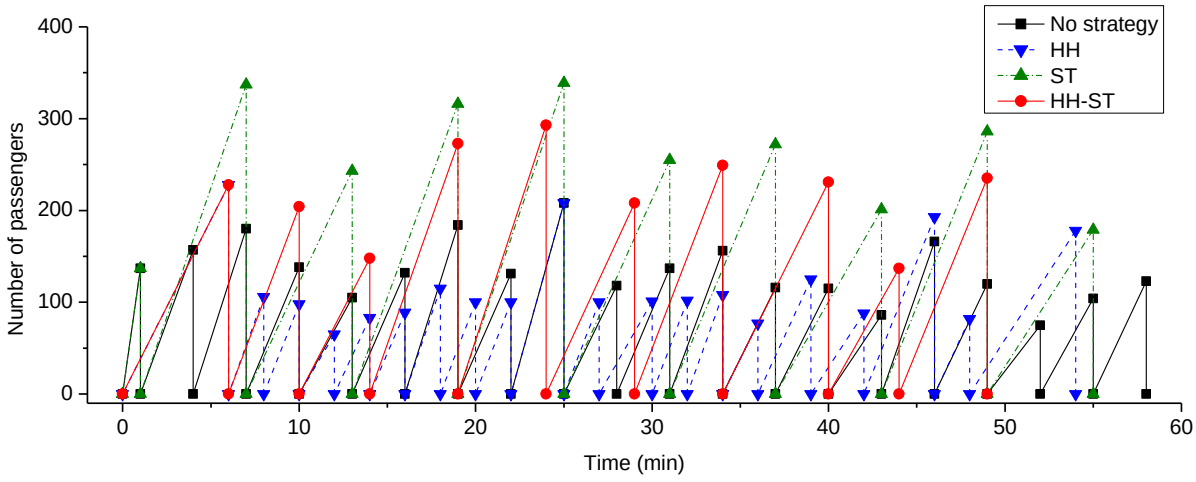
To further analyze the change of gathering passengers on the platform over time after applying different strategies, we select two typical stations, Station 12 inside the short cycle and Station 3 outside the short cycle, as research objects. Assuming that the process of passengers leaving the station can be ignored, the corresponding change of passenger flow on the platform is shown in Figure 10.

It can be observed from Figure 10(a), the regular schedule (no strategy) is the worst because that the sustained large number of passengers are left-behind by the first coming train in the first 30 minutes at Station 12, and the maximum number of gathering passengers reaches 656 people, causing serious safety risks to station operation. After implementing the ST strategy, although the largest gathering passengers drop to 416 people, there are still continuing stranded passengers. The effects of HH-ST strategy and HH strategy are similar, the number of left-behind passengers has been significantly reduced and the number of gathering passengers remains below 200 people after 10 minutes, indicating that the time-varying train schedules can mitigate the phenomenon that passengers cannot board in time due to limited service capacity, and effectively improve the waiting environment of the station platform. However, as can be seen from Figure 10(b), the number of gathering passengers on the platform of Station 3

increases after applying ST strategy and HH-ST strategy, despite the growth range of HH-ST is smaller than that of ST. Due to the adoption of short-tuning strategy, only full-length trains pass through the non-common sections, which makes longer service headway, and leads to an increase in the number of gathering passengers at station outside the short cycle, but there is no left-behind passengers because less traffic volume and adequate train capacity. This explains that the short-turning strategy alleviates the passenger flow pressure in the high-demand sections by sacrificing the waiting time of passengers in the low-demand sections, contributing to adapt the spatial distribution of passenger demand.



(a) Station 12 (Inside the short cycle)



(b) Station 3 (Outside the short cycle)

Figure 10. Number of gathering passengers on the platform

6.2.4. Comparison of performance indicators

A comparison of the performance of the four strategies is illustrated in Table 2, where the values of the total initial waiting time, total left-behind waiting time, total users' cost, total operators'

cost, penalty cost for overload, penalty cost for wasted capacity and the total cost are listed. Additionally, the benefits of the solutions obtained by optimization strategies are presented via the relative deviations of the optimal results and the base results.

For the HH strategy, the total initial waiting time is reduced by 19.28%, and the total extra waiting time is even reduced by 66.62%. This clearly demonstrates the effectiveness of HH strategy in terms of reducing passengers' waiting time, especially mitigate the left-behind phenomenon. The penalty for overload is reduced by 51.59%, indicating that the schedule with heterogeneous headways can effectively alleviate congestion in the cabin. Although the penalty for wasted capacity increased with 17.47%, this loss can easily be compensated by other benefits. As a result, the total cost is 10.07% smaller than that of the base schedule. From a theoretical point of view, the non-fixed headway schedule is more flexible than the fixed headway schedule, can better adapt to the time-varying passenger demand by adjusting service headways, and contribute to improve the level of service.

For the ST strategy, the operators' cost decreases by 13.12%, and the plenty for wasted capacity is reduced by 34.56%. It is observed that the ST strategy has a significant advantage in saving operator cost, reducing the waste of transport capacity on low-demand sections, and improving the capacity utilization. However, the total users' cost obtained by ST is 23.04% higher than that of the base schedule. One reason is that the longer train service interval outside the short cycle results in an increase in the initial waiting time for suburban passengers. Another reason is that equal offset between full-length trip and short-turn trip lead each route pattern carries one-half of the passengers within the short cycle, then it will aggravate traffic pressure on the full-length trains, causing an increase of left-behind passengers. Nevertheless, the benefit of operating cost savings is sufficient to cover the increase of passenger wait costs, the total cost is reduced by 5.10%.

Compared with the two separate strategies, the solution obtained by HH-ST strategy minimizes the total cost by 23.17%, yields the best performance for this case study. All indicators have been reduced with varying degrees. For example, the optimized total users' cost is 21.14% less than that of the base schedule, because the waiting time savings by heterogeneous headways strategy outweigh the waiting time increases by short-turning strategy. Besides, 13.12% of the operating costs are saved due to application of short-turning strategy. Furthermore, both congestion penalty and waste penalty are reduced, which represents that the load rates of trains are optimized to the reasonable range. Therefore, for a congested urban rail line, a demand-oriented train service plan with HH-ST strategy can effectively coordinate train

services with complex spatial and temporal demand distributions, achieve the optimal matching between transport capacity and traffic demand, and benefit both the passengers and the train operators.

Table 2. Performance comparison for different strategies

Indicator	HH		ST		HH-ST		No strategy (Base schedule)
	Optimal Results	Variations	Optimal Results	Variations	Optimal Results	Variations	
Total initial waiting time (min)	86092	-19.28%	124710	16.93%	98504	-7.64%	106650
Total left-behind waiting time (min)	3842	-66.62%	17415	51.29%	1880	-83.67%	23022
Total users' cost ($)	93776	-27.68%	159540	23.04%	102260	-21.14%	129670
Total operators' cost ($)	165096	0.00%	143430	-13.12%	143430	-13.12%	165096
Penalty cost for overload ($)	31280	-51.59%	73420	13.63%	28736	-55.53%	64614
Penalty cost for wasted capacity ($)	140988	17.47%	78536	-34.56%	93893	-21.77%	120017
Total cost ($)	431140	-10.07%	454930	-5.10%	368330	-23.17%	479400

6.3. Sensitivity analysis

In order to explore the trade-off between passenger benefit and corporate benefit, the coefficients in objective function are divided into two groups, a group of time value λ and overload penalty coefficient δ_1, a group of operating cost coefficients θ_1, θ_2 and capacity waste penalty coefficient δ_2. We doubled and quadrupled the values of the two sets of coefficients, respectively, the results of the HH-ST model are shown in Table 3. When λ and δ_1 increases, passengers' goals are given higher priority, so the passenger waiting time and overloaded passengers decrease. In the third instance, the best ratio of short-turn and full-length trains becomes 1:2, which means a higher number of full-length trains than short-turn trains is more beneficial to passengers. In the fourth and fifth instances, when θ_1, θ_2 and δ_2 become larger, the total train running mileage, train travel time and wasted capacity decrease, so we can

conclude that more short-turn trains and shorter length of the short-turn route are more helpful in reducing operator cost.

Table 3. Effects of coefficients values on the results

No.	$\lambda/\delta_1/\theta_1/\theta_2/\delta_2$	Waiting time (min)	Train running mileage (vehicle·km)	Train travel time (min)	Overloaded passengers (people)	Wasted capacity (people)	Route Ratio	Short -turn
1	1/2/15/50/1	102260	5166	1319	14368	93893	1:1	7-26
2	2/4/15/50/1	99519	5166	1319	13184	93733	1:1	7-26
3	4/8/15/50/1	93919	5486	1396	11513	114587	1:2	7-26
4	1/2/30/100/2	135140	4666	1181	18684	73455	1:1	7-21
5	1/2/60/200/4	243080	4275	1082	32574	70037	2:1	7-21

7. Conclusions

In this paper, we studied the demand-oriented train scheduling problem for a congested urban rail line under temporal-spatial heterogeneous passenger demand conditions, combing heterogeneous headways and short-turning strategy as an integrated train management strategy (HH-ST). Taking into account the train capacity constraint and passenger selection behavior, a passenger-train interaction algorithm is developed to describe the operation of the trains and calculation of the number of waiting passengers at platforms, left-behind passengers and in-vehicle passengers. Based on these intermediate variables, a mixed integer nonlinear programming model is formulated to jointly benefit the operators and the passengers, and balancing vehicle occupancy levels over all the segments. The optimization variables are the departure times of each train, the stations where the short-turn strategy begins and ends, and the ratio between short-turn trains and full-length trains. Due to the complexity of this model, a two-stage heuristic algorithm based on an integer coding approach has been proposed to decompose the primal problem into a set of sub-problems and thus enables to find a good solution in short computational time.

To verify the effectiveness of the proposed train scheduling approach, we have calculated four train schedules for Shanghai Metro Line 9 with different strategies, which include integrated HH-ST strategy, separate HH strategy, separate ST strategy and regular schedule (no strategy). The computational results showed that, a demand-oriented train service plan obtained by HH-ST method yields the best performance compared with other strategies. Firstly, the train schedule obtained by the proposed HH-ST model makes the train loads more uniform, showing the best match between transport capacity and passenger demand both in time and space

dimensions. Secondly, the HH-ST method can effectively mitigate the left-behind phenomenon, and coordinate the gathering passengers at platform between oversaturated stations and unsaturated stations, contributing to improve the waiting environment and guarantee station safety. In addition, HH-ST strategy provides a better trade-off between users' cost and operators' cost than other strategies, thus evidently reducing the total passenger waiting time and keeping operation costs relatively low.

Although we consider the dynamic characteristics of passenger flow in this paper, in practice the variations of demand can be more complex, passenger flows fluctuate day by day. Thus, analyzing OD demand data of more days, and extending the proposed model to a stochastic optimization model to improve the robustness of the train schedule will be further studied. Besides, we will extend the proposed model by relaxing the assumption that the full-length route must be operated. In the absence of full-length trains, several short routes jointly serve the whole line, then transfers might be needed for some passengers. So a more general model to permit better route flexibility is needed and the key is to formulate a unified equation for calculating the total passenger waiting time, which includes initial waiting time, extra waiting time due to denied boarding and transfer waiting time. Moreover, our future research will extend our work to an urban rail transit network, where the transfer coordination between connected lines, and the influence of train operation plan on passenger route choice should be considered. Correspondingly, we will investigate other solution approaches to solve a large-scale problem efficiently.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No. 71701152, 71601110), the National Key Research and Development Plan of China (Grant No. 2017YFC0804900) and the Shanghai Committee of Science and Technology (Grant No. 19030501400).

References

Barrena, E., D. Canca, L. C. Coelho, and G. Laporte. 2014a. "Single-line rail rapid transit timetabling under dynamic passenger demand." *Transportation Research Part B: Methodological* 70: 134-150.

Barrena, E., D. Canca, L. C. Coelho, and G. Laporte. 2014b. "Exact formulations and algorithm for the train timetabling problem with dynamic demand." *Computers & Operations Research* 44: 66-74.

Canca, D., E. Barrena, A. Zarzo, F. Ortega, and E. Algaba. 2012. "Optimal Train Reallocation Strategies under Service Disruptions." *Procedia - Social and Behavioral Sciences* 54: 402-413.

Canca, D., E. Barrena, G. Laporte, and F. A. Ortega. 2016. "A short-turning policy for the management of demand disruptions in rapid transit systems." *Annals of Operations Research* 246 (1-2): 145-166.

Ceder, A. 1989. "Optimal design of transit short-turn trips." *Transportation Research Record: Journal of the Transportation Research Board* 1221: 8-22.

Chang, Y., C. Yeh, and C. Shen. 2000. "A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line." *Transportation Research Part B: Methodological* 34 (2): 91-106.

Cortés, C. E., S. Jara-Díaz, and A. Tirachini. 2011. "Integrating short turning and deadheading in the optimization of transit services." *Transportation Research Part A: Policy and Practice* 45 (5): 419-434.

Delle Site, P., and F. Filippi. 1998. "Service optimization for bus corridors with short-turn strategies and variable vehicle size." *Transportation Research Part A: Policy and Practice* 32 (1): 19-38.

Furth, Peter G. 1987. "Short turning on transit routes." *Transportation Research Record: Journal of the Transportation Research Board* 1108: 42-52.

Ghoneim, N. S. A., and S. C. Wirasinghe. 1986. "Optimum Zone Structure during Peak Periods for Existing Urban Rail Lines." *Transportation Research Part B: Methodological* 20 (1): 7-18.

Kang, L., J. Wu, H. Sun, X. Zhu, and Z. Gao. 2015. "A case study on the coordination of last trains for the Beijing subway network." *Transportation Research Part B: Methodological* 72: 112-127.

Kang, L., and X. Zhu. 2015. "A simulated annealing algorithm for first train transfer problem in urban railway networks." *Applied Mathematical Modelling* 40 (1):419-435.

Kroon, L., D. Huisman, E. Abbink, P. J. Fioole, M. Fischetti, G. Maróti, A. Schrijver, A. Steenbeek, and R. Ybema. 2009. "The New Dutch Timetable: The OR Revolution." *Interfaces* 39 (1):6-17.

Li, S., R. Xu, Z. Jiang, and W. Zhu. 2017. "Evaluating Matching Degree between Passenger Demand and Transportation Capability of Train Timetable for Urban Rail Transit." Transportation Research Board 96th Annual Meeting, Washington, D.C., January.

Liebchen, C. 2008. "The First Optimized Railway Timetable in Practice." *Transportation Science* 42 (4):420-435.

Meng, X., L. Jia, and Y. Qin. 2010. "Train Timetable Optimizing and Rescheduling Based on Improved Particle Swarm Algorithm." *Transportation Research Record: Journal of the Transportation Research Board* 2197: 71-79.

Niu, H., X. Tian, and X. Zhou. 2015. "Demand-Driven Train Schedule Synchronization for High-Speed Rail Lines." *IEEE Transactions on Intelligent Transportation Systems* 16 (5): 2642-2652.

Niu, H., and X. Zhou. 2013. "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions." *Transportation Research Part C: Emerging Technologies* 36: 212-230.

Niu, H., X. Zhou, and R. Gao. 2015. "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* 76: 117-135.

Parbo, J., O. A. Nielsen, and C. G. Prato. 2014. "User perspectives in public transport timetable optimisation." *Transportation Research Part C: Emerging Technologies* 48: 269-284.

Serafini, P., and W. Ukovich. 1989. "A Mathematical Model for Periodic Scheduling Problems." *SIAM Journal on Discrete Mathematics* 2 (4): 550-581.

Shafahi, Y., and A. Khani. 2010. "A practical model for transfer optimization in a transit network: Model formulations and solutions." *Transportation Research Part A: Policy and Practice* 44 (6): 377-389.

Sparing, D., and R. M. P. Goverde. 2017. "A cycle time optimization model for generating stable periodic railway timetables." *Transportation Research Part B: Methodological* 98: 198-223.

Sun, L., J. Jin, D. Lee, K. W. Axhausen, and A. Erath. 2014. "Demand-driven timetable design for metro services." *Transportation Research Part C: Emerging Technologies* 46: 284-299.

Sun, Y., P. M. Schonfeld, Y. Lu, and M. Zhou. 2016. "Redesigning Rail Transit Short Turn Operations: A Case Study of Line 2 of Shanghai Metro." *Transportation Research Record: Journal of the Transportation Research Board* 2540: 46-55.

Tirachini, A., C. E. Cortés, and S. R. Jara-Díaz. 2011. "Optimal design and benefits of a short turning strategy for a bus corridor." *Transportation* 38 (1): 169-189.

Ulusoy, Y., S. Chien, and C. Wei. 2010. "Optimal All-Stop, Short-Turn, and Express Transit Services under Heterogeneous Demand." *Transportation Research Record: Journal of the Transportation Research Board* 2197: 8-18.

Wang, Y., T. Tang, B. Ning, T. J. J. van den Boom, and B. D. Schutter. 2015. "Passenger-demands-oriented train scheduling for an urban rail transit network." *Transportation Research Part C: Emerging Technologies* 60: 1-23.

Wu, J., M. Liu, H. Sun, T. Li, Z. Gao, and D. Z. W. Wang. 2015. "Equity-based timetable synchronization optimization in urban subway network." *Transportation Research Part C: Emerging Technologies* 51: 1-18.

Wu, Y., J. Tang, Y. Yu, and Z. Pan. 2015. "A stochastic optimization model for transit network timetable design to mitigate the randomness of traveling time by adding slack time." *Transportation Research Part C: Emerging Technologies* 52: 15-31.

Yin, J, L. Yang, T. Tang, Z. Gao, and B. Ran. 2017. "Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches." *Transportation Research Part B: Methodological* 97: 182-213.

Zhu, Y., B. Mao, and H. Zhou. 2017. "A Short Turning Strategy for Urban Rail Line with Capacity Constraint: A Case Study of a Commuting Subway Line in China." Transportation Research Board 96th Annual Meeting, Washington, D.C., January.