Delft University of Technology

Delft Center for Systems and Control

Technical report 16-006

Integrated line planning and train scheduling for an urban rail transit line*

Y. Wang, X. Pan, S. Su, F. Cao, T. Tang, B. Ning, and B. De Schutter

If you want to cite this report, please use the following reference instead:
Y. Wang, X. Pan, S. Su, F. Cao, T. Tang, B. Ning, and B. De Schutter, "Integrated line planning and train scheduling for an urban rail transit line," *Transportation Research Record*, no. 2540, pp. 66–75, 2016.

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: `https://www.dcsc.tudelft.nl`

Integrated Line Planning and Train Scheduling for an Urban Rail Transit Line

Y. Wang[*], X. Pan[†], S. Su[‡], F. Cao[§], T. Tang[¶], B. Ning[‖], B. De Schutter[**]

Abstract

In this paper, an integrated line planning and train scheduling model based on the circulation of trains is proposed to reduce the passenger dissatisfaction and the operation costs for an urban rail transit line. In the integrated model, the turnaround operations of trains and the departures and arrivals of trains at the depot are included. Furthermore, binary variables are introduced to indicate whether a circulation of a train (i.e., a train service) exists or not and a discrete-event model is used to determine the order of the train services. In addition, a bi-level optimization approach is proposed to solve the integrated line planning and train scheduling problem, where the number of required train services, the headways between train services, the departure times, and the arrival times are optimized simultaneously based on the passenger demands. The performance of the proposed integrated model and bi-level approach is illustrated via a case study of the Beijing Yizhuang line.

1 Introduction

Urban rail transit systems can provide safe, convenient, and punctual passenger services, which are important for the stability and sustainability of public transportation, especially in large cities like Beijing, Shanghai, Tokyo, New York, and Paris. To satisfy the increasing passenger demand, nowadays the headway between trains is often less than 10 minutes and sometimes even close to 2 minutes. When train services are operated with such a short headway, urban rail transit planning becomes more and more important for reducing operation costs and passenger dissatisfaction.

The planning process for rail transit networks has a hierarchical structure (*1*), which is divided into several steps: demand analysis, line planning, train scheduling, rolling stock circulation, and crew scheduling. In this paper, we focus on line planning and train scheduling with known passenger demands. Traditionally, the line plans and the train schedules are optimized sequentially in the hierarchical structure. The line planning problem for the urban rail transit system involves determining the types of line services (i.e., full-length service and short-turning service) and the headways of these line services. The train scheduling process does not use the passenger demands directly but is based on the type of line services and headways obtained in the line planning process. This hierarchical process may result in less optimal results when compared with those obtained by optimizing the two process

[*]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, email: yihui.wang@bjtu.edu.cn
[†]Beijing Subway Operation Company, email: kyyxb@sina.com
[‡]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, email: 10111043@bjtu.edu.cn
[§]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, email: caofang@bjtu.edu.cn
[¶]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, email: ttang@bjtu.edu.cn
[‖]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, email: bning@bjtu.edu.cn
[**]Delft Center for Systems and Control, Delft University of Technology, email: b.deschutter@tudelft.nl

simultaneously. So we propose an integrated line planning and train scheduling approach for an urban rail transit line, where the headways between train services and the departure and arrival times of train services are optimized simultaneously based on the passenger demands. Urban rail transit lines have the following characteristics in general: (a) the urban rail transit lines are usually separate from each other and each direction of a line has a separate rail track; (b) most of the urban rail transit lines contain only one type of line service, namely, full-length train services (which means that all the train services go from the start station of the line to the end station, then turn back at the end station and go towards the start station); (c) the set-up of a train does not change in the normal operation of an urban rail transit line and a train can keep running throughout the operation of the whole day. Note that we only consider full-length train services, so line planning in this paper only determines the headways between trains services.

Line planning and train scheduling for rail transit systems have been studied for decades. Train schedules can be divided into two main categories: (a) periodic (or regular) train schedules and (b) non-periodic train schedules. In periodic train schedules, the whole operation period in a day is split into several small periods, e.g., peak hours, off-peak hours. The departure times at stations follow a certain pattern and the headways of train services are fixed for each small period in the periodic train schedule. Ceder (*2*) proposed four methods to compute the headways for each hour during the operation period using passenger counts. Based on the obtained headways, Ceder provided alternative methods for constructing schedules in (*3*). Liebchen (*4*) formulated a periodic event-scheduling problem and obtained periodic schedules for the Berlin subway system using genetic algorithms and integer programming. As stated in (*5*), periodic train schedules are still not fully sensitive and responsive to the time-varying passenger demands even though the headways are calculated based on the passenger demands. A non-periodic train schedule is obtained to minimize passenger dissatisfaction and operation costs in (*6*), where a Lagrangian relaxation approach is used to solve the resulting nonlinear problem in a hierarchical manner. A model predictive control method is proposed in (*7*) to generate non-periodic train schedules for a whole day. Niu and Zhou optimized non-periodic train schedules for an urban rail transit line with consideration of time-varying origin-destination passenger demands in heavily congested situations in (*8*). Moreover, in (*9*) we proposed an efficient bi-level approach for an urban rail transit line with consideration of time-varying passenger demand and stop-skipping to minimize the total travel time and the total energy consumption. However, in (*8,9*)only one operation direction of an urban rail transit line is considered and the details about the turnaround operation at terminals and the train circulation between the depot and terminals are not considered, although they are usually the bottlenecks of the line and they are critical for the train schedule.

In this paper, an integrated line planning and train scheduling approach is proposed to minimize the operation cost and the passenger dissatisfaction. The optimized train schedule is non-periodic, where the headway varies with the passenger demands. The current paper extends our previous research (*9*) in the following aspects:

1. The integrated model used in the approach is based on the circulation of available trains, where the turnaround operation at the terminals and the departures and arrivals of trains at the depot are included in the model formulation.

2. The number of required train services is not fixed in the scheduling period, but is optimized based on the passenger demands.

Note that the description of passenger demands is different from the origin-destination passenger demand in our previous work (*9*). Here, the passenger demands is described based on the cumulative data (i.e., the numbers of passengers traveling between two consecutive stations in a certain time interval) obtained in practice.

This paper is structured as follows. Section 2 formulates an integrated line planning and train scheduling model, which describes the operation of trains and the passenger demand characteristics. Section 3 describes the objective function and the constraints of the integrated line planning and train scheduling problem. Section 4 proposes a bi-level optimization approach for the integrated problem. Section 5 illustrates the performance of the proposed integrated model and solution approach with a case study of the Beijing Yizhuang line. Finally, Section 5 concludes the paper.

2 Model formulation

This section formulates the model of the operation of trains and the passenger demand characteristics for the integrated line planning and train scheduling problem.

2.1 Assumptions

When formulating the passenger demand oriented train scheduling model, we make the following assumptions:

A.1 The running times and dwell times for all the trains are constant but still station-dependent.

A.2 There exists only one depot for the urban rail transit line.

A.3 The operation of trains in the maintenance depot is not included in the train scheduling model.

Assumption A.1 generally holds for advanced urban rail transit systems, where automatic train control systems are implemented to control the operation of trains and the door opening/closing process automatically. For some urban rail transit systems, the dwell times may depend on the number of alighting and boarding passengers, the number of passengers onboard trains, the number of passengers waiting at platforms, the number of doors, etc. In particular, when the train and/or the platforms are oversaturated, the dwell times could be much longer than expected. However, since the passenger demands of urban rail transit are characterized certain patterns corresponding to e.g. weekdays, weekends, and holidays, the dwell times at stations could be estimated using historical data for different passenger demand patterns throughout a day. The estimates of the dwell times can then be used in the train scheduling approach to generate different train schedules for different passenger demand patterns. Assumption A.2 is made for the sake of the simplicity in the description of the circulation of trains. In addition, Assumption A.3 is made to simplify the model of train operations since the operation of trains in the depot is normally decided after the train scheduling phase, see (*10*) for more information.

2.2 Operation of trains

We consider an urban rail transit line as shown in Figure 1(a), which has J stations with $S_{\mathrm{sta}} = \{1,2,3,\ldots,J\}$ the set of stations. Station 1 is assumed to be connected with the depot of this urban rail transit line. The turnaround operation of trains happens at station 1 and at station J. The direction from station 1 to station J is defined as the down direction and is indicated as 'dn'. In addition, the direction from station J to station 1 is defined as the up direction and is indicated as 'up'. The total number of available trains for operation in the depot is denoted as I and the set of trains is denoted as $S_{\mathrm{tra}} = \{1,2,3,\ldots,I\}$. Note that the capacity of the depot is larger than or equal to I.

2.2.1 Circulation of trains

As illustrated by the red dashed line in Figure 1(a), a circulation of a train, i.e., a train service, consists of the following four phases:

P1: A train (that could come out from the depot or turn around from the previous cycle operation) departs from station 1 and goes towards station J;

P2: The train turns around at station J;

P3: The train departs from station J and goes towards station 1;

P4: That train goes back to the depot or turns around at station 1.

The time-space diagram for the circulation of trains is illustrated in Figure 1(b), where the four phases, i.e., P1, P2, P3, and P4, are denoted by the dark lines, dash-dotted lines, dashed lines, and dotted lines. Furthermore, the operation of trains between the depot and station 1 is illustrated with grey lines. Recall that the operation of trains in the depot is not considered in this paper as indicated in Assumption A.3. The detailed information about the departures, arrivals, and turnaround operations of trains will be given in Section 2.2.2.

A train could operate for the whole operation period and it could also operate a few cycles and then go into the depot. The train schedule of an urban rail transit line can be described by the circulations of all the available trains and the order of these train services as illustrated in Figure 2. In particular, the order of the train services will be determined by the discrete-event model introduced in Section 2. As shown in Figure 2, a cycle index is assigned for each circulation of a train. Let $C_{\max}$denote the maximum number of circulations for the operation period, which can be calculated by

$$C_{\max} = \text{ceil}\left(\frac{t_{\text{end}} - t_{\text{start}}}{T_{\text{cycle,min}}}\right), \tag{1}$$

where ceil$(\cdot)$ is the ceiling function, $[t_{\text{start}}, t_{\text{end}}]$ is the whole operation period for trains, and $T_{\text{cycle,min}}$ is the minimal time for a cycle operation for a train. The minimal circulation time is determined by the running times, dwell times, and the minimum turnaround times at terminal stations. In addition, we introduce binary variables $\delta_{i,c}$ to denote the operation status of train i in cycle c:

$$\delta_{i,c} = \begin{cases} 1 & \text{if train } i \text{ operates in cycle } c, \\ 0 & \text{if train } i \text{ waits at the depot in cycle } c. \end{cases} \tag{2}$$

Furthermore, we define the operation of train i in cycle c as train service (i,c). So if train i operates in cycle c, i.e., $\delta_{i,c} = 1$, train service (i,c) exists, otherwise, train service (i,c) does not exist. Whether a train service exists or not, i.e., $\delta_{i,c}$ equals 1 or 0, is decided by the passenger demands, the load factor of train services, the operation costs, and the irregularity of train services (see Section 2.3 and Section 3.1 for details).

2.2.2 Departures and arrivals

The arrival times and departure times of the down direction are denoted as $a^{\text{dn}}_{i,c,j}$ and $d^{\text{dn}}_{i,c,j}$, respectively, where i is the train index, c is the cycle index, and j is the station index. In a similar way, the arrival times and departure times of the up direction are denoted as $a^{\text{up}}_{i,c,j}$ and $d^{\text{up}}_{i,c,j}$. Based on Assumption

A.1, the running times and dwell times are independent from train index i and cycle index c. If train i operates in cycle c, i.e., $\delta_{i,c} = 1$, then we have

$$a^{\text{dn}}_{i,c,j+1} = d^{\text{dn}}_{i,c,j} + r^{\text{dn}}_{j,j+1}, \tag{3}$$

$$d^{\text{dn}}_{i,c,j+1} = a^{\text{dn}}_{i,c,j+1} + \tau^{\text{dn}}_{j+1}, \tag{4}$$

where $j \in S_{\text{sta}}/\{J\}$, $r^{\text{dn}}_{j,j+1}$ is the running time from station j to station $j+1$ in the down direction and τ^{dn}_{j+1} is the dwell time at station $j+1$ in the down direction. When train i with cycle c departs at station J in the down direction, the train will turn around and go to the up direction, which can be formulated as

$$a^{\text{up}}_{i,c,J} = d^{\text{dn}}_{i,c,J} + r^{\text{turn}}_{J}, \tag{5}$$

$$d^{\text{up}}_{i,c,J} = a^{\text{up}}_{i,c,J} + \tau^{\text{up}}_{J}, \tag{6}$$

where r^{turn}_{J} is the turnaround time at station J from the down direction to the up direction and $\tau^{\text{up}}_{i,c,J}$ is the dwell time at station J for the up direction. The turnaround time r^{turn}_{J} is a variable, which should satisfy

$$r^{\text{turn}}_{J,\min} \leq r^{\text{turn}}_{J} \leq r^{\text{turn}}_{J,\max}, \tag{7}$$

where $r^{\text{turn}}_{J,\min}$ is the minimum turnaround time decided by the line structure, train characteristics, signaling systems, etc., and $r^{\text{turn}}_{J,\max}$ is the maximum turnaround time specified by the rail operator. Similarly, the arrival time $a^{\text{up}}_{i,c,j-1}$ and the departure time $d^{\text{up}}_{i,c,j-1}$ at station $j-1$ in the up direction is

$$a^{\text{up}}_{i,c,j-1} = d^{\text{up}}_{i,c,j} + r^{\text{up}}_{j,j-1}, \tag{8}$$

$$d^{\text{up}}_{i,c,j-1} = a^{\text{up}}_{i,c,j-1} + \tau^{\text{up}}_{j-1}, \tag{9}$$

with $j \in S_{\text{sta}}/\{1\}$. Furthermore, if train i operates both in cycle c and $c+1$, i.e., $\delta_{i,c} = 1$ and $\delta_{i,c+1} = 1$, train i needs to turn around at station 1 after the operation of cycle c and then start the operation of cycle $c+1$, which can be written as

$$a^{\text{dn}}_{i,c+1,1} = d^{\text{up}}_{i,c,1} + r^{\text{turn}}_{1}, \quad \text{if } \delta_{i,c} = 1 \text{ and } \delta_{i,c+1} = 1, \tag{10}$$

$$d^{\text{dn}}_{i,c+1,1} = a^{\text{dn}}_{i,c+1,1} + \tau^{\text{dn}}_{1}, \quad \text{if } \delta_{i,c} = 1 \text{ and } \delta_{i,c+1} = 1, \tag{11}$$

where the turnaround time r^{turn}_{1} should satisfy

$$r^{\text{turn}}_{1,\min} \leq r^{\text{turn}}_{1} \leq r^{\text{turn}}_{1,\max}. \tag{12}$$

2.2.3 Discrete-event model for the train services

A discrete-event model similar to that proposed in (*11*) is introduced to describe the order of train services in the urban rail transit line. Since in this paper we only consider the full-length line service and there is only one depot as given in Figure 1(a), the order of train services can only be changed at station 1. Therefore, we only consider the departures of train services at station 1 of the down direction in the discrete-event model. The n-th event e_n occurring in the discrete-event system is denoted as

$$e_n = (i_n, c_n, t_n), \tag{13}$$

where n is the index of the event and t_n is the event time, i.e., the departure time of train service (i_n, c_n) at station 1 in the down direction. The total number of events is equal to the number of all train services, which is denoted as N_{service} and is calculated by equation (22) in Section 3.1. Furthermore, we define an event list E_{dep} which contains all the events and can be written as $\{e_1, e_2, \ldots, e_{N_{\text{service}}}\}$.

If event $e_m = (i_m, c_m, t_m)$ is a successor of event $e_n = (i_n, c_n, t_n)$, then train service (i_m, c_m) departs from station 1 after train service (i_n, c_n). These two train services should satisfy the headway constraints for the up and down direction as follows:

$$h_{\min} \leq d^{\text{dn}}_{i_m,c_m,1} - d^{\text{dn}}_{i_n,c_n,1} \leq h_{\max}, \tag{14}$$

where $h_{\min}$ and $h_{\max}$ are the minimal and maximal headway between two consecutive train services. Because the running times and dwell times between and at stations are the same for all train services in the urban rail transit line, if the headway constraint is satisfied at station 1 in the down direction, then the headway constraints at other stations for the down direction will be satisfied automatically. However, the turnaround times at the station J are not fixed for all train services, so the headway at station J in the up direction should be included, i.e.,

$$h_{\min} \leq d^{\text{up}}_{i_m,c_m,J} - d^{\text{up}}_{i_n,c_n,J} \leq h_{\max}. \tag{15}$$

Similarly, the headway constraints for other stations in the up direction will also be satisfied automatically. Furthermore, the turnaround operation of the two consecutive train services (i_n, c_n) and (i_m, c_m) at station J should satisfy

$$d^{\text{dn}}_{i_m,c_m,J} - d^{\text{up}}_{i_n,c_n,J} > 0, \tag{16}$$

which means that train service (i_m, c_m) can start the turnaround operation at station J only after train service (i_n, c_n) has already departed from station J in the up direction. The turnaround constraint at station 1 also depends on whether train service $(i_n, c_n + 1)$ and $(i_m, c_m + 1)$ exist or not, i..e, the value of binary variables δ_{i_n,c_n+1} and δ_{i_m,c_m+1}. If train services $(i_n, c_n + 1)$ and $(i_m, c_m + 1)$ exist, then the following constraint should be satisfied

$$d^{\text{up}}_{i_m,c_m,1} - d^{\text{dn}}_{i_n,c_n+1,1} > 0, \quad \text{if } \delta_{i_n,c_n+1} = 1 \text{ and } \delta_{i_m,c_m+1} = 1. \tag{17}$$

If one of the train services $(i_n, c_n + 1)$ and $(i_m, c_m + 1)$ does not exists or both of them do not exist, then the turnaround constraint at station 1 as given in (17) is not needed.

2.3 Passenger demand

The passenger demand of an urban rail transit line varies systematically with the day of the week, the time of a day, station, operation direction, etc. With the employment of advanced data collection systems, more and more accurate passenger information is available. As an illustration, the number of passengers traveling between Songjiazhuang station and Xiaocun station in the down direction of the Beijing Yizhuang line is given in Figure 3, where the number of traveling passengers is counted for every half an hour. The solid line and the dashed line illustrate the passenger demand on a weekday and a weekend day, respectively. The crowdedness of train services can be measured by the load factor, which is defined as a ratio of the number of onboard passengers to the capacity of trains (that includes the number of seats and the maximum allowed number of standing passengers). The load factor of trains in the busy urban lines could be larger than 0.8 and even larger than 1 in some extreme cases. Therefore, it is important to consider the load factors of trains in the train scheduling problem to reduce the passenger dissatisfaction.

In this paper, we propose an integrated line planning and train scheduling approach for an urban rail transit line, where the number of required train services, the headways between these train services, and the departure and arrival times are optimized simultaneously based on the passenger demands. The passenger demands can be analyzed and estimated based on historical and real-time data collected by the automatic fare systems. As illustrated in Figure 3, we do not consider the origins and destinations of passengers, but we translate the passenger demands into the number of passengers *wanting* to take a train from station j to $j+1$ in the down direction and from station j to $j-1$ in the up direction for each time slot based on the historical data. After the translation process, we define the passenger traveling rates between two consecutive stations of the up and down direction as piecewise constant functions, where we assume that whenever a train service arrives all passengers can board the train even if there is no space; however, this will later on be penalized via the load factor. The piecewise constant function for the down direction can be written as follows:

$$\lambda^{\mathrm{dn}}_{j,j+1}(t) = \begin{cases} \beta^{\mathrm{dn}}_{j,j+1,1}/(t^{\mathrm{dn}}_1 - t_{\mathrm{start}}), & \text{for } t \in [t_{\mathrm{start}}, t^{\mathrm{dn}}_1) \\ \cdots & \cdots \\ \beta^{\mathrm{dn}}_{j,j+1,k}/(t^{\mathrm{dn}}_k - t^{\mathrm{dn}}_{k-1}), & \text{for } t \in [t^{\mathrm{dn}}_{k-1}, t^{\mathrm{dn}}_k) \\ \cdots & \cdots \\ \beta^{\mathrm{dn}}_{j,j+1,K}/(t_{\mathrm{end}} - t^{\mathrm{dn}}_K), & \text{for } t \in [t^{\mathrm{dn}}_K, t_{\mathrm{end}}] \end{cases} \tag{18}$$

where the operation period $[t_{\mathrm{start}}, t_{\mathrm{end}}]$ is split into K time slots with the splitting time instants $t^{\mathrm{dn}}_1, t^{\mathrm{dn}}_2, \ldots,$ t^{dn}_K for the down direction, $\beta^{\mathrm{dn}}_{j,j+1,k}$ with $k \in \{1,2,\ldots,K\}$ is the total numbers of passengers traveling between two consecutive stations during the splitting time slots.. A similar piecewise constant function $\lambda^{\mathrm{up}}_{j,j-1}(\cdot)$ can be defined for the up direction. If event $e_m = (i_m, c_m, t_m)$ is a successor of event $e_n = (i_n, c_n, t_n)$, then the numbers of passengers $P^{\mathrm{up}}_{i_m,c_m,j,j-1}$ and $P^{\mathrm{dn}}_{i_m,c_m,j,j+1}$ on board train services (i_m, c_m) when departing from station j in the up and down direction can be computed by

$$P^{\mathrm{up}}_{i_m,c_m,j,j-1} = \int_{d^{\mathrm{up}}_{i_n,c_n+1,j}}^{d^{\mathrm{up}}_{i_m,c_m,j}} \lambda^{\mathrm{up}}_{j,j-1}(t)\mathrm{d}t,$$

and

$$P^{\mathrm{dn}}_{i_m,c_m,j,j+1} = \int_{d^{\mathrm{dn}}_{i_n,c_n+1,j}}^{d^{\mathrm{dn}}_{i_m,c_m,j}} \lambda^{\mathrm{dn}}_{j,j+1}(t)\mathrm{d}t,$$

respectively. In order to reduce the passenger dissatisfaction, the load factors for all the train services at all stations are considered in the integrated model. Let $\sigma^{\mathrm{up}}_{i,c,j,j-1}$ and $\sigma^{\mathrm{dn}}_{i,c,j,j+1}$ denote the load factors for train service (i,c) between station j and the next station in the up and down direction. The load factors can be calculated by

$$\sigma^{\mathrm{up}}_{i,c,j,j-1} = \frac{P^{\mathrm{up}}_{i,c,j,j-1}}{C_{\mathrm{train}}} \text{ and } \sigma^{\mathrm{dn}}_{i,c,j,j+1} = \frac{P^{\mathrm{dn}}_{i_m,c_m,j,j+1}}{C_{\mathrm{train}}},$$

where C_{train} is the capacity of trains including the number of seats and the maximum allowed number of standing passengers. In the train scheduling process, the load factor of train service (i,c) between any two consecutive stations should be smaller than the predefined load factor $\sigma_{\max}$, i.e.,

$$\begin{aligned} &\sigma^{\mathrm{up}}_{i,c,j,j-1} \leq \sigma_{\max}, \quad \text{for } j \in S_{\mathrm{sta}}/\{1\}, \\ &\sigma^{\mathrm{dn}}_{i,c,j,j+1} \leq \sigma_{\max}, \quad \text{for } j \in S_{\mathrm{sta}}/\{J\}. \end{aligned} \tag{19}$$

3 The integrated line planning and train scheduling problem

We first formulate the objective function of the integrated problem, which involves the operation costs and the passenger dissatisfaction cost. Next, the constraints of the integrated problem are also defined in this section.

3.1 Performance criteria

The operation costs of rail operators depend on the number of trains required for the operation and the number of train kilometers, i.e., the total running distance of all train services. The number of required trains affects the maintenance schedule of trains and also the number of trains equipped for the urban rail transit line. Based on the train scheduling and circulation model in Section 0, the number of trains put into operation, i.e., the maximum of train services among all cycles, can be calculated by

$$N_{\text{train}} = \max_{c \in S_{\text{cyc}}} \Big(\sum_{i \in S_{\text{tra}}} \delta_{i,c} \Big), \tag{20}$$

where $\sum_{i \in S_{\text{tra}}} \delta_{i,c}$ is the number of trains operating in cycle c. Furthermore, the number of train kilometers D_{train} can be calculated by

$$D_{\text{train}} = L N_{\text{service}}, \tag{21}$$

where L is the running distance for a circulation of a train and N_{service} is the total number of train services

$$N_{\text{service}} = \sum_{c \in S_{\text{cyc}}} \sum_{i \in S_{\text{tra}}} \delta_{i,c}. \tag{22}$$

The passenger dissatisfaction is affected by the crowdedness of trains and the waiting time at stations, where the load factor constraints in (19) are introduced to limit the crowdedness of train services in the integrated model. In this paper, we assume that the passenger arrivals in urban rail transit systems are uniformly distributed. According to the statements in (*6*), the minimum passenger waiting time is achieved when the headway variations between consecutive train services are minimized. Therefore, we also include the headway variations between consecutive train services in the performance criterion of the train scheduling problem, where the headway variation is defined as the square of the difference between the current headway and the mean of the neighboring headways. We sort the discrete events in the event list E_{sep} and denote it as the *ordered* set $\{e'_1, e'_2, \ldots, e'_{N_{\text{service}}}\}$ with N_{service} the total number of events. When calculating the cost of irregularity, the numbers of neighboring events involved are n_1 and n_2 for the left side and right side, respectively. So the cost of irregularity for the up and down direction can be written as

$$\begin{aligned} P_{\text{irregularity}} = & \sum_{n \in \{1,2,\ldots,N_{\text{service}}\}} \left(\left(d^{\text{dn}}_{i'_n, c'_n, 1} - d^{\text{dn}}_{i'_{n-1}, c'_{n-1}, 1} \right) - \frac{\sum\limits_{m=\max(n-n_1,1)}^{\min(n+n_2, N_{\text{service}})} \left(d^{\text{dn}}_{i'_m, c'_m, 1} - d^{\text{dn}}_{i'_{m-1}, c'_{m-1}, 1} \right)}{n_1 + n_2 + 1} \right)^2 \\ & + \sum_{n \in \{1,2,\ldots,N_{\text{service}}\}} \left(\left(d^{\text{up}}_{i'_n, c'_n, 1} - d^{\text{up}}_{i'_{n-1}, c'_{n-1}, 1} \right) - \frac{\sum\limits_{m=\max(n-n_1,1)}^{\min(n+n_2, N_{\text{service}})} \left(d^{\text{up}}_{i'_m, c'_m, 1} - d^{\text{up}}_{i'_{m-1}, c'_{m-1}, 1} \right)}{n_1 + n_2 + 1} \right)^2 . \end{aligned} \tag{23}$$

Note that the max and min function are introduced to ensure event $e'_m = (i'_m, c'_m, t'_m)$ belongs to the event list.

The objective of the integrated train scheduling and circulation problem is to minimize the operation costs and the irregularity of the train services, i.e., the overall cost function is

$$f = \xi_1 N_{\text{train}} + \xi_2 D_{\text{train}} + \xi_3 P_{\text{irregularity}}, \tag{24}$$

where ξ_1 is the cost for using a physical train in the operation period, ξ_2 is the cost for a train running for a kilometer, and ξ_3 is the cost of the irregularity of the train services.

3.2 Constraints

The constraints of the train scheduling problem consist of the departure/arrival constraints (3)-(12), headway constraints in the event model (14)-(16), and passenger demand constraints (19). In addition, the departure/arrival times of the operating train service (i,c) should be in the operation period, i.e.,

$$\begin{aligned} t_{\text{start}} \leq d^{\text{dn}}_{i,c,j} \leq t_{\text{end}}, \quad t_{\text{start}} \leq a^{\text{dn}}_{i,c,j} \leq t_{\text{end}}, \\ t_{\text{start}} \leq d^{\text{up}}_{i,c,j} \leq t_{\text{end}}, \quad t_{\text{start}} \leq a^{\text{up}}_{i,c,j} \leq t_{\text{end}}. \end{aligned} \tag{25}$$

Moreover, rail transit operators have rigid constraints for the departure times of the first train service and the last train service. So we have the following constraints:

$$d^{\text{dn}}_{i'_1,c'_1,1} = d_{\text{start}}, \tag{26}$$

$$d^{\text{up}}_{i'_{N_{\text{service}}},c'_{N_{\text{service}}},J} = d_{\text{end}}, \tag{27}$$

where train services (i'_1, c'_1) and $(i'_{N_{\text{service}}}, c'_{N_{\text{service}}})$ are the first and last train services in the scheduling period, d_{start} is the required departure time of the first train service at station 1 in the down direction, and d_{end} is the required departure time of the last train service at station J in the up direction.

4 Solution approach

The integrated line planning and train scheduling problem proposed in Section 3 is a mixed integer nonlinear programming (MINLP) problem, where the load factor constraints (19) are nonlinear. The discrete-event model proposed in Section 2, where the order of all train services is modeled, introduces additional nonlinearities into the problem. The resulting MINLP problem can be solved using a direct MINLP approach, implemented via solvers such as MINLP BB (*12*) and SCIP (*13*). However, a direct MINLP approach can only solve small-sized scheduling problems. Therefore, we apply the bi-level optimization approaches proposed in our previous paper (*14*) to solve the integrated line planning and train scheduling problem.

The bi-level optimization method consists of two levels of optimization. The high level decides whether a train service exists or not, i.e., it optimizes the binary variables $\delta_{i,c}$. Integer programming approaches can be adopted for the high-level optimization, such as genetic algorithms. The low level optimizes the order of the existing train services and the departure times of these train services at stations, which results in a real-valued nonlinear problem since the binary variables are now fixed. For the low-level optimization, multi-start sequential quadratic programming algorithms (*15*) or pattern search (*16*) can be applied. See (*14*) for more detailed information about the procedure of the bi-level optimization method.

5 Case study

The performance of the integrated line planing and train scheduling approach is demonstrated via a case study based on the Beijing Yizhuang line. This line has 14 stations, where we introduce indices for these 14 stations. Yizhuang is station 1 and Songjiazhuang is station 14. The depot is connected with station 1. The running times between any two consecutive stations and the dwell times at stations for the up and down direction of the Yizhuang line are given in Table 1. The passenger demands on a weekend day for the up and down direction of the Beijing Yizhuang line are given in Figure 4, where the number of passengers traveling between any two consecutive stations is counted every half hour. Based on the passenger demands on a weekend day, 10 trains are sufficient for the operation and the capacity of a train is 1440 passengers. The maximum load factor of train services is chosen as 0.75 to guarantee passenger satisfaction. In addition, the maximum and minimum headway between train services on a weekend day are defined as 660 s and 240 s by the rail operator. The required departure time d_{start} of the first train service at station 1 in the down direction is 5:20:00 and the required departure time d_{end} of the last train service at station J in the up direction is 22:45:00. Furthermore, the turnaround time at station 1 and station 14 should be larger than 120 s and less than 720 s. In addition, the weights ξ_1, ξ_2, and ξ_3 in the objective function are chosen as 10, 0.1, and 0.005 in this case study.

For the high-level binary optimization problem we use a genetic algorithm for the binary optimization, where the ga function in the global optimization toolbox of Matlab is applied. The low-level nonlinear programming optimization problem is solved by the sequential quadratic programming algorithm implemented in the fmincon function of the Matlab optimization toolbox. The obtained train schedule by the bi-level optimization for a weekend day is compared with the current train schedule used in the Beijing Yizhuang line as illustrated in Figure 5. The train schedule currently used in practice is regular, where the scheduling period is split into peak hours and off-peak hours and the headways in the peak and off-peak hours are 660 s and 465 s, respectively. However, the train schedule obtained by the integrated model is a non-regular train schedule, where the headways between train services are changing with the passenger demands as shown in Figure 5, where the maximum headway is 660 s and the minimum headway is around 500 s. A comparison between the train schedule used in practice and the optimized train schedule is given in Table 2. When compared with the train schedule used in practice, the number of required trains in the optimized schedule is reduced from 10 to 9, i.e., one train is saved for the train circulation in daily operation. In addition, the number of train services is reduced from 108 in the train schedule used in practice to 102 in the optimized train schedule. The maximum number of cycles of each train is 13 for both the train schedule used in practice and the optimized train schedule. Moreover, the maximum load factor for all train services between two consecutive stations is 0.98 in the train schedule used in practice and there are 7 train services, the load factors of which are larger than 0.75 as shown in Figure 5(b). However, the maximum load factor of the optimized train schedule is guaranteed to be less than or equal to 0.75, even though the number of train services is reduced.

Based on the comparison with the train schedule used in practice of the Beijing Yizhuang line, the train schedule obtained using the integrated line planning and train scheduling model achieves a better performance both for the rail operator and for the passengers. In addition, the proposed bi-level optimization approach can solve the integrated train scheduling and circulation problem and result in an acceptable train schedule. However, the computation time of the bi-level approach for 10 trains and 13 cycles is $6.85 \cdot 10^5$ s on a 2.9 GHz Intel Core i7-3520M CPU running on a 64-bit windows operating system, which is slow for rail operators. However, the train schedule is obtained using off-line computations, and therefore the computation time could be further reduced if using parallel processing, since both genetic algorithms and sequential quadratic programming algorithms can be performed in parallel. Since the headway between trains is small, the train schedules of the Beijing subway are not published to all passengers. There are screens at platforms to provide the train service information to passengers. So the train schedule with varying headways will not affect the passengers' behaviors and service management. In addition, because the Beijing Yizhuang line is equipped with a communication-based train control system, where a automatic train operation system and a automatic train supervision system are being used in practice, the train schedule with varying headways can be applied directly.

6 Conclusions and future work

We have considered the integrated line planning and train scheduling problem for an urban rail transit line while taking the passenger demands into account. A circulation of a train on an urban rail transit line is defined as a train service. The train schedule can then be formulated as a sequence of train services, i.e., the circulation of available trains, and the order of these train services, which is described by a proposed discrete-event model. The integrated line planning and train scheduling problem is essentially a mixed-integer nonlinear programming problem, which is solved by a proposed bi-level optimization approach. The simulation results show that the optimized train schedule yields a better performance than the train schedule currently used in practice. We will perform various case studies for the proposed train schedule approach and analyze the robustness, stability, etc. of the optimized train schedules. In addition, a quantitative approach will be proposed to evaluate the prices of the crowdedness dissatisfaction and the waiting time of passengers to obtain a better train schedule. The test of the optimized train schedule will be carried out in the Beijing Yizhuang line in future. Moreover, the proposed bi-level optimization approach can be applied to small-sized and medium-sized train scheduling problems. However, for large-scale train scheduling problems hierarchical and/or distributed optimization approaches could be investigated, where the urban rail transit line can be decomposed into small parts to decrease the size of the problem and to reduce the computational time. Furthermore, we optimize the train schedule based on a static passenger demand profile in this paper. We will develop receding horizon and/or robust control approaches that can deal with dynamic variations of passenger demands.

Acknowledgments

Research supported by the National Natural Science Foundation of China (No. 61503020), the research foundation of Beijing Jiaotong University (Contract No. 2015RC003), and the Beijing Municipal Commission of Education (Contract No. I15H100010). The authors would like to thank Prof. Ingo Hansen for his constructive suggestions, which improved the content substantially.

References

1. Ghoseiri, K., F. Szidarovszky, and M. Asgharpour, A multi-objective train scheduling model and solution. *Transportation Research Part B: Methodological*, vol. 38, no. 10, Dec. 2004, pp. 927–952.

2. Ceder, A., Bus frequency determination using passenger count data.*Transportation Research Part A*, vol. 18A, no. 5/6, May 1984, pp. 439–453.

3. Ceder, A., Methods for creating bus timetables. *Transportation Research Part A*, vol. 21A, no. 1, Jan. 1986, pp. 59–83.

4. Liebchen, C., The first optimized railway timetable in practice. *Transportation Science*, vol. 42, no. 4, Nov. 2008, pp. 420–435.

5. Niu, H., 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, Nov. 2015, pp. 117–135.

6. Cury, J., F. Gomide, and M. J. Mendes, A methodology for generation of optimal schedules for an underground railway systems. *IEEE Transactions on Automatic Control*, vol. 25, no. 2, Apr. 1980, pp. 217–222.

7. Assis, W., and B. Milani, Generation of optimal schedules for metro lines using model predictive control. *Automatica*, vol. 40, no. 8, Aug. 2004, pp. 1397–1404.

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

9. Wang, Y., B. De Schutter, T. van den Boom, B. Ning, and T. Tang, Efficient bi-level approach for urban rail transit operation with stop-skipping. *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, Dec. 2014, pp. 2658–2670.

10. Flamini, M., and D. Pacciarelli, Real time management of a metro rail terminus. European Journal of Operational Research, vol. 189, no. 3, Sep. 2008, pp. 746–761.

11. Wang, Y., T. Tang, B. Ning, T. 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, Oct. 2015, pp. 1–23.

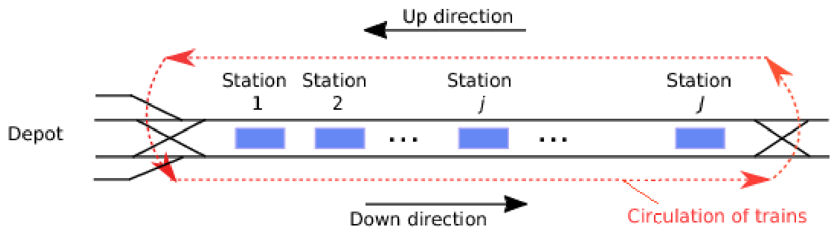
12. Holmström, K., A.O.Göran, and M. Edvall, User's Guide for Tomlab/MINLP, Feb. 2007, URL: http://tomopt.com/docs/TOMLAB_MINLP.pdf

13. Berthold, T., G. Gamrath, A. Gleixner, S. Heinz, T. Koch, and Y. Shinano, *Solving mixed integer linear and nonlinear problems using the SCIP Optimization Suite*. ZIB-Report 12-17, Zuse Institute Berlin, Berlin, Germany, 2012.

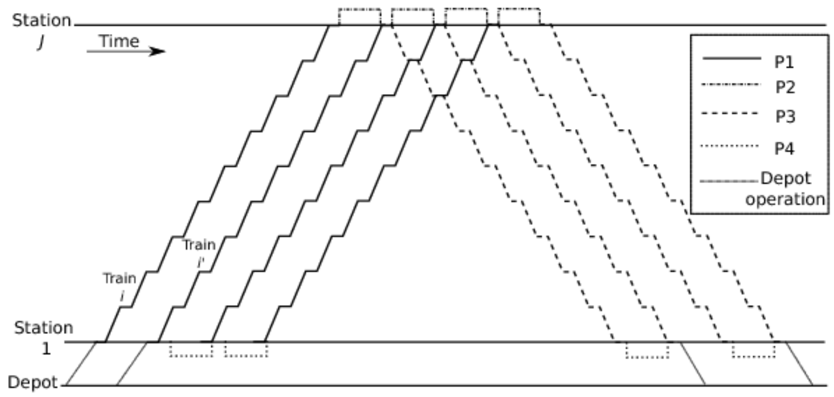
14. Wang, Y., B. De Schutter, T. van den Boom, B. Ning, and T. Tang, Origin-destination dependent train scheduling problem with stop-skipping for Urban Rail Transit Systems. In *Proceedings of the 93nd Annual Meeting of the Transportation Research Board*, Washington, DC, Jan. 2014.

15. Boggs, P., and J. Tolle, Sequential Quadratic Programming. *Acta Numerica*, vol. 4, no. 1, Jan. 1995, pp. 1–51.

16. Hooke, R., and T. Jeeves, "Direct search" solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, vol. 8, no. 2, Apr. 1961, pp. 212–229.



(a) The structure of an urban rail transit line



(b) Four phases of a circulation of a train

Figure 1: The urban rail transit line structure and the four phases of a cyclic train operation

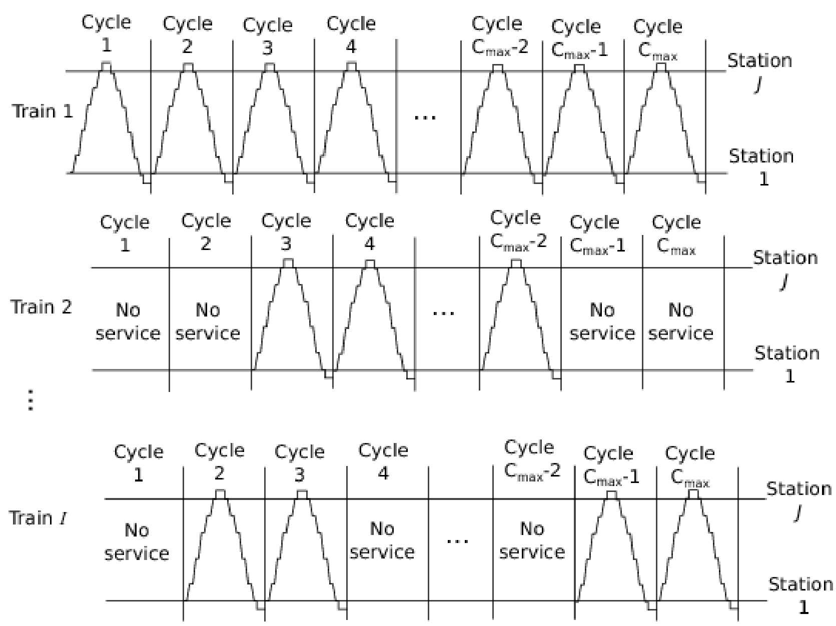


Figure 2: The train schedule described by cycle operations

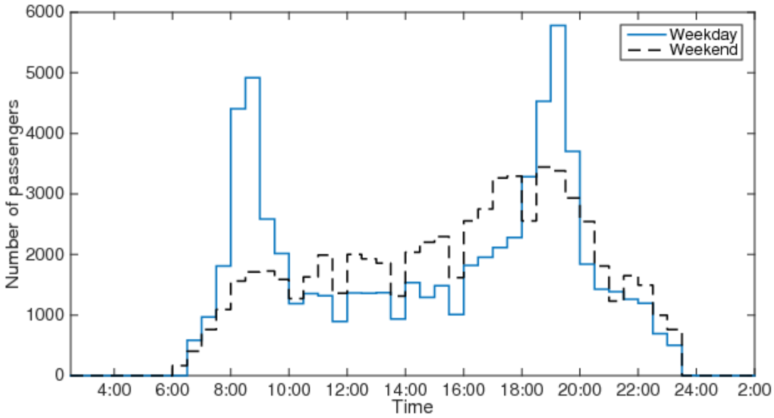
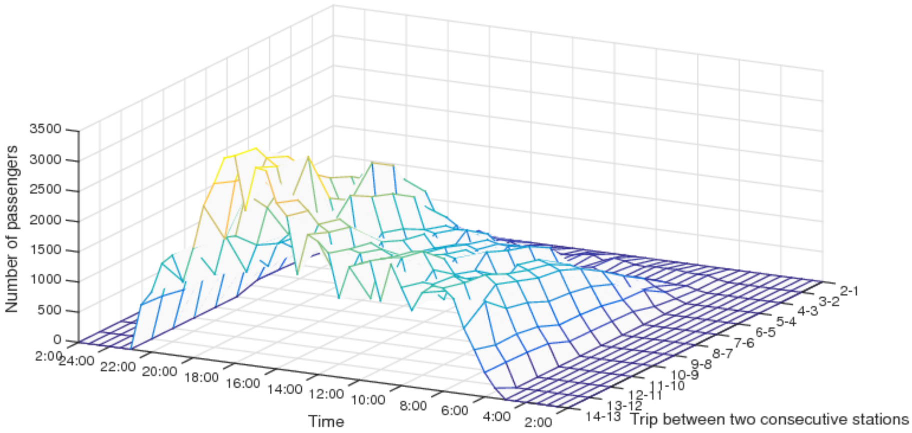
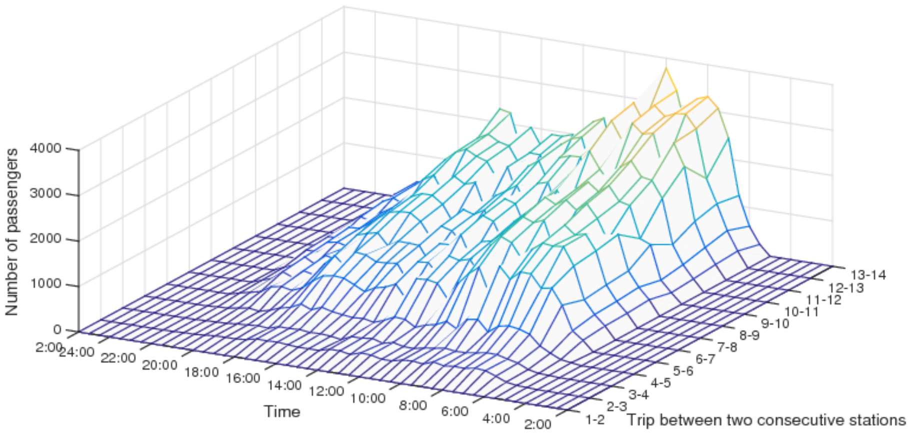


Figure 3: The number of passengers traveling from Songjiazhuang to Xiaocun in the down direction of Beijing Yizhuang line

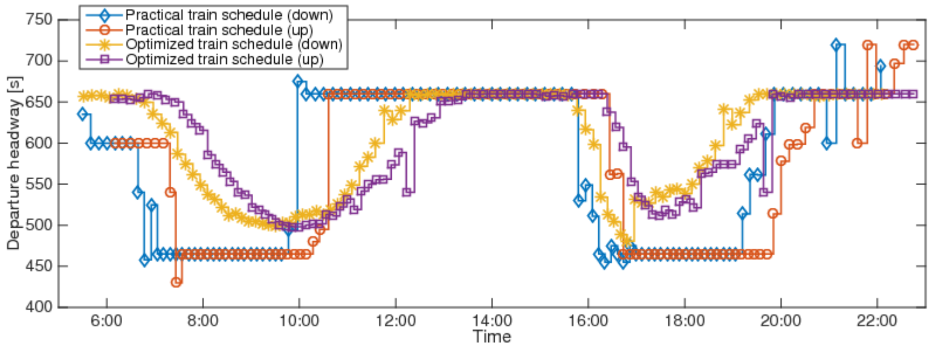


(a) Passenger demand for the up direction

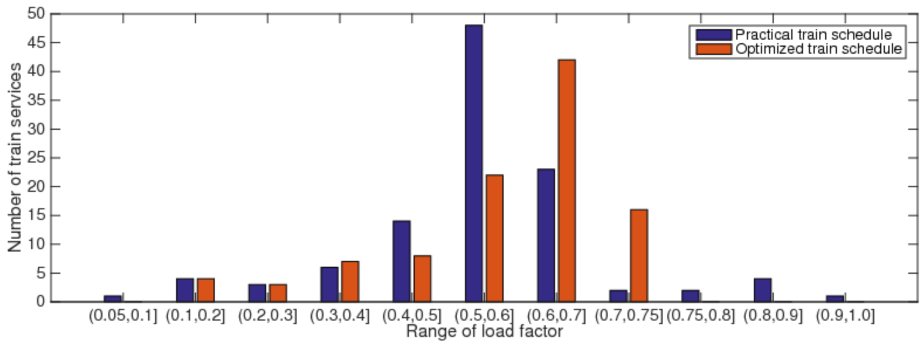


(b) Passenger demand for the down direction

Figure 4: The number of passengers [passengers/half hour] traveling between two consecutive stations for the Beijing Yizhuang line



(a) Headways for the up and down direction



(b) Number of train services in different ranges of the load factor

Figure 5: The comparison between the train schedule used in practice and the optimized train schedule for the Beijing Yizhuang line

Table 1: The running times and dwell times of the Beijing Yizhuang line

Segment between stations	1→2	2→3	3→4	4→5	5→6	6→7	7→8
Running time (down) [s]	110	100	141	150	162	103	101
Segment between stations	8→9	9→10	10→11	11→12	12→13	13→14	-
Running time (down) [s]	111	90	135	157	105	195	-
Segment between stations	14→13	13→12	12→11	11→10	10→9	9→8	8→7
Running time (up) [s]	190	108	157	135	90	114	103
Segment between stations	7→6	6→5	5→4	4→3	3→2	2→1	-
Running time (up) [s]	104	164	150	140	102	105	-

Station index	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Dwell time [s]	45	45	35	30	30	30	30	30	30	35	30	30	30	45

Table 2: Comparison between the train schedule used in practice and the train schedule obtained by the integrated train scheduling and circulation approach for the Beijing Yizhuang line

Property	Train schedule (used in practice)	Train schedule (integrated & optimized)
Number of required trains	10	9
Number of train services	108	102
Maximum cycles of a train	13	13
Maximum load factor	0.98	0.75