A novel two-stage approach for energy-efficient timetabling for an urban rail transit network

Transportation Research Part E 176 (2023) 103212

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

A novel two-stage approach for energy-efficient timetabling for an urban rail transit network

Kang Huang [a,b], Feixiong Liao [b,*]

[a] School of Systems Science, Beijing Jiaotong University, Beijing, China
[b] Urban Planning and Transportation Group, Eindhoven University of Technology, Eindhoven, The Netherlands

ARTICLE INFO

Keywords:
Energy-efficient timetabling
Path choice behavior
Tree knapsack problem
Urban rail transit

ABSTRACT

Urban rail transit (URT) is the backbone transport mode in metropolitan areas to accommodate large travel demands. The high energy consumption of URT becomes a hotspot problem due to the ever-increasing operation mileages and pressing agendas of carbon neutralization. The high model complexity and inconsistency in the objectives of minimizing passenger travel time and operational energy consumption are the main challenges for energy-efficient timetabling for a URT network with multiple interlinked lines. This study proposes a general model framework of timetabling and passenger path choice in a URT network to minimize energy consumption under passenger travel time constraints. To obtain satisfactory energy-efficient nonuniform timetables, we suggest a novel model reformulation as a tree knapsack problem to determine train running times by a pseudo-polynomial dynamic programming algorithm in the first stage. Furthermore, a heuristic sequencing method is developed to determine nonuniform headways and dwell times in the second stage. The suggested model framework and solution algorithm are tested using a real-world URT network, and the results show that energy consumption can be considerably reduced given certain travel time increments.

1. Introduction

With the high level of capacity, safety, punctuality, and environmental friendliness, urban rail transit (URT) becomes a popular transport mode for addressing urban mobility issues in metropolises (Yang et al., 2020). The total operating mileage of the URT worldwide has been increasing steadily over 30,000 km by the end of 2022. Although URT is more oriented for low carbon emissions for serving the same passenger demand, energy efficiency has become a significant issue for the carbon neutralization initiative. Many studies have addressed the energy efficiency in URT systems in recent years. One traditional research line considers only one train running on a track between two stations for improving the driving strategy and finding the optimal speed profiles to reduce energy consumption (Howlett, 1996; Albrecht et al., 2013; Scheepmaker et al., 2017). Another research line concerns the scheduling of multiple trains of one URT line from the perspective of energy-efficient timetabling (Wang et al., 2021; Zhan et al., 2022; Zhou et al., 2023). Energy-efficient timetabling optimizes timetables of arrival times, departure times, running times, and dwell times of the train fleet to serve the passenger demand with the lowest energy consumption. Yet, at a higher level that receives little attention in the literature, the energy efficiency of a URT network (Huang et al., 2021) extends the energy-efficient timetabling from a single line to multiple interlinked lines involving the interactions between timetabling and passenger path choice. Below we provide a brief review

* Corresponding author.
E-mail address: f.liao@tue.nl (F. Liao).

https://doi.org/10.1016/j.tre.2023.103212
Received 5 February 2023; Received in revised form 1 June 2023; Accepted 14 June 2023
Available online 27 June 2023

of the relevant energy efficiency studies at the three levels.

The one-train energy efficiency models derive the driving strategies on a track on the basis of optimal control theory (Khmelnitsky, 2000; Albrecht et al., 2016a, 2016b; Lai et al., 2020). The optimal control for speed profile covers three phases, i.e., acceleration, coasting, and braking, and the key is to find the optimal switching points of the speed profile (Howlett and Pudney, 1998). Howlett et al. (2009) concluded that the optimal speed profile switching points are the solution to an intrinsic local energy consumption function. Using artificial intelligence and massive monitoring data, some research utilized data-driven methods to adapt the train speed profiles and minimize energy consumption. For instance, support vector machines and random forests were applied for estimating the energy consumption of speed profiles by Huang et al. (2019), and deep neural networks were taken as a sub-module for optimizing the speed profiles by Yin et al. (2020). Based on the recurrent neural network, Kuppusamy et al. (2020) incorporated the algorithm into the long short-term memory architecture to optimize speed profiles given a fixed running time. To improve the cooperative control of traction and regenerative energy, multi-agent reinforcement learning was suggested by Su et al. (2021) to optimize the driving strategy in a discrete time interval. Focusing on the energy efficiency of only one train, these studies did not attend to the interdependencies between multiple trains.

As an effective means to synchronize the trains for saving energy, energy-efficient timetabling of multiple trains running on a URT line has been increasingly discussed recently (Yang et al., 2017; Kang et al., 2020; Li et al., 2021). Representatively, Li and Lo (2014a, 2014b) combined the optimization of train control and schedule to coordinate train braking and acceleration to utilize the regenerative energy. Considering stop-skipping patterns, Yang et al. (2019) proposed a convex quadratic model formulation to determine the timetable and speed profile at the same time. To consider time window constraints for train arrivals and departures, Wang and Goverde (2017, 2019) incorporated a multi-train trajectory with the objectives of optimizing the total energy consumption and eliminating conflicts between trains. External factors, (e.g., line conditions and rolling stocks) are also important for the total energy consumption. Mo et al. (2019a, 2019b) suggested a double-layer model to determine the plan of the rolling stock at the first layer and then improve the overlapping time between braking traction at the same power supply section stations to obtain more regenerative energy. Similarly, Yang et al. (2020) integrated passenger assignment into a timetabling model and applied a non-dominated sorting genetic algorithm to find non-dominated solutions for sufficiently utilizing regenerative energy to reduce energy consumption. Considering the flow conservation theory and inserting the speed profile into the space–time network, Xu et al. (2020) developed a two-step method by first identifying energy-efficient train trajectory templates and then searching for the least generalized cost path to obtain the optimal timetable. Combining timetabling and stop plans, Xie et al. (2021) presented triple objectives to minimize the energy consumption, the probability of train delays, and the total train trip times in the model. In the event of power interruptions, Yang et al. (2022) proposed many-to-many energy allocation mechanisms to improve regenerative energy and minimize potential power interruption. As seen, energy efficiency in these studies is unexceptionally limited to a single line without the consideration of passenger path choices.

It is recognized that timetabling for a single line is an NP-hard problem (Cai and Goh, 1994; Qi et al., 2018). After incorporating energy efficiency, the timetabling problem has a higher complexity. Ignoring passenger path choice and fixing passenger arrival rates, a few studies (e.g., Yin et al., 2017; Lv et al., 2019; Mo et al., 2021) proposed mixed-integer linear programming (MILP) in a URT line of fewer than 20 stations. Some studies (e.g., Liu et al., 2020; Qu et al., 2020; Yang et al., 2019) presented non-linear formulations and resorted to metaheuristic solution algorithms. When path choice and transfer are incorporated, the energy-efficient timetabling for a URT network of multiple interlinked lines has an even higher complexity than that for a single URT line. Very few studies addressed the energy-efficient timetabling at such a challenging level. For instance, to optimize energy consumption and transfer connections for the last train, i.e., one train per URT line, Wang et al. (2022) presented a bi-objective model for the last train timetabling problem in a URT network. However, their focus of attention was the transfer accessibility of the last trains given the application context. As modern URT systems often have multiple trains running on multiple URT lines, Canca et al. (2018) developed an energy-efficient model to determine the frequencies and uniform timetable in a URT network given passenger arrivals and stochastic assignment of passenger paths. However, the uniform passenger arrivals and headways are not applicable in the period with time-invariant passenger demand without the time index for the passenger path. Huang et al. (2021) presented a bi-level model, in which the upper level is responsible for energy-efficient timetabling while the lower level for passenger path choice upon the timetable generated from the upper level. Their bi-level model has a few limitations in modeling and application contexts. First, the trade-off between energy consumption and passenger travel time was overlooked, which would result in unrealistic operations that sacrifice passenger travel time for energy consumption reduction. Second, the timetable was optimized with time-invariant passenger demands and the same headways for each URT line. The timetable with uniform headways fits the context in the peak hours that are associated with steady and high passenger demands. The timetable with non-uniform headways can accommodate uneven passenger demands typically in non-peak hours. Therefore, we argue that it is more demanding for energy-efficient timetabling during non-peak hours, while mobility-efficiency is more important in peak hours. Third, the suggested heuristic solution algorithm found acceptable timetable solutions for computational considerations but ignored the domain knowledge that could generate insights into methodological advancement.

In view of the limitations of Huang et al. (2021), this paper develops a general model framework for energy-efficient timetabling for a URT network of multiple interconnected lines. The framework considers a nonuniform timetable structure, and the time horizon of timetabling focuses on non-peak hours, during which energy efficiency is a prioritized objective as opposed to mobility-efficiency as a prioritized objective during peak hours. Accordingly, different behavioral mechanisms are incorporated to capture passenger path choice realistically in a URT network. Based on the explored relationships between train energy consumption and travel time components, we transform the total passenger travel time constraint into the total running time constraint and then reformulate the original problem as a tree knapsack problem (TKP). We suggest a novel two-stage approach of pseudo-polynomial dynamic programming (DP) and heuristics based on domain knowledge to find a satisfactory timetable solution iteratively. The suggested model framework and solution algorithm are validated in a real-world URT network. The results show that the two-stage approach can

achieve considerable energy efficiency while respecting the trade-off with passenger travel times.

The rest of this paper is structured as follows. Section 2 describes the energy-efficient timetabling problem for a URT network considering time-dependent passenger demands and path choice. Section 3 presents the modeling of energy-efficient timetabling. Section 4 designs the two-stage solution algorithm based on the proposed model framework. Section 5 verifies the solution algorithm. Lastly, conclusions of this study and plans for future work are made in Section 6.

2. Problem description

In a URT network, a train runs from one terminal station to the other terminal station in the up or down direction of a line and does not cross lines. The essential difference between timetabling for single-line and multi-line resides in the consideration of transfer and path choice. When the network has a well-designed topological structure, passengers may have more than one path to reach their destinations. Several factors affect passenger path choice behavior, such as travel time, ticket price, the convenience of transfer, and heterogenous travel preferences. Given the speed profile on a track, the total weight of a train, including empty train mass and passenger loading weight, determines the energy consumption. Since passenger loading weight is the outcome of path choice, energy-efficient timetabling for a URT network should incorporate passenger path choice behavior. Hence, timetabling and passenger path choice are two key aspects of energy-efficient timetabling for a URT network.

To show the timetable elements of a URT network, suppose there are N_l and $N_{l'}$ stations, $2N_l$ and $2N_{l'}$ platforms, and $2N_l-2$ and $2N_{l'}-2$ tracks on lines l and l', respectively. For an arbitrary example, station 3 of line l and station 4 of line l' denote the same station, then passengers can make transfers at platforms 3 and $2N_l-2$ of line l and platforms 4 and $2N_{l'}-3$ of line l'. In a URT network of more lines, passengers choose their path and spend time on different path components (e.g., waiting, in-train, and transfer) to reach the destination platform. Fig. 1 shows a simplified URT network of four lines, in which there are two paths for the specified pair of origin and destination (OD). Passengers choose different paths going through different transfer stations and take different trains in Fig. 2, which is a space–time network representation (Liao, 2016, 2019) of the simple URT network. In each orange dotted box, there is an operation line, and the platforms with the same label represent a transfer station. These two paths in Fig. 1 may have different travel times and the trains have different passenger loadings in Fig. 2.

To clearly describe path components, path r of an OD pair can be defined by a series of sub-trips separated by transfers. For instance, the path is expressed as a specific combination of tracks or platforms.

respectively as $r = \oplus\{(l,t_0)\cdots(l_n,t_1)\}_\Lambda$ or $r = \oplus\{(l,p_0)\cdots(l,p_1)\}_\Lambda$, where Λ is the index of a sub-trip and Λ_r is the set of sub-trips of path r whose cardinality is equal to the number of transfers plus one; $\Lambda \in \Lambda_r$; t_0 and t_1 are the first and last tracks of the sub-trip Λ, while p_0 and p_1 are the first and last platforms, and those in-between are intermediary tracks and platforms of the sub-trip Λ, respectively; $\oplus$ is an operator interlinking sub-trips. If there is more than one sub-trip in r, two neighboring sub-trips have platforms sharing a common transfer station. Note that in a well-designed URT network, URT trips usually include less than three times of transfer, i.e., no more than three sub-trips in a path.

On a track, a train normally runs through the stages of acceleration, coasting, and deceleration subject to constraints of operational conditions. The speed profiles of different operation levels are often pre-defined in the automatic train operation system (ATO). Holding other operation conditions (e.g., train fleet and type) unchanged, the operation level of speed profile and passenger loading determines the running time and energy consumption. As stated above, passenger loading depends on passenger path choices. In light of the interactive process of timetabling and passenger path choice, we present a general model framework of energy-efficient timetabling for a URT network in Fig. 3. First, given the time-dependent passenger demands, an initial passenger loading in the URT network is performed given a path choice behavior mechanism and a timetable. Usually, the currently used timetable is utilized in the initial assignment for the sake of comparison with the optimized timetable solution. Then, taking the passenger loading (i.e., passenger volumes in the network) as input, the timetabling optimization reaches energy efficiency subject to constraints of the URT

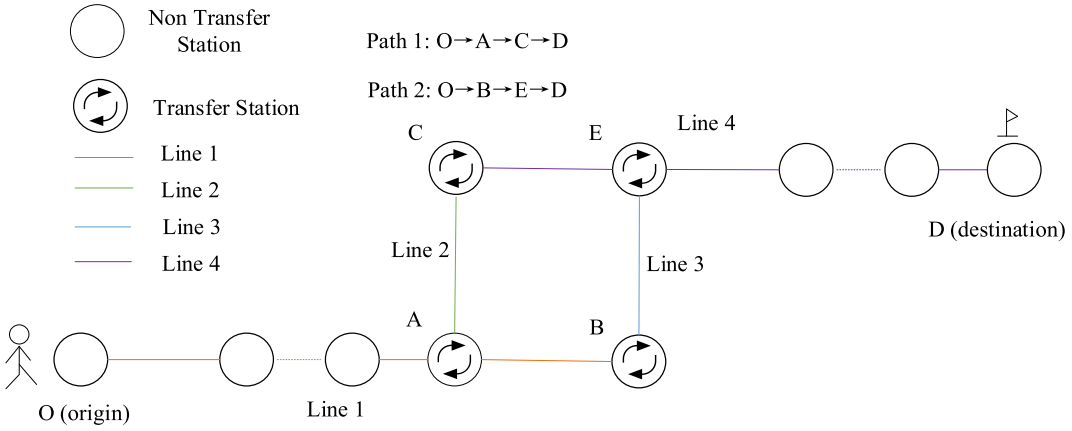


Fig. 1. Illustration of passenger path choice.

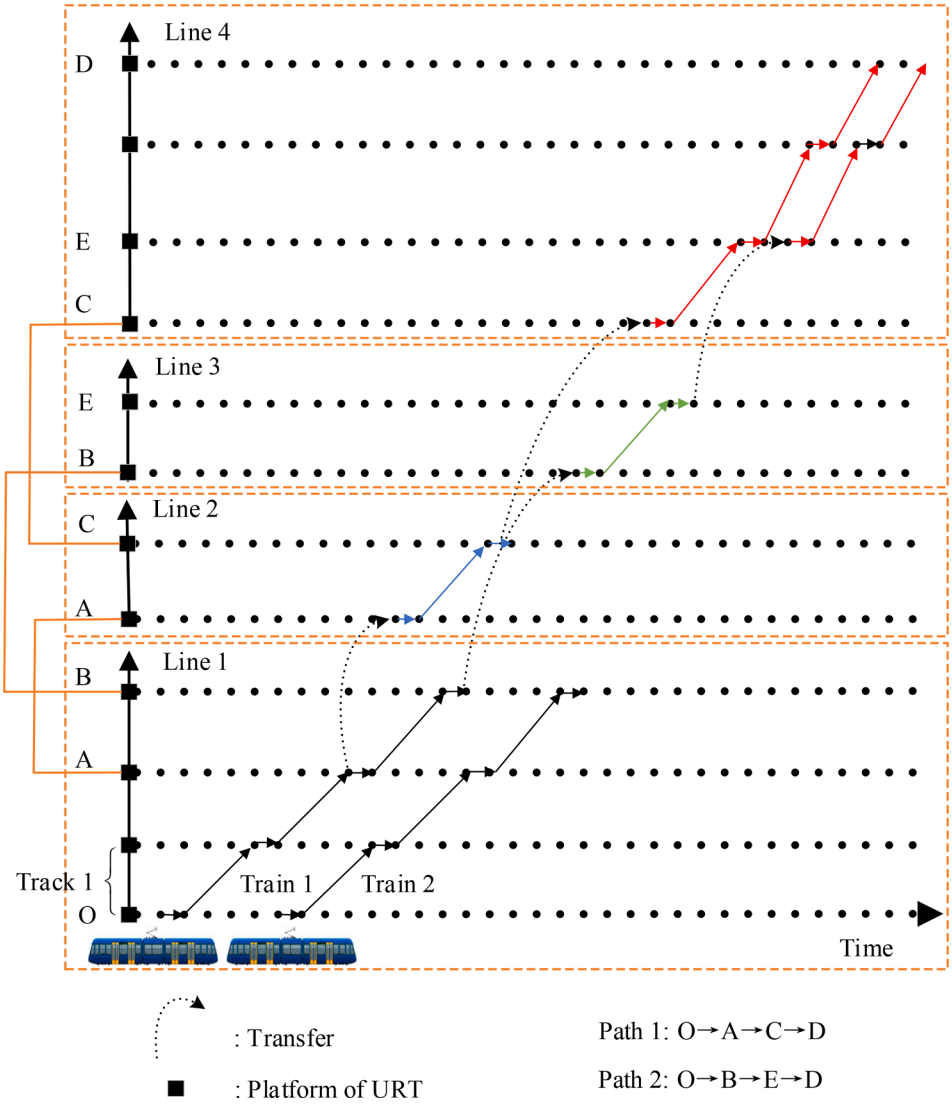


Fig. 2. Components of passenger space–time path.

characteristics and passenger travel times. Next, passenger flow adjustment is considered to reflect the passenger travel choice responding to the produced timetable. Finally, an energy-efficient timetabling solution is obtained when the iterative process converges. In short, energy-efficient timetabling determine a timetable with specified time components and operation levels that minimize energy consumption considering passenger path choices.

3. Modeling

In this section, we present the modeling of energy-efficient timetabling of a URT network, including primary assumptions, notations, and formulations of passenger travel time and train energy consumption.

3.1. Assumptions and primary notations

The model development is based on the following five assumptions. The explanations and motivation for the assumptions are also provided.

Assumption 1. The energy-efficient timetabling problem concerns a planning period during the non-peak hours, in which the trains of the URT are scheduled parallelly and nonuniformly (Shang et al., 2018). The time-dependent passenger arrival during a time interval is given by the statistical Automatic Fare Collection System (AFC) data (Yin et al., 2020). After arriving at the departure platform, a passenger can always board the first incoming train during non-peak hours.

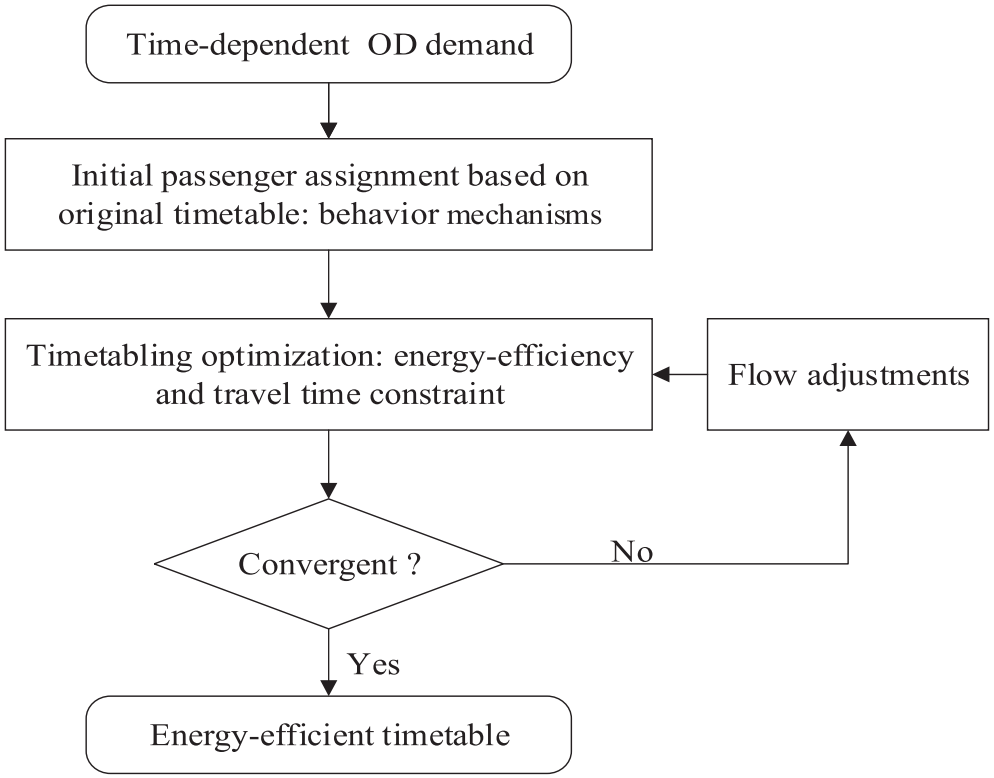


Fig. 3. A general model framework of energy-efficient timetabling for a URT network.

Assumption 2. Transfer time in a URT trip includes walking time and waiting time. Waiting time is determined by the timetable while walking time is unchanged as the average depending on the size of the transfer station (Chen et al., 2019).

Assumption 3. The URT operator seeks to minimize energy consumption with the consideration of passenger path choice behavior and travel time (Yang et al., 2020).

Assumption 4. For any OD pair, passengers have limited path choice alternatives in the URT network, where cyclic or overlong paths are excluded (Berggren et al., 2021).

Assumption 5. Given the running time on a track, the energy consumption of a train is in a linear relationship with the passenger weights based on the energy consumption of an empty train.

For Assumption 1, parallel timetables are commonly applied for safety considerations and operational convenience in a URT network (Zhang et al., 2019). A nonuniform timetable has fluctuated headways depending on passenger arrival rates. The delay due to waiting for the next incoming trains may occur during peak hours, in which the URT operator often pursues the maximum transport capacity and thus chooses a uniform and minimum headway (Parkinson and Fisher, 1996). Therefore, Assumption 1 is reasonable during the non-peak hours, in which energy-efficient timetabling matters for the URT operator. Assumptions 2–3 are commonly adopted in the timetabling literature. Path travel time does not consider the possible delay caused by overcrowding at the platforms. Assumption 4 is made in line with the broad travel behavior literature stating that individuals usually consider only a few path alternatives.

For Assumption 5, given running time v and based on the train traction dynamics, the train energy consumption on a track shows a linear relationship with the total train weight m, represented as $e(m, v)$ (Canca and Zarzo, 2017). To precisely calculate the total energy consumption, the passenger loadings on different trains and tracks should be distinguished, which causes a high model complexity. We show below that Assumption 5 does not only lead to negligible calculation errors but also simplifies the model complexity. Based on the real data from the Xi'an URT, for a track with a length of 1.5 km (the average track length in most URT networks) and a full capacity of passenger loading (1468 passengers), the average relative differences in energy consumption using a linear function with the passenger weights are less than 4% (see Section S1 in the supplementary document). Suppose two scenarios of passenger loadings in one hour of 20 trains running over a track: (1) all passengers are on the last 10 trains and all other trains are empty; (2) passengers are evenly loaded on the 20 trains. The error for calculating the total energy consumption caused by Assumption 5 is less than 2%. When the train energy consumption is linear with passenger loadings (Assumption 5), it is sufficient to obtain the total energy consumption by calculating the total weights on a track. For example, given the same speed profile, passengers who go through track 1 either in train 1 or 2 in Fig. 2 have the same effects on the total energy consumption during the planning period. Hence, the train index can be ignored

in passenger loading. The linearized treatment in Assumption 5 may slightly overestimate the real total energy consumption but can efficiently reduce the model complexity.

The primary notations of the model framework are listed below.

Parameters:	
P	set of platforms
$\widetilde{P}$	set of platforms belonging to transfer stations
l, L	index of line and set of lines
t, T_l	index of track and set of tracks on line l
p, P_l	index of the platform and set of platforms on line l
k	index of train
K_l	train set of line l
K_l^-	train set of line l excluding the last train
u, U	index of time intervals and set of time intervals in the planning period
r	index of passenger path
$\bar{r}, \underline{r}$	origin and destination stations of path r
$\bar{r}_1, \bar{r}_2$	starting line and platform of path r
R	set of alternative paths in the URT network,$r \in R$
$q_{\bar{r}\underline{r}u}$	passenger demand of OD $\bar{r}\underline{r}$ at time u
$[w^{\mathrm{O}}, w^{\mathrm{E}}]$	feasible time range for the dwell time
$[h^{\mathrm{O}}, h^{\mathrm{E}}]$	feasible time range for the headways
$\mathrm{s}_1, \mathrm{s}_2$	specific combinations of train, platform, and line
s	concatenation of platforms and lines for transfer arc $s = \langle \mathrm{s}_1, \mathrm{s}_2 \rangle = \langle (l,p), (l',p') \rangle,\ p, p' \in \widetilde{P}$
S	set of concatenations of platforms and lines for transfer arcs
s_{zr}	the z-th concatenation in path r
λ_{rlp}	0–1 parameter: $\lambda_{rlp} = 1$ if platform p of line l belongs to path r; otherwise,$\lambda_{rlp} = 0$
Decision variables:	
a_{lkp}	arrival time of train k at platform p of line l
v_{lp}	running time from platform p to the next platform $(\bar{p})$ of line l in one direction
d_{lkp}	departure time of train k at platform p of line l
w_{lp}	dwell time at platform p of line l
h_{lk}	headway between train k and the next train $(\bar{k})$ of line l
Intermediate variables:	
η_{lur}	probability that path r is selected by time-dependent passengers of OD $\bar{r}\underline{r}$ at time u
q_{ur}	passenger volume of OD $\bar{r}\underline{r}$ choosing path r at time u
c_{ur}	travel time of path r for the passenger that arrives at $\bar{r}$ at time u
ξ_{urk}	0–1 variable: $\xi_{urk} = 1$ if passengers arriving at u for r can board train k; otherwise, $\xi_{urk} = 0$
$\gamma_{skk'}$	0–1 variable: $\gamma_{skk'} = 1$ if transfer synchronization occurs between train k and train k' at s; otherwise, $\gamma_{skk'} = 0$
q_{lt}	passenger volume in all trains on track t of line l
m_{lt}	passenger weight in all trains on track t of line l
E_l^1	energy consumption by an empty train of line l
E_l^2	energy consumption by passenger weight of line l

3.2. Passenger path travel time

For a parallel and nonuniform timetable, the trains on the same URT line are scheduled cyclically with different headways. Denote the headway of train k on line l ($l \in L$) by h_{lk}; thus, h_{lk} may differ across k ($k \in K_l$). However, for any k, the dwell time w_{lp} at platform p of line l and the running time v_{lp} between platform p ($p \in P_l$) and the next platform $(\bar{p})$ are uniform. Therefore, there is no need to attach train index k to v_{lp} and w_{lp}. The definitional relationship among arrival time a_{lkp}, dwell time w_{lp}, and departure time d_{lkp} is given in Eq. (1). To avoid delay or over-waiting at the platforms for alighting and boarding, dwell time is constrained to a time range $[w^{\mathrm{O}}, w^{\mathrm{E}}]$ in Eq. (2). For platform p on line l, the relationship among departure time, running time, and arrival time is described as Eq. (3). The train headways between train $k (k \in K_l^-)$ and its next train $\bar{k}$ at the same platform in a parallel timetable are described by Eq. (4). Similarly, there is also a feasible range $[h^{\mathrm{O}}, h^{\mathrm{E}}]$ in Eq. (5) for operational requirements (e.g., related to the train signaling system and required service level).

$$a_{lkp} + w_{lp} = d_{lkp}, \qquad \forall p \in P_l, k \in K_l, \forall l \in L \tag{1}$$

$$w^{\mathrm{O}} \le w_{lp} \le w^{\mathrm{E}}, \qquad \forall p \in P_l, \forall\, l \in L \tag{2}$$

$$d_{lkp} + v_{lp} = a_{lk\bar{p}}, \qquad \forall p \in P_l, k \in K_l, \forall l \in L \tag{3}$$

$$d_{lkp} + h_{lk} = d_{l\bar{k}p}, \qquad \forall p \in P_l, k \in\ k \in\ K_l^-, \forall l \in L \tag{4}$$

$$h^{\mathrm{O}} \le h_{lk} \le h^{\mathrm{E}}, \qquad \forall k \in K_l, \forall l \in L \tag{5}$$

where K_l is the train set on line l, P_l is the set of platforms on line l, and K_l^- is the train set of line l excluding the last train in the planning period.

The travel time of a path may involve three parts, i.e., waiting before the first platforms, transfer, and in-train (including running time on the tracks and dwell time at the platforms). Whether a passenger arriving at time u and selecting path r can board train k at platform p of line l depends on the relationship between u and train departure time d_{lkp}. If u falls between the departure times of train k and its previous train $\underline{k}$, passengers can board train k. Therefore, a 0–1 variable ξ_{urk} is introduced to denote if a passenger can board a train in Eq. (6) where $\bar{r}_1$ and $\bar{r}_2$ denote the starting line and platform of path r for convenience, respectively. As stated in Assumption 1, the passenger boards the first incoming train during non-peak hours. To unify the expression of ξ_{urk}, we introduce a virtual departure train, $k = 0$ with its departure time fixed as 0 ($d_{l0p} = 0$). Eq. (6) can be linearized as Eq. (7) with auxiliary 0–1 variables ξ^1_{urk} and ξ^2_{urk}, where M is a large number.

$$\xi_{urk} = \begin{cases} 1, & \text{if } d_{l\underline{k}p} < u \le d_{lkp} \\ 0, & \text{otherwise} \end{cases} \quad ,\forall l = \bar{r}_1, p = \bar{r}_2, k \in K_l \tag{6}$$

$$\begin{cases} d_{lkp} - u \ge \left(\xi^1_{urk} - 1\right)M \\ d_{lkp} - u < \xi^1_{urk} M \\ u - d_{l\underline{k}p} > \left(\xi^2_{urk} - 1\right)M \\ u - d_{l\underline{k}p} < \xi^2_{urk} M \\ \xi_{urk} \le \xi^1_{urk}; \xi_{urk} \le \xi^2_{urk} \\ \xi_{urk} \ge \xi^1_{urk} + \xi^2_{urk} - 1 \end{cases} \quad ,\forall l = \bar{r}_1, p = \bar{r}_2, k \in K_l \tag{7}$$

The waiting time at the starting platform of path r for a passenger arriving at time u, c^1_{ur}, is calculated by Eq. (8).

$$c^1_{ur} = \sum_{k \in K_{\bar{r}_1}} \xi_{urk}(d_{lkp} - u), \forall p = \bar{r}_2 \tag{8}$$

where $K_{\bar{r}_1}$ is the train set of line $\bar{r}_1$.

Eq. (8) consists of the product of binary variable ξ_{urk} and $\left(d_{lkp} - u\right)$. To linearize Eq. (8), a new variable φ_{urk} is introduced: $\varphi_{urk} = \left(d_{lkp} - u\right)$ if $\xi_{urk} = 1$; otherwise, $\varphi_{urk} = 0$. φ_{urk} imposes the reformulations as Eq. (9) for $\forall r \in R, u \in U, l = \bar{r}_1, p = \bar{r}_2, k \in K_l$. c^1_{ur} could be re-formatted in Eq. (10).

$$\begin{cases} \left(d_{lkp} - u\right) - M(1 - \xi_{urk}) \le \varphi_{urk} \le \left(d_{lkp} - u\right) + M(1 - \xi_{urk}) \\ \varphi_{urk} \le \xi_{urk} M \\ \varphi_{urk} \ge 0 \end{cases} \tag{9}$$

$$c^1_{ur} = \sum_{k \in K_{\bar{r}_1}} \varphi_{urk} \tag{10}$$

The in-train time of a path is composed of train running and dwell times across all tracks and at all associated platforms, respectively. For passengers that arrive at time u, in-train time of path r, c^2_{ur}, is formulated as

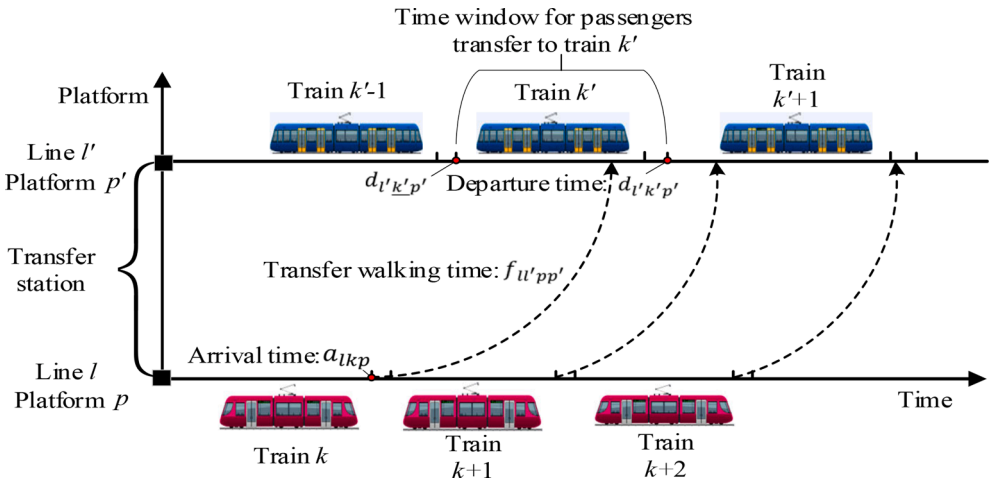


Fig. 4. Demonstration of time windows for transfer synchronization.

$$c_{ur}^{2} = \sum_{\forall l \in L, p \in P_l} \lambda_{rlp}(v_{lp} + w_{\overline{lp}}) \tag{11}$$

where λ_{rlp} is an incidence variable: $\lambda_{rlp} = 1$ if platform p of line l belongs to path r; otherwise, $\lambda_{rlp} = 0$.

The notion of transfer synchronization is used to describe a movement chain that passengers transfer from the alighting train to the connecting transfer platform and thereafter board the first incoming train. As shown in Fig. 4, the time window for passengers transferring from train k at platform p of line l to train k' at platform p' of line l' should satisfy $d_{\underline{lk}p'} \le a_{lkp} + f_{ll'pp'} \le d_{lk'p'}$ to reach synchronization, where $f_{ll'pp'}$ is the walking time during the transfer. For passengers arriving at time u, we use a 0–1 variable $\gamma_{skk'} = 1$ if transfer synchronization occurs between train k and k' at the concatenation of platforms and lines for transfer arc s; otherwise, $\gamma_{skk'} = 0$, where $\gamma_{skk'}^{1}$ and $\gamma_{skk'}^{2}$ are auxiliary 0–1 variables. One concatenation of platforms and lines for transfer arc s is constructed as Eq. (12).

$$s = \langle s_1, s_2 \rangle = \langle (l,p), (l',p') \rangle, p, p' \in \widetilde{P} \tag{12}$$

$\gamma_{skk'}$ can be expressed as

$$\gamma_{skk'} = \begin{cases} 1, & \text{if } d_{l'\underline{k}'p'} \le a_{lkp} + f_{ll'pp'} \le d_{l'k'p'} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Eq. (13) can be linearized by a couple of constraints:

$$\begin{cases} d_{l'k'p'} - a_{lkp} - f_{ll'pp'} \ge (\gamma_{skk'}^{1} - 1)M \\ d_{l'k'p'} - a_{lkp} - f_{ll'pp'} < \gamma_{skk'}^{1} M \\ a_{lkp} + f_{ll'pp'} - d_{l'\underline{k}'p'} \ge (\gamma_{skk'}^{2} - 1)M \\ a_{lkp} + f_{ll'pp'} - d_{l'\underline{k}'p'} < \gamma_{skk'}^{2} M \\ \gamma_{skk'} \le \gamma_{skk'}^{1}; \gamma_{skk'} \le \gamma_{skk'}^{2} \\ \gamma_{skk'} \ge \gamma_{skk'}^{1} + \gamma_{skk'}^{2} - 1 \end{cases}, \forall (l,p) = s_1, (l',p') = s_2, \langle s_1, s_2 \rangle = s \tag{14}$$

Moreover, there is one feasible space–time transfer arc for train k of a path at most in Fig. 4, which imposes a combination constraint as

$$\sum_{k' \in K_{l'}} \gamma_{skk'} \le 1, \forall s \in S, (l,p) \in s_1, (l',p') \in s_2, k \in K_l \tag{15}$$

To guarantee passengers arriving at time u can complete path r, the passenger should successfully board the trains $|\Lambda_r|$ times, imposing the following constraint

$$\xi_{urk} + \sum_{z \in Z_r (k_1 = k, k_z' = k_{z+1})} \gamma_{s_{zr} k_z k_z'} = |\Lambda_r| \tag{16}$$

where z and Z_r are index and the set of transfer of path r. $|\Lambda_r|$ is the number of sub-trips in path r. For path r, $|\Lambda_r| = |Z_r| + 1$. s_{zr} is the z-th concatenation of platforms and lines of transfer arc in path r. k_z and k_z' are the train indexes of the feeder line and the connecting line in the z-th transfer arcs respectively.

We introduce a new binary variable σ_{ur} to reformulate Eq. (16) as

$$\sigma_{ur} = \begin{cases} 1, & \text{if } \sum_{k \in K_{\overline{r}_1}} \left(\xi_{urk} + \sum_{z \in Z_r (k_1 = k, k_z' = k_{z+1})} \gamma_{s_{zr} k_z k_z'} \right) = |\Lambda_r| \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

Eq. (17) can be linearized as

$$M(\sigma_{ur} - 1) \le \sum_{k \in K_{\overline{r}_1}} \left(\xi_{urk} + \sum_{z \in Z_r (k_1 = k, k_z' = k_{z+1})} \gamma_{s_{zr} k_z k_z'} \right) - |\Lambda_r| \le \sigma_{ur} - \chi \tag{18}$$

where χ is a sufficiently small positive number.

The transfer time of path r for passengers arriving at time u, c_{ur}^{3}, can be represented as

$$c_{ur}^{3} = \sigma_{ur} \sum_{z \in Z_r} \left(d_{lk_z'p'} - a_{lk_zp} \right), z \in Z_r, k_1 \in K_{\overline{r}_1}, k_z' = k_{z+1} \in K_{s^z} \tag{19}$$

where s^z indicates the departure line in the z-th transfer arc.

Eq. (19) is a product of a binary variable σ_{ur} and variable $\sum_{z\in Z_r}(d_{lk'_zp'} - a_{lk_zp})$. To linearize the left part of Eq. (19), variable ψ_{ur} is introduced: $\psi_{ur} = \sum_{z\in Z_r}(d_{lk'_zp'} - a_{lk_zp})$, if $\sigma_{ur} = 1$; otherwise, $\psi_{ur} = 0$. Then, a set of constraints is represented in Eq. (20).

$$\begin{cases} \sum_{z\in Z_r}(d_{l'k'_zp'} - a_{lk_zp}) - M(1-\sigma_{ur}) \le \psi_{ur} \le \sum_{z\in Z_r}(d_{l'k'_zp'} - a_{lk_zp}) + M(1-\sigma_{ur}) \\ \psi_{ur} \le \sigma_{ur} M \\ \psi_{ur} \ge 0 \end{cases} \tag{20}$$

Combing the above, we can simplify the transfer time of path r as

$$c^3_{ur} = \psi_{ur} \tag{21}$$

The perceived travel time of a passenger on path r, c_{ur}, can be expressed as a function of the above three parts ($c^1_{ur}, c^2_{ur}, c^3_{ur}$) and passenger loading (q_{ur}) as

$$c_{ur} = C(c^1_{ur}, c^2_{ur}, c^3_{ur}, q_{ur}) \tag{22}$$

3.3. Passenger flow loading and travel time

We use η_{ur} to denote the probability that path r is chosen by passengers during time interval u for OD pair $\bar{r}\underline{r}$ ($\bar{r}, \underline{r}$: origin and destination of path r) following a path choice behavior. To capture the effects of path choice behaviors on energy-efficient timetabling, we consider three passenger loading mechanisms, namely, all-or-nothing assignment, linear proportional assignment, and one-off stochastic assignment (Liu et al., 2010). The all-or-nothing is widely applied in practice to non-congested transport networks (Hui, 2014). Linear proportional assignment is utilized for distributing passenger flow evenly (Yang et al., 2020). The one-off stochastic assignment takes into account the perceived travel time and captures non-linear effects. These three types of passenger loading mechanisms are valid alternatives in a non-crowded scenario during non-peak hours. In the all-or-nothing assignment, all the passengers choose the least travel time path. η_{ur} for OD pair $\bar{r}\underline{r}$ is described as Eq. (23.1).

$$\eta_{ur} = \begin{cases} 1, & \text{if } c_{ur} = \min\left\{c_{ur'} \middle| r' \in R_{\bar{r}\underline{r}}\right\} \\ 0, & \text{otherwise} \end{cases} \tag{23.1}$$

where $R_{\bar{r}\underline{r}}$ is the set of paths of OD pair $\bar{r}\underline{r}$.

In the linear proportional assignment, considering an imaginative reference of travel time $\overline{c_r}$ as the maximum possible travel time of OD pair $\bar{r}\underline{r}$, which may be 1.5–2 times the maximum travel time in a given timetable, η_{ur} is described in a linear function as Eq. (23.2).

$$\eta_{ur} = \frac{\overline{c_r} - c_{ur}}{\left|R_{\bar{r}\underline{r}}\middle|\overline{c_r} - \sum_{r'\in R_{\bar{r}\underline{r}}} c_{ur'}\right.} \tag{23.2}$$

where $|R_{\bar{r}\underline{r}}|$ is the number of paths of OD pair $\bar{r}\underline{r}$.

In the stochastic assignment, c_{ur} is stochastic due to perception error and thus represents the perceived travel time with a scaling parameter β. Assuming the error term follows an extreme value distribution, η_{ur} for a one-off stochastic assignment is formulated as

$$\eta_{ur} = \frac{e^{-\beta c_{ur}}}{\sum_{r'\in R_{\bar{r}\underline{r}}} e^{-\beta c_{ur'}}} \tag{23.3}$$

The above three assignment mechanisms involve only one-off passenger loadings. During non-peak hours, the passenger loadings on a track would not exceed the train capacity, which is in line with Assumption 1. However, with Eq. (23), passengers of a certain class may perceive longer travel time due to the low tolerance for crowdedness and thus adjust path choices in response to a generated timetable.

For passengers arriving at time u, the passenger volume choosing path r is equal to $q_{\bar{r}\underline{r}u} \bullet \eta_{ur}$, where $q_{\bar{r}\underline{r}u}$ is the passenger demand of OD $\bar{r}\underline{r}$ at time u. Considering the track and path incidences, the passenger volume of all trains on track t of line l can be calculated by Eq. (24), which determines the total energy consumption on track t involving all operating trains in the planning period.

$$q_{lt} = \sum_{\forall r\in R, u\in U} q_{\bar{r}\underline{r}u} \bullet \eta_{ur} \bullet \lambda_{rlt} \tag{24}$$

where λ_{rlt} is a binary incidence variable to indicate the relationship between paths and tracks, $\lambda_{rlt} = 1$ if track t of line l belongs to path r; otherwise, $\lambda_{rlt} = 0$.

Given an average passenger weight τ (e.g., Canca and Zarzo, 2017), the passenger load m_{lt} on track t is formulated as

$$m_{lt} = q_{lt} \bullet \tau \tag{25}$$

Given all path travel times and volumes, the total passenger travel times in the URT network, $\widetilde{C}$, can be calculated as

$$\widetilde{C} = \sum_{\forall r \in R, u \in U} q_{\overline{r}\underline{r}u} \bullet \eta_{ur} \bullet c_{ur} \tag{26}$$

where c_{ur}, in a special case, can be defined as the sum of the waiting time at the starting platform (c_{ur}^1), in-train time (c_{ur}^2), and the transfer time (c_{ur}^3) as

$$c_{ur} = c_{ur}^1 + c_{ur}^2 + c_{ur}^3 \tag{27}$$

From the operator's perspective, energy efficiency should be achieved within a tolerable increment in passenger travel times. Hence, a travel time constraint is added as

$$\widetilde{C} \le \overline{C} \bullet (1 + \Delta c) \tag{28}$$

where $\overline{C}$ is the total passenger travel times associated with the original timetable, and Δc is the allowed rate of travel time increment. From the modeling perspective, it is possible to consider passenger travel time and energy consumption in a bi-objective optimization model. In this paper, the objective of passenger travel time is transformed into the constraint, with which the relationship between passenger travel time and energy consumption can also be identified.

Remark 1. Given different time components in a URT trip, perceived travel time is adopted in Eqs. (22)–(23) and (26) to account for subjective travel time due to taste differences in the time components. The objective travel time formulated in Eq. (27) is a special case by equalizing the coefficients of the time components. Furtherly, the perceived travel time may be extended to perceived travel cost considering passenger heterogeneity and other factors in a URT trip, for example, transfer times, fare, and seat availability. With perceived travel cost, the passenger assignment mechanisms in Eq. (23) are still applicable.

3.4. Formulation of the energy consumption

We formulate the total energy consumption of all URT lines based on the relationships among passenger volume (weight), running time, and energy consumption. Typically, one level of speed profile in the ATO system corresponds to one specific running time on a track. A higher operation level of the speed profile is associated with a shorter running time and needs more energy consumption. A heavier train, given the speed profile level, involves more energy consumption. In other words, operation level g and train loading m_{lt} are the factors of train running time and energy consumption. Denote the base energy consumption of an empty train (weight m_l^0) at level g on track t by e_{ltg}^0. G_{lt} is provided as the set of operation levels on track t of line l. Then, a binary variable θ_{ltg} is introduced: $\theta_{ltg} = 1$ when level g is selected; otherwise, $\theta_{ltg} = 0$. For each track t, only one level is chosen, formulated in Eq. (29).

$$\sum_{g \in G_{lt}} \theta_{ltg} = 1, \ \forall l \in L, \forall t \in T_l \tag{29}$$

For the empty train of line l, the energy consumption, E_l^1, equals to

$$E_l^1 = |K_l| \bullet \sum_{t \in T_l} \sum_{g \in G_{lt}} \theta_{ltg} \bullet e_{ltg}^0 \tag{30}$$

With operation level g, the running time for track t of line l, v_{lt}, is re-specified as

$$v_{lt} = \sum_{g \in G_{lt}} \theta_{ltg} v_{ltg}^0, \ \forall l \in L, \forall t \in T_l \tag{31}$$

where v_{ltg}^0 is the running time on track t of line l at level g.

After passengers board the train, the total weight of the train is $m_l^0 + m_{lt}$. Therefore, the energy consumption equals $\left(1 + \frac{m_{lt}}{m_l^0}\right) \bullet e_{ltg}^0$ in a linear form. The energy consumption $e(m_{lt}, v_{lt})$ caused by the passenger volume can be formulated as

$$e(m_{lt}, v_{lt}) = \sum_{g \in G_{lt}} \theta_{ltg} \bullet \frac{m_{lt}}{m_l^0} \bullet e_{ltg}^0, \ \forall l \in L, \forall t \in T_l \tag{32}$$

Eq. (32) is also a product of binary variable θ_{ltg} and variable $\frac{m_{lt}}{m_l^0}$. To linearize it, variable ρ_{ltg} is created: $\rho_{ltg} = \frac{m_{lt}}{m_l^0}$ if $\theta_{ltg} = 1$; otherwise, $\rho_{ltg} = 0$. Then, a group of constraints is introduced as

$$\begin{cases} \frac{m_{lt}}{m_l^0} - M\left(1 - \theta_{ltg}\right) \le \rho_{ltg} \le \frac{m_{lt}}{m_l^0} + M\left(1 - \theta_{ltg}\right) \\ \rho_{ltg} \le \theta_{ltg} M \\ \rho_{ltg} \ge 0 \end{cases}, \forall l \in L, \ t \in T_l, \ g \in G_{lt} \tag{33}$$

Accordingly, $e(m_{lt}, v_{lt})$ can be replaced by $\sum_g \rho_{ltg} e_{ltg}^0, \ \forall g \in G_{lt}$. Therefore, the total energy caused by the passenger weight of line l, E_l^2, is presented as

$$E_l^2 = \sum_{t\in T_l}\sum_{g\in G_{lt}} \rho_{ltg} \bullet e^0_{ltg} \tag{34}$$

Combing the above, the total energy consumption of a URT network is expressed as

$$E_l = E_l^1 + E_l^2 = |K_l| \bullet \sum_{t\in T_l}\sum_{g\in G_{lt}} \theta_{ltg} \bullet e^0_{ltg} + \sum_{t\in T_l}\sum_{g\in G_{lt}} \rho_{ltg} \bullet e^0_{ltg} \tag{35}$$

Taken together, the objective function for energy-efficient timetabling and constraints are summarized as

$$\min E = \sum_{l\in L} E_l \tag{36}$$

s.t..Eqs.$(1)-(5),(7),(9)-(11),(14)-(21),(23-29),(31)-(33)$

The timetabling, i.e., the specification of train dwell times, arrival times, running times, and departure times, involves constraints Eqs. $(1)-(5),(31)-(33)$; the passenger assignment, i.e., determining the passenger volumes on every path or track, involves constraints Eqs. $(9)-(11),(14)-(21),(23)-(27)$. The two parts are coupled by path choices dictated by Eq. (23). The objective function and all the constraints are linearized by introducing auxiliary variables. Related to Assumption 5, the following remark is made.

Remark 2. Eqs. (32)–(35) are formulated in a linear relationship between energy consumption and train weights according to existing studies. However, with the complex real-world conditions, a nonlinear relationship should be applied. As an extension to capture the nonlinear relationship, the weights of different trains on the same track should be distinguished. Hence, the passenger volume on each train q_{klt} instead of q_{lt} should be calculated in Eq. (24), and the energy consumption should be formulated with $e(m_{klt}, v_{lt})$ instead of $e(m_{lt}, v_{lt})$ in Eq. (32). While it is technically feasible modeling-wise, the non-linear objective and constraints add considerable complexity to finding a satisfactory timetable solution.

Except for the passenger assignments, the linearization of the constraints by the Big *M* method and the introduction of auxiliary variables make the model in MILP formulations. Given the results of any passenger assignment, the exact energy-efficient timetable solution can be found by optimization solvers (e.g., Gurobi and Cplex) in theory. The scales of variables and constraints are provided in Table 1 to show the complexity of the proposed model framework. Compared with the parallel and uniform timetabling setup in Huang et al. (2021), there is an increased dimension of trains ($|K_l|$) for timetabling and an increased dimension of space–time paths ($|U|$) for passenger assignment. Also, there is an additional group of auxiliary variables $\gamma_{skk'}$ to respect the space–time constraints for transfer synchronization in the URT network. Comparatively, the energy-efficient timetabling and constraints formulated above have much higher complexity than their counterparts. It is impractical to find the exact optimal timetable even in a small URT network of two URT lines after adding the hard constraints related to passenger path choice and transfers. Therefore, it is advisable to devise an effective approximation algorithm for real-world applications.

Remark 3. In existing timetabling studies, the majority concern the operations of one single URT line and apply metaheuristic algorithms (e.g., PSO, GA) to address the nonuniform timetabling setup owing to the high complexity of this NP-hard problem, for which a polynomial-time algorithm does not exist. When the timetabling is extended to the network level with multiple interlinked lines, even in a simplified application context of one train per URT line (e.g., first-train or last-train problems), it is hard to find an exact timetable

Table 1
The scale of variables and constraints in the nonuniform and uniform timetable ($|\bullet|$ is the size of a set).

Variables or constraints	Scale	
	Uniform	Nonuniform
Arrival time, a_{lkp}; departure time,d_{lkp}	$\sum_{l\in L}\|P_l\|$	$\sum_{l\in L}\|P_l\|\|K_l\|$
Dwell time,w_{lp}	$\sum_{l\in L}\|P_l\|$	N/C
Running time, v_{lt}; loading weight, m_{lt}; passenger volume,q_{lt}	$\sum_{l\in L}\|T_l\|$	N/C
Headway,h_{llk}	$\|L\|$	$\sum_{l\in L}\|K_l\|$
Waiting time c^1_{ur}; in-train time c^2_{ur}; transfer time c^3_{ur}	$\|R\|$	$\|U\|\|R\|$
Intermediate binary variable:ξ_{urk}, φ_{urk}	N/A	$\|U\|\sum_{r\in R}\|K_{\bar{r}_1}\|$
Intermediate binary variable,$\gamma_{skk'}$	N/A	$\sum_{s\in S}\|K_{s_{1,1}}\|\|K_{s_{2,1}}\|$
Intermediate binary variable,θ_{ltg}	$\sum_{l\in L}\sum_{t\in T_l}\|G_{lt}\|$	N/C
Intermediate continuous variable,ρ_{ltg}	$\sum_{l\in L}\sum_{t\in T_l}\|G_{lt}\|$	N/C
Constraints (1), (3)-(4)	$\sum_{l\in L}\|P_l\|$	$\sum_{l\in L}\|P_l\|\|K_l\|$
Constraint (2)	$\sum_{l\in L}\|P_l\|$	N/C
Constraint (5)	$\|L\|$	$\sum_{l\in L}\|K_l\|$
Constraints (7), (9), (18)	N/A	$2\|U\|\sum_{r\in R}\|K_{\bar{r}_1}\|$
Constraints (8), (10), (11), (21)	$\|R\|$	$\|U\|\|R\|$
Constraint (14)	N/A	$3\sum_{s\in S}\|K_{s_{1,1}}\|\|K_{s_{2,1}}\|$
Constraint (20)	$3\|R\|$	$3\|U\|\|R\|$
Constraint (28)	$\sum_{l\in L}\sum_{t\in T_l}\|G_{lt}\|$	N/C
Constraint (30)	$\sum_{l\in L}\|T_l\|$	N/C
Constraint (32)	$3\sum_{l\in L}\sum_{t\in T_l}\|G_{lt}\|\|\Omega_l\|$	N/C

(N/C: Not changed; N/A: Not applicable).

solution (Kang et al., 2020), not to mention the application context of multi-train and multi-line raised in this study. Hence, an efficient approximation algorithm that can produce satisfactory timetable solutions is preferable.

4. Solution algorithm

This section presents an algorithmic framework enriching the timetable generation framework depicted in Fig. 3 to solve the formulated problem above. Considering the domain knowledge of energy consumption in a URT network, we decompose the algorithmic framework into three modules and suggest efficient solution methods.

4.1. Algorithmic framework

Responding to a timetable, the time-dependent passenger demands of a certain period adapt path choices that together result in a certain amount of energy consumption. The URT operator intends to reduce energy consumption by designing an energy-efficient timetable that in turn leads to new path choices. The interactive relationship between passenger path choice and timetabling is depicted in Fig. 5, in which the solution methods for different modules are also shown.

For module 1, the time-dependent OD demands are taken as input in the passenger assignment according to a certain behavioral mechanism (Eq. (23)). The passenger distribution in the URT network as a result of the passenger assignment is the input of timetabling optimization subject to various constraints. The travel time of a path may involve three parts, i.e., waiting at the first platform c_{ur}^1, in-train time c_{ur}^2, and transfer time c_{ur}^3. Typically, the dwell time at a platform is short and falls within a tight range (e.g., between 30 s and 60 s) in a URT network. Given the running time on each track, c_{ur}^2 is stable and fluctuates only in a narrow range. Moreover, with a moderate train fleet size in the planning period, the total waiting and transfer time is stable. Therefore, the total passenger travel time, $c_{ur}^1 + c_{ur}^2 + c_{ur}^3$, is stable and fluctuates in a narrow range, of which c_{ur}^2 takes a large proportion. Based on this domain knowledge, we make a conjecture that the total passenger travel time is stable once the fleet size and track running times (speed profiles) are given in the planning period.

In the second module, we design a two-stage approach for energy-efficient timetabling based on the conjecture. The calculation of the path travel times is transformed into track-based running times. Also, we replace the total passenger travel time constraint Eq. (28) by an approximate total running time constraint. In the first stage, given passenger distribution in the URT network, the running times and speed profiles on all tracks determine the total running time and train energy consumption. Specifically, the track running times are determined by adjusting the level of the speed profile successively from the level with the largest energy consumption and minimum running time to save energy while satisfying the travel time constraint. This problem can be reduced to a knapsack problem with

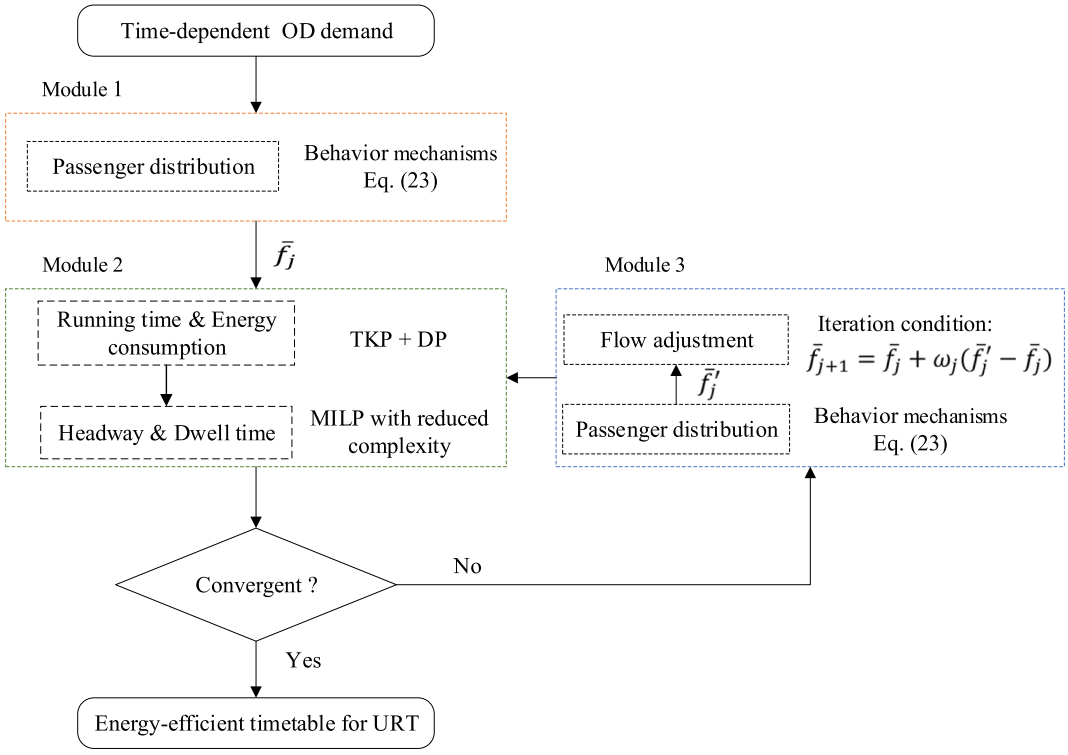


Fig. 5. Flowchart of the solution algorithm.

a tree structure, i.e., a tree knapsack problem. Applying the established DP method, we can find the optimal track running times and speed profiles on all tracks under the total running time constraint within pseudo-polynomial computation time. In the second stage, given the track running times derived from the first stage, a novel heuristic rule is developed for determining headways and dwell times. The remaining timetabling is a MILP problem with significantly reduced complexity.

In the third module, passengers may adjust their path choices given a new timetable produced from module 2. Passenger flow adjustments are incorporated and demonstrated to converge after an iterative adjustments process induced by the method of successive average (Chen et al., 2015). Passenger flows are adjusted according to a widely applied projection $\bar{f}_{j+1} = \bar{f}_j + \omega_j(\dot{\bar{f}}_j - \bar{f}_j)$, where $\bar{f}_j$ is the j-th iteration of passenger flow pattern, $\dot{\bar{f}}_j$ is the passenger flow pattern after adjustment in module 2, and $\omega_{j+1}\,(0 < \omega_j < 1)$ is the adjustment coefficient. The flow adjustments are subject to flow conservation. When j gets large, ω_j approaches 0 and $\bar{f}_{j+1} \approx \bar{f}_j$, which guarantees convergence.

Given that modules 1 and 3 are standard procedures, the following part focuses on the two-stage approach in module 2.

4.2. Reformulation as a tree knapsack problem

Based on the stipulated conjecture, we transform the total passenger path travel time constraint (28) into the total track running time constraint as

$$\sum_{\forall r\in R, u\in U} q_{\bar{r}\underline{r}u} \bullet \eta_{ur} \bullet v_{ur} \le \overline{V} \bullet (1 + \Delta c) \tag{37}$$

where $\overline{V}$ is the total passenger running times associated with the original timetable, v_{ur} is the total track running time of a passenger on path r, and $\overline{V} \bullet \Delta c$ represents the allowed running time increment.

Given the constraint of total running times, the objective is to select a set of dependent track running times that minimize energy consumption, where the speed profile and running time on each track are the decision variables. This problem resembles the knapsack problem in a tree structure, which is a classic combinatorial optimization problem (Martello and Toth, 1990). Given a capacitated knapsack and a set of items with certain weights (or sizes) and values, the objective is to determine which items to be included in the knapsack for achieving the maximum values.

For the train operations, between every two adjacent speed profile levels, there is an energy consumption reduction Δe_i from a high speed level to a lower speed level, which corresponds to a running time increment Δv_i. This relationship can be conceptualized as an item in the knapsack problem, where the value is the energy consumption reduction due to speed level adjustment and the weight (or size) is the running time increment. The knapsack capacity corresponds to the allowed total running time increment. On all tracks, the speed profile level adjustment starts from the levels with the largest energy consumption and minimum running time, and then successively between adjacent speed profile levels. Each speed profile level adjustment relies on the previous level and acquires energy consumption reduction while incurring running time increment. Hence, there is dependency among these speed profile level ad-

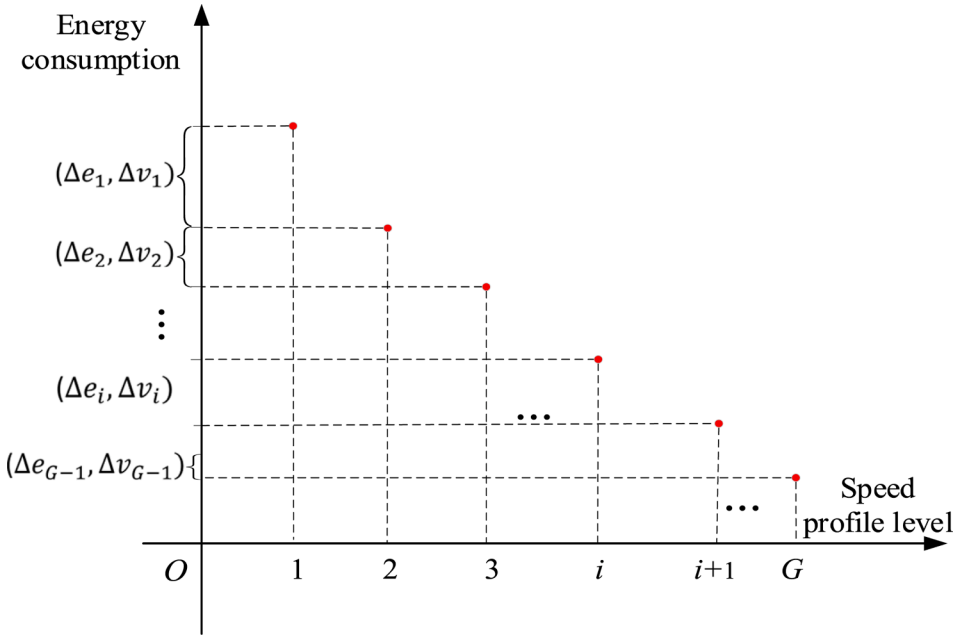


Fig. 6. Energy consumption changes over the adjacent speed profile level.

justments. As shown in Fig. 6, on each track, given a total of $|G|$ speed profile levels, $|G|-1$ items are generated. If item i is selected, item $i-1$ must be picked up first. For instance, if level 3 is selected, two items (i.e., level 1 is shifted to level 2, then level 2 is shifted to level 3) must be selected successively to gain an energy reduction of $\Delta e_1+\Delta e_2$ and incur a running time increment of $\Delta v_1+\Delta v_2$. The dependency can be seen as a relationship like one father node and one son node specified in a tree structure. To connect the nodes, we create a dummy root node in the tree, representing a dummy item with zero weight and zero value. As illustrated in Fig. 7, the dummy root node has $|T|$ son nodes, and every node except the last one on each track has only one son node. Based on these principles, the knapsack problem with a tree structure, TKP, can be modeled as a knapsack problem with groups of items as follows.

$$\text{Max} \sum \delta_{xy} \bullet \beta_{xy} \tag{38}$$

$$\sum \alpha_{xy} \bullet \delta_{xy} \le C_0 \tag{39}$$

$$\delta_{xy} \le \delta_{x(y-1)}, \forall x \in X, y \le |Y_x| \tag{40}$$

$$\delta_{xy} \in \{0,1\}, \forall x \in X, y \le |Y_x| \tag{41}$$

where x and X are the group index and set of groups; y is the index of the item and $Y=\{Y_x|x\in X\}$ is the item set in the group x; δ_{xy} is a binary variable: $\delta_{xy}=1$ indicating item y in group x is selected to reflect the speed profile level adjustment on a track, otherwise, $\delta_{xy}=0$; β_{xy} is the value of item y of group x representing energy consumption reduction for the level adjustments; α_{xy} is the weight of item y of group x representing running time increment. Objective (38) maximizes the total value of selected items. Constraint (39) represents the capacity constraint, where $C_0=\overline{V}\bullet\Delta c$ corresponds to the allowed total passenger travel time increment. For the dependency among the items in a group, when item y of group x is selected ($\delta_{xy}=1$), the previous item $(y-1)$ must have been picked up as stated by Constraint (40).

Since the standard knapsack problem is NP-hard (Caccetta and Kulanoot, 2001), TKP is also NP-hard. Considering integral item weights and capacity, the allowed running time increment can also be treated as integral using the time interval Δv_0 as one time unit in the model. We use a DP approach with pseudo-polynomial computation time to solve the TKP. The recursive formulation for the TKP is succinctly given as

$$dp(o,\mathscr{C})=\begin{cases}\max\{dp(o,\mathscr{C}),dp(o,\mathscr{C}-\mathscr{C}')+dp(son,\mathscr{C}')\}, & C\in[0,C_0-\alpha(o)],\ \mathscr{C}'\in[0,C]\\ dp(o,\mathscr{C}-\alpha(o))+\beta(o), & C\in[\alpha(o),C_0]\\ 0, & C\in[0,\alpha(o)]\end{cases} \tag{42}$$

where $dp(o,\mathscr{C})$ is the maximum value of the items given weight $\mathscr{C}$ and father node o; son denotes one son node item dependent node o; $\alpha(o)$ is the weight of node o representing running time increment; $\beta(o)$ is the value of node o representing energy consumption reduction; $\mathscr{C}'$ represents the remaining weight; in the recursive process, $\mathscr{C}$ decreases with a step of one unit while $\mathscr{C}'$ increases with a

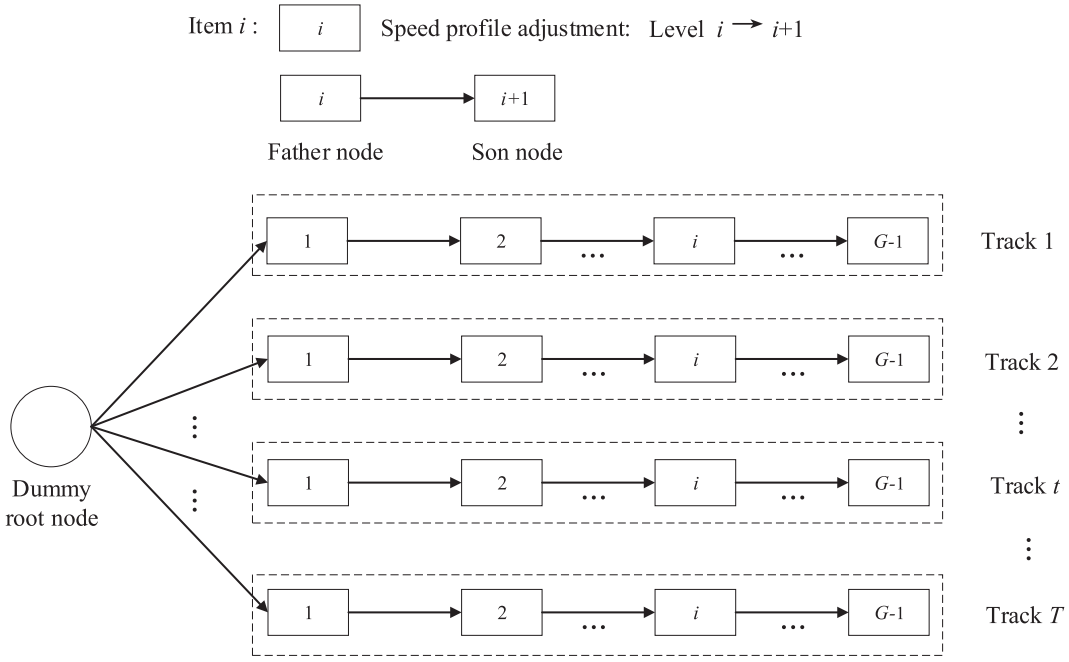


Fig. 7. Illustration for dependency in the TKP.

step of one unit. The running time complexity of this algorithm is $O(|T| \bullet (|G|-1) \bullet \lfloor \overline{V} \bullet \Delta c/\Delta v_0 \rfloor^2)$, where $|T|$ is the number of tracks in the URT network, $|G|$ is the number of speed profile levels, $(\lfloor \overline{V} \bullet \Delta c/\Delta v_0 \rfloor)$ is the allowed total increased running time in the unit of the specified time interval Δv_0, and $\lfloor \cdot \bullet \rfloor$ is an operator to round down a number. Finally, $dp(o_0, C_0)$ is the maximum value given capacity C_0 and root node o_0, corresponding to the maximum energy consumption reduction given an allowed running time increment. By backtracking the labeling process, we can find all the track running times and speed profiles (see Section S3 in the supplemental document for the source code).

Remark 4. The TKP exquisitely incorporates the interdependencies between train energy consumption reduction and allowed total running time increment. The scale in the temporal dimension, $\lfloor \overline{V} \bullet \Delta c/\Delta v_0 \rfloor$, can be adjusted with different lengths of Δv_0 to balance the computation efficiency and accuracy. With a lower resolution, the efficiency of this algorithm is higher, but the accuracy is lower,

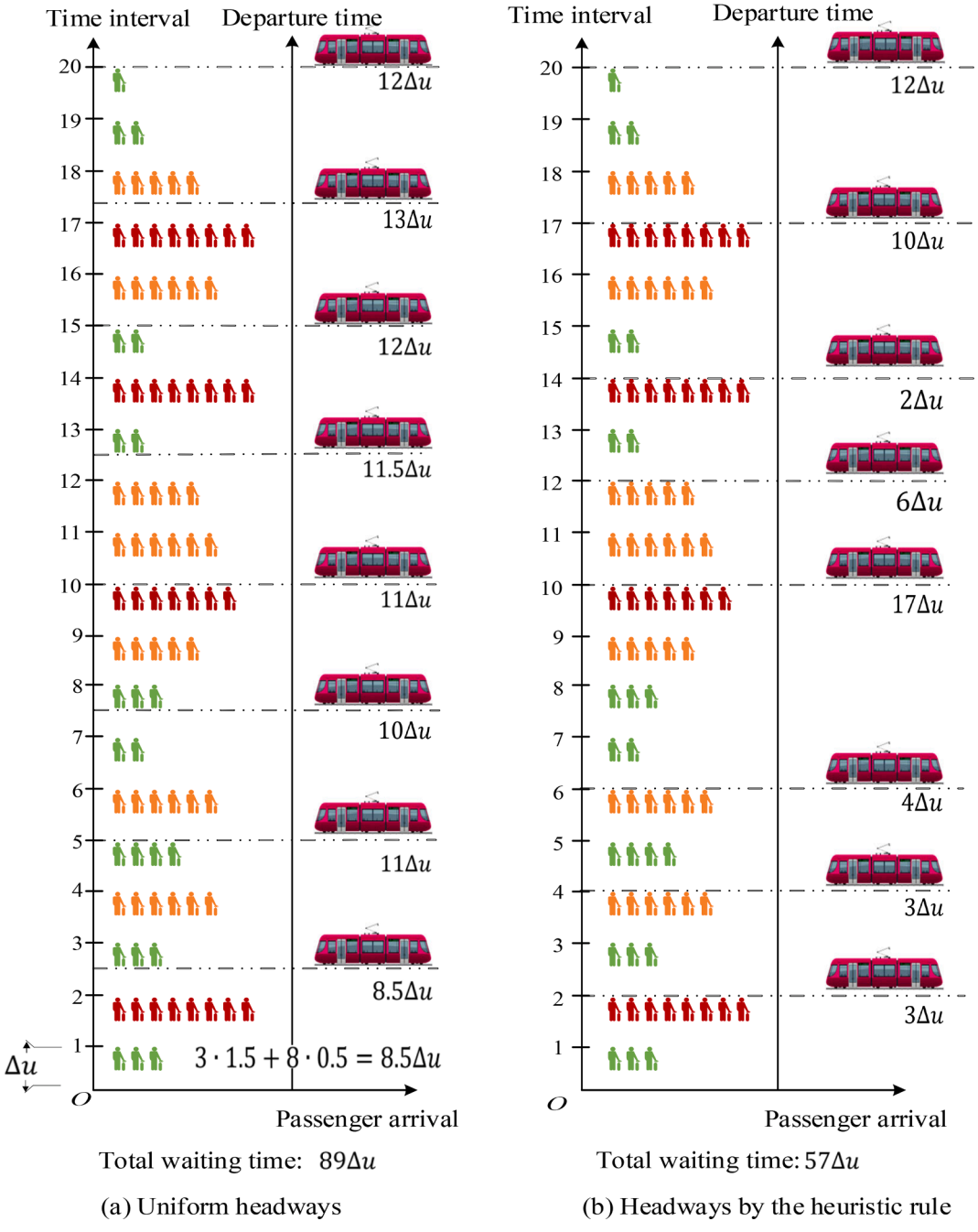


Fig. 8. Illustration of the combination of train departure and passenger arrival.

and vice-versa. The recursive formulation (42) performs in a topological ordering and needs to accommodate the dependencies among the items, which results in a higher order of time complexity than that for the ordinal knapsack problem. Nevertheless, since the number of tracks $|T|$ and the number of speed profile levels $|G|$ in the URT network have bounded values in reality (for example, the Shanghai URT has around 980 tracks, being the largest among all metropolis, and the total speed profile levels are usually less than 6), the DP is feasible to solve the TKP for large-scale URT networks with high accuracy.

4.3. Heuristic method

After the track running times are determined by Eq. (42), solving the MILP in the second stage of timetable optimization (Eq. (36)) still requires unaffordable computational time for a large number of binary variables. Since this part is iterated in the flow adjustments, we propose an approximation algorithm based on the domain knowledge for computational efficiency. As set up in Section 2, the planning period of timetabling is equally divided into a set of discrete time intervals of passenger arrivals. There are an explosive number of combinations of train dwell times and departure times to serve the discrete time-dependent passenger arrivals. To reduce the overall passenger waiting times, the typical uniform timetable should be relaxed to be nonuniform to serve the uneven passenger arrivals. It is instrumental to determine the train dwell times and departure times based on the density of passenger arrivals. Therefore, the guiding heuristic rule is to place the train departures at time points immediately or slightly after those associated with higher volumes of passenger arrivals.

To implement the heuristic rule effectively, two steps are involved to find a satisfactory solution. In the first step, we insert train departures in the timeline of the planning period based on the constraints of fixed fleet size and headways. Specifically, for one train to be scheduled, its departure is positioned right after the time interval with the highest volume of passenger arrivals that cannot be served by a neighboring train, while satisfying the headway constraint. Then, the next train is positioned in the timeline to serve the remaining highest volume of passenger arrivals. Following this rule, all the train departures are determined one by one, and an order of train departures and discrete passenger arrivals is generated for producing tighter time windows for departure times. In the second step, according to the generated order, a MILP model with significantly reduced complexity is formulated to find the specific departure times and dwell times. Since the track running times are already known from Eq. (42), the objective of the MILP model is simplified to minimize the total passenger waiting time for each train at every platform. The simplified MILP model only has $|K_l| + |P_l|$ decision variables ($|K_l|$ for departure times and $|P_l|$ for dwell times). The departure times have a smaller feasible solution space given the discrete interval of passenger arrival (e.g., 180 s). Moreover, the MILP can be solved in parallel for each URT line. The input of this simplified MILP is the volume of passengers waiting for each train. The equivalent passenger flow waiting for the train is counted in the unit of time interval Δu.

Given the order of train departures and discrete passenger arrivals, the passenger waiting for each train can be formulated according to the assumption that the passengers can board the first incoming trains (Assumption 1). Fig. 8 illustrates the comparison between a uniform timetable and a nonuniform one applying the heuristic rule. For flexible headways of 8 trains and 20 discrete time-dependent passenger arrivals, there are 75,582 (or $\frac{19!}{8!(19-8)!}$) combinations of discrete departures and passenger arrivals. The waiting times are calculated for those passengers waiting between two dotted train departure lines. Compared to Fig. 8(a), Fig. 8 (b) shows one set of headways based on the heuristic rule, which produces a lower total passenger waiting time. For illustration purposes, the passengers in red, orange, and green arrive in high, medium, and low densities respectively. For calculating the waiting time, we consider that the discrete passenger arrivals occur during a time interval. For instance, the uniform headway in Fig. 8 (a) is 2.5 Δu and the waiting time is calculated as 8.5 Δu (or $3 \bullet 1.5 + 8 \bullet 0.5$) for the first two discrete groups of passenger arrivals, while the counterpart waiting time is 3 Δu in Fig. 8 (b). The heuristic rule applies to all URT lines to generate sequences of all train departures. The simplified MILP model can efficiently obtain the optimized departure and dwell times. Consequently, all determined time components form a satisfactory nonuniform timetable solution. Notably, the waiting times for transfers can be reduced by the heuristic rule since the transfer stations often have high volumes of passenger arrivals.

Remark 5. Given all the track running times, the determination of headways and dwell times for good synchronization to reduce the total passenger travel time is still a hard problem in a URT network (Ceder and Tal, 2001; Ibarra-Rojas et al., 2012). The proposed heuristic rule attaches more importance to the higher volumes of passenger arrivals by positioning train departures closer to the denser passenger arrivals. The method by default attends to transfer stations of the URT network, which often have high volumes of passenger arrivals. Once the passenger arrivals are not uniform in the timeline, the nonuniform headways applying the proposed heuristic rule can always save waiting time, compared with the uniform headways. When there is a large variance in the density of the discrete passenger arrivals, the heuristic method is supposed to save waiting time considerably. Since track running times are determined in a prior step, the total passenger travel time is also reduced.

Overall, the three decomposed modules (shown in Fig. 5) together tackle the high complexity of energy-efficient timetabling for a URT network with multiple interlinked lines. The two-stage approach in module 2 underlies the core of the iterative algorithmic framework. The TKP formulation and heuristic method proposed in module 2 cater to speeding up the iterations to find a satisfactory timetable solution.

5. Case study

Considering a real-world application, we demonstrate the effectiveness of the proposed model framework and solution algorithm

using the URT network in the Xi'an metropolitan area (China) in Fig. 9. Detailed information on the URT is provided in Section S2 (in the supplementary document). The operational parameters of each line from the operator are provided in Table 2.

The AFC data of passenger demand were collected between 10:00 and 11:00 during the morning non-peak time on an average working day in 2019. The track running times and speed profile levels are provided in Table S3 (in the supplementary document). The algorithm is programmed in the MATLAB platform on a personal computer (8 G RAM and Intel Core i7-6700 CPU), while the sub-module for the MILP to determine the optimal headways and dwell times is coded in Yalmip + Gurobi. The mass of the train is 205,000 kg and the average passenger weight is set to 65 kg (Canca and Zarzo, 2017). Based on preliminary tests, the minimum total passenger running time increment of all tracks is more than 10^5 s during the planning period. Hence, we set $\Delta v_0 = 10^4$ s (equivalently about 1/11 of the average minimum track running time increment) as the time unit in the DP algorithm to solve the TKP. The setup can efficiently reduce one order of computation time of the DP while maintaining near-equal fidelity compared to using 10^3 s as the time unit.

The case study is developed in an accumulative way. Using the existing operational data, Subsection 5.1 verifies the model framework and solution algorithm and particularly tests different passenger behavior mechanisms. Subsection 5.2 discusses two sets of sensitivity analyses. First, we consider the influence of different allowed travel time increments to reveal the relationship between optimal train energy consumption and total passenger travel time. Second, to highlight the effects of passenger demand and train supply in the URT network, we consider the influence of different passenger demands and train fleet sizes on train energy consumption and passenger travel time.

5.1. Optimal energy consumption given allowed travel time increments

Based on the URT network structure and operations, a *k*-shortest path procedure is utilized to generate feasible physical paths in the

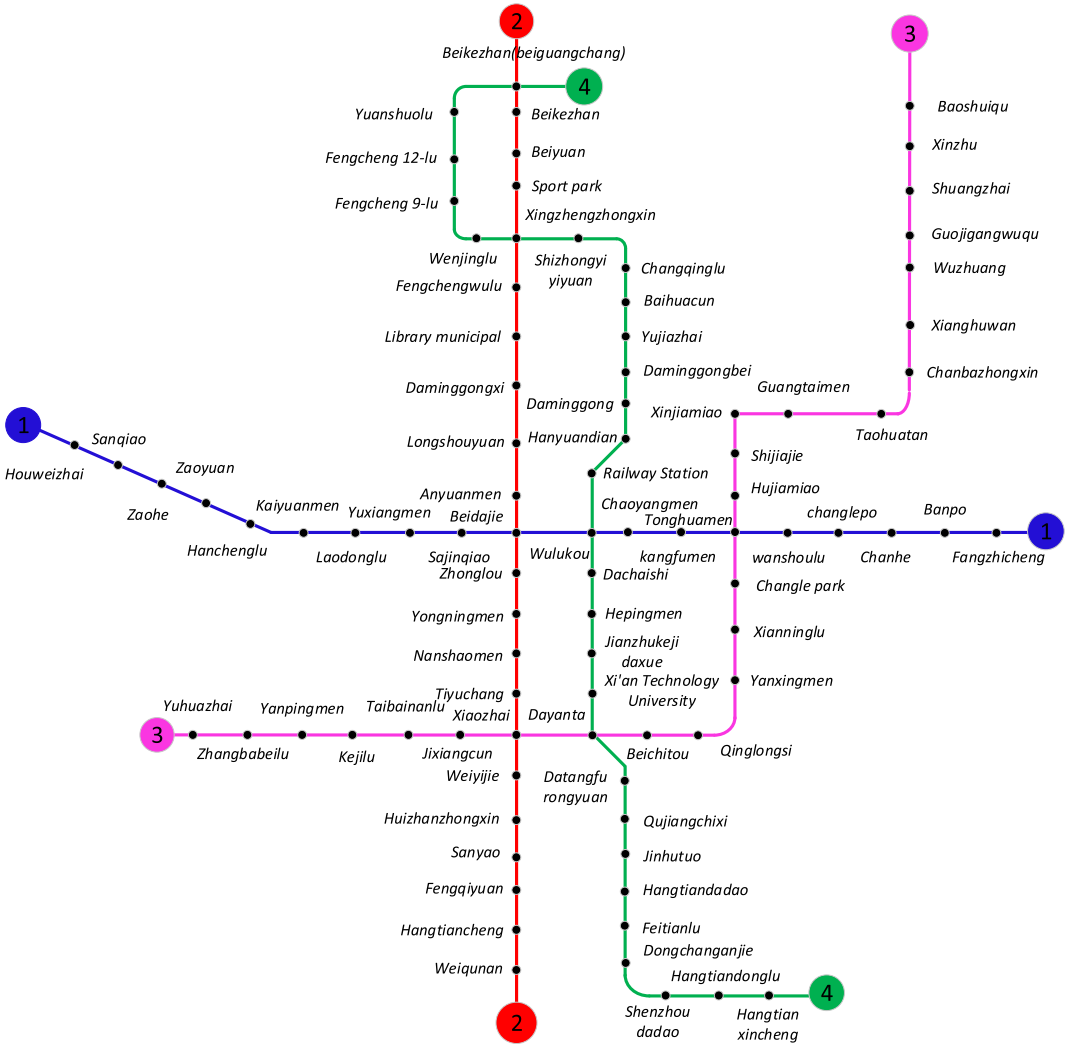


Fig. 9. The Xi'an URT network.

Table 2
The operational parameters for each URT line.

Line	Train number/fleet size	Minimum headway	Maximum headway
Line 1	13	240 s	300 s
Line 2	18	200 s	240 s
Line 3	20	180 s	210 s
Line 4	12	270 s	300 s

URT network. According to Assumption 4, we set constraints such that k equals 3 in the k-shortest path procedure and an alternative path should neither involve more than two transfers nor travel time 1.5 times more than the shortest travel time. In total, there are 9324 valid alternative paths for 8730 OD pairs in the choice set, of which 6750 paths need at least one transfer. The discrete passenger demands are provided per time interval of 180 s, making 20 (or 3600/180) discrete OD matrices. The time interval (180 s) of discrete passenger demand is based on the existing statistical precision provided by the URT operator. Since the passenger demand for URT may be very large, focusing on every passenger's precise time in the lower level model would cause a huge computation burden. Usually, the passenger demand is counted in groups in a time interval. It is not a critical issue modeling-wise to set a smaller time interval (e.g., 60 s). With smaller time intervals, the calculation precision is supposed to be higher. In the one-off stochastic logit assignment, $\overline{\beta}$ is set as $\pi^2/6\overline{\sigma}$, where $\overline{\sigma}$ is the travel time variance of all valid paths (Chen et al., 2015). Based on the original timetable, we have $\overline{\beta} = 0.0015$ and the passenger flow adjustment coefficient $\omega_j = 0.75/j$.

In module 2, it takes about 30 s to solve the TKP by DP, and the simplified MILP for headways and dwell times is solved within 1 s. The passenger assignment takes about 5 s in each iteration. The total average computation time under each behavior mechanism is about 500 s for 20 iterations of modules 2 and 3. Under different passenger behavioral mechanisms (Eq. (23)), the initial assignment results are provided in Table 3. We can see that there are some different effects on transfer and waiting times but very limited effects on the total passenger travel time, track running time, and energy consumption. For addressing a timetable with nonuniform headways, it is meaningful to consider the non-peak hour with uneven passenger arrivals. As found, there is little difference among these three mechanisms in the non-peak hour, which is somehow reasonable. The detailed comparison reveals that all-or-nothing assignment (ANA) shows the best performance with the present passenger demand in the non-peak hour. It can be foreseen that the linear proportional assignment (LPA) and one-off stochastic logit assignment (LA) may involve less energy consumption with more passenger demand. After running the two-stage approach with the three passenger behavioral mechanisms, the proportions of different time components in the URT are shown in Fig. 10. Although the passenger volumes on different paths are changed, all the time proportions under the three passenger behavioral mechanisms have little difference. These indicate the stability of different time proportions and confirm the conjecture stipulated in Section 4.1 (see Section S5 in the supplementary document for a verification of the conjecture).

The comparison between the original and a selected optimized timetable with $\Delta c = 10\%$ can be seen in Fig. 11. The timetables of Lines 1 and 3 are taken for illustrations. In the optimized timetable, the departure times of Line 1 are slightly adjusted later, whereas those of Line 3 are shifted later to a slighter extent for catering to passenger arrivals. The trip time from the starting platform to the ending platform shows no obvious increments for the optimized timetable in both Lines 1 and 3, but the optimized energy consumption is decreased to 61366.9 kWh, standing for a 13.8% reduction in energy consumption compared to 71193.5 kWh associated with the original timetable (see Table 3). Therefore, it is concluded that a nonuniform timetable with flexible headways to serve the time-varying passenger demands can sufficiently facilitate energy consumption reduction.

To further optimize energy consumption, we set a higher Δc with $\Delta c = 20\%$, the energy consumption and travel time can converge after a dozen of passenger flow adjustments as shown in Fig. 12(a)-(b). Note that to clearly show the differences in each iteration, the scale of the Y-axis is magnified. Taking the ANA as the path choice mechanism for instance, the optimized energy consumption is 56048.2 kWh, which means a 21.27% of energy consumption reduction compared to the original timetable. It is seen that ANA shows the best performance in both energy consumption and travel time. Similarly, with $\Delta c = 20\%$, the convergent energy consumption and travel time under each behavior mechanism show only a little difference, implying that the three passenger behavior mechanisms are representative during non-peak hours. This result signifies that the two-stage approach in module 2 can efficiently reduce energy consumption. In sum, the results verify the effectiveness of the presented model framework and solution algorithm in achieving energy efficiency subject to the constraint of given travel time increments.

Table 3
Initial assignment results under different passenger behavioral mechanisms.

Mechanisms	EC (kWh)	TT (h)	RT (h)	TST (h)	WT (h)	DWT (h)
ANA	71193.5	94245.46	52113.14	9527.72	16759.35	15845.24
LPA	71331.6	94820.75	52662.62	9272.08	15947.83	16938.21
LA	71289.6	94610.94	52495.27	9345.76	15886.73	16883.18

ANA: all-or-nothing assignment; LPA: linear proportional assignment; LA: one-off stochastic logit assignment; EC: Energy consumption; TT: Total passenger travel time; RT: Total running time; TST: Total transfer time; WT: Total waiting time; DWT: Total dwell time.

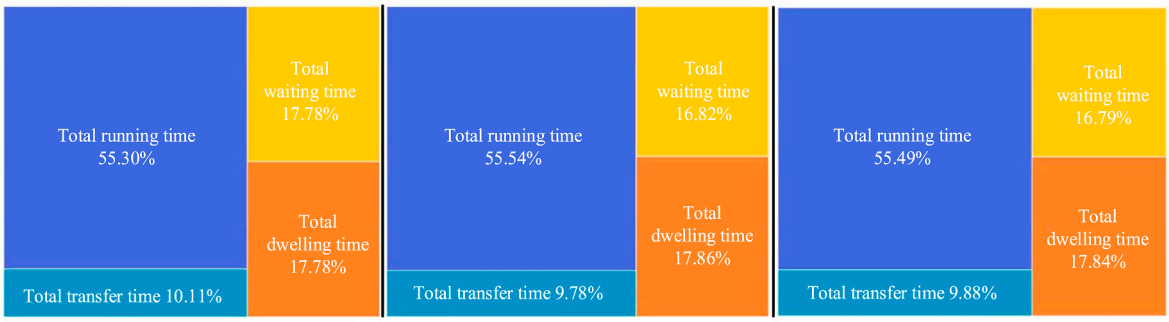


Fig. 10. Proportions of different time components under three passenger behavioral mechanisms.

5.2. Sensitivity analyses

5.2.1. Influence of allowed travel time increments (Δc)

Besides $\Delta c = 10\%$ and $\Delta c = 20\%$, we conduct a sensitivity analysis on Δc. The relationship between optimal energy consumption and different travel time increments is shown in Fig. 13, which includes nine points of travel time increments, i.e., 3%, 5%, 7%, 10%, 12%, 15%, 18%, 20%, and 24%, in Fig. 13(a). The results conform to the observation that train energy consumption and passenger travel time are two contradictory objectives in real operations. The maximum passenger travel time occurs when trains run with the lowest speed profiles on all tracks, corresponding to the longest track running times and a 24% increment compared with the original timetable. To reflect the relationship between train energy consumption and passenger travel time, they are normalized in the range [0, 1] by $\frac{\phi-\phi^{min}}{\phi^{max}-\phi^{min}}$, where ϕ is the specific value of energy consumption or travel time, ϕ^{max} and ϕ^{min} are their maximum and minimum values, respectively. Suppose that the weights of energy consumption and total passenger travel time add up to 1. Then, the objective becomes the sum of energy consumption and total passenger travel time with their respective weights. Considering a common difference of 0.1 and nine points of travel time increments, results are shown in Fig. 13(b) in different weight combinations. When both weights are 0.5, the lowest objective value is obtained when the travel time increment is coincidentally at 10%.

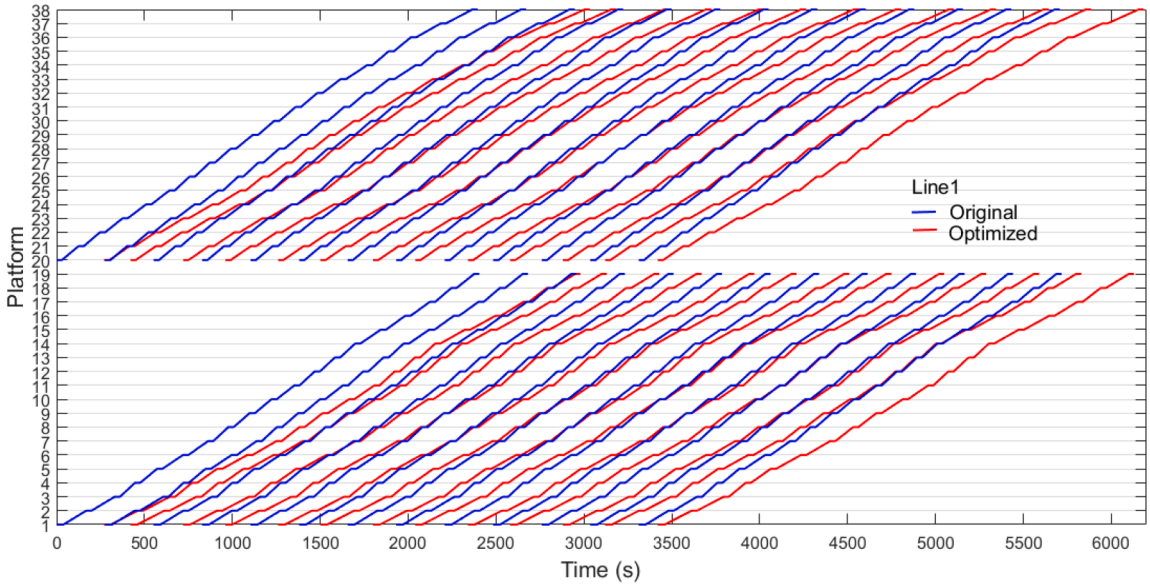
The relationship between energy consumption reduction and travel time increment is shown in Fig. 14. Energy consumption reduction increases with the increase of allowed travel time increment in Fig. 14(a). The ratio of energy consumption reduction to travel time increment decrease progressively as shown in Fig. 14(b). After the travel time increment is around 10%, the ratio decreases slowly, meaning that exchanging energy consumption by travel time becomes ineffective marginally. Hence, $\Delta c = 10\%$ produces a satisfactory compromised timetable solution. Notice that when the allowed travel time increment is around 20%, the ratio of energy consumption reduction to travel time increment tends to be close to 1. This indicates a near equivalent exchange in percentage between the two key indicators.

With $\Delta c = 10\%$, the proportions of different time components under the three passenger behavioral mechanisms after optimization are listed in Table 4. These proportions have smaller differences compared to those of the original timetables in Fig. 10. The proportions of the total transfer time slightly decrease despite the presence of a modest rise in waiting time, indicating better transfer synchronizations. The optimal energy consumption for ANA, LPA, and LA are 61366.96 kWh, 61337.38 kWh, and 61451.63 kWh, all about 13.8% energy consumption reduction. From this sensitivity analysis, we observe downward efficiencies in exchanging energy consumption by travel time and a 10% of travel time increment seems reasonable and beneficial in this particular case.

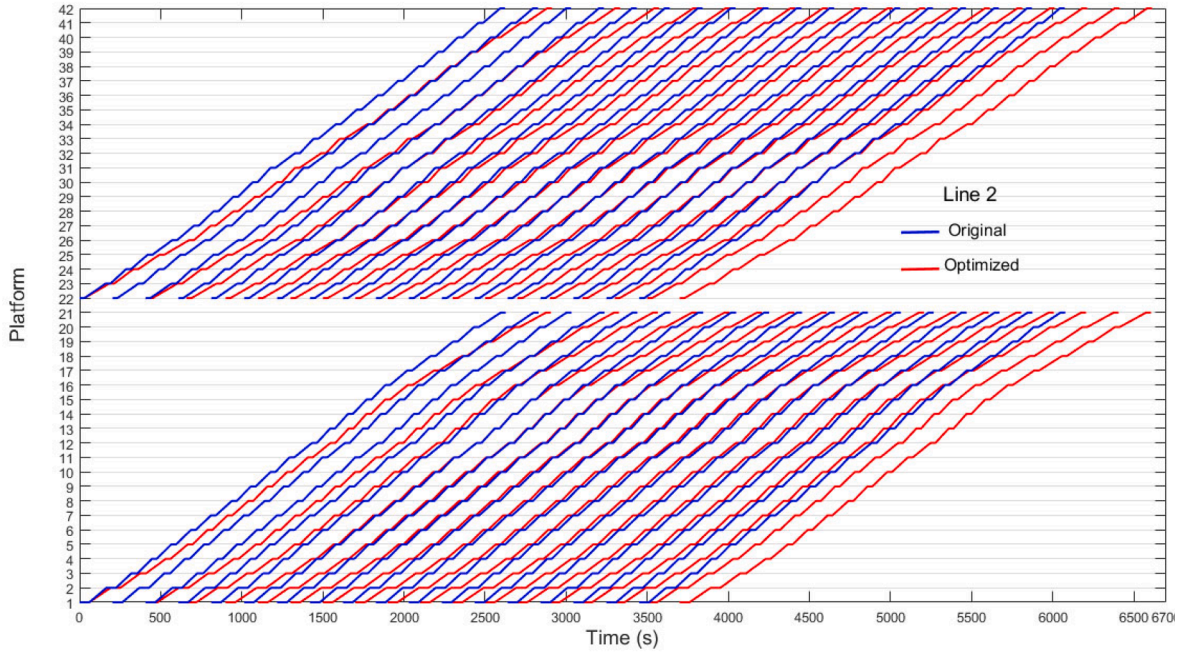
5.2.2. Influence of travel demand and fleet size

We further conduct a sensitivity analysis to evaluate the effects of passenger demand and fleet size in the URT network. As found by Huang et al. (2021), the proportion of energy consumption by passenger weights is about 23% on average in a uniform timetable during peak hours. It is interesting to examine to what extent the passenger demand influences energy consumption during a non-peak hour. We suppose that the passenger demand proportionally changes (from 0% to 150% of the original demand with 10% as the common difference) and the train fleet size has variations in each line (1 and 2 positive and negative changes respectively) under five scenarios. Given the non-peak hour context, the maximum demand is set to 150% of the original travel demand. For the same reason, the train fleet size only has minor changes for each line. Expectedly, in Fig. 15, both the train energy consumption and passenger travel time increase with the increasing passenger demand. It is noteworthy that when the passenger demand is 0 (empty trains), the energy consumption corresponds to the train weights only. With the original passenger demand (corresponding to 1.0 on the x-axis), the energy consumed by train weights is about 72% (or 44122/61366.96), meaning 28% of the energy consumed by passenger weights. A slight increase in train fleet size causes a significantly larger increase in energy consumption, while only a modest decrease takes place in passenger travel time. This finding suggests that the operator may remove up to two trains in the non-peak hour for more energy consumption reduction on the condition that the passenger travel times are only slightly affected.

Specifically, we observe a nearly linear increase in train energy consumption and passenger travel time with the increase of passenger demand in Fig. 15(a)-(b). When the passenger demand is close to 0, the travel time is close to 0, expectedly. In Fig. 15(a),



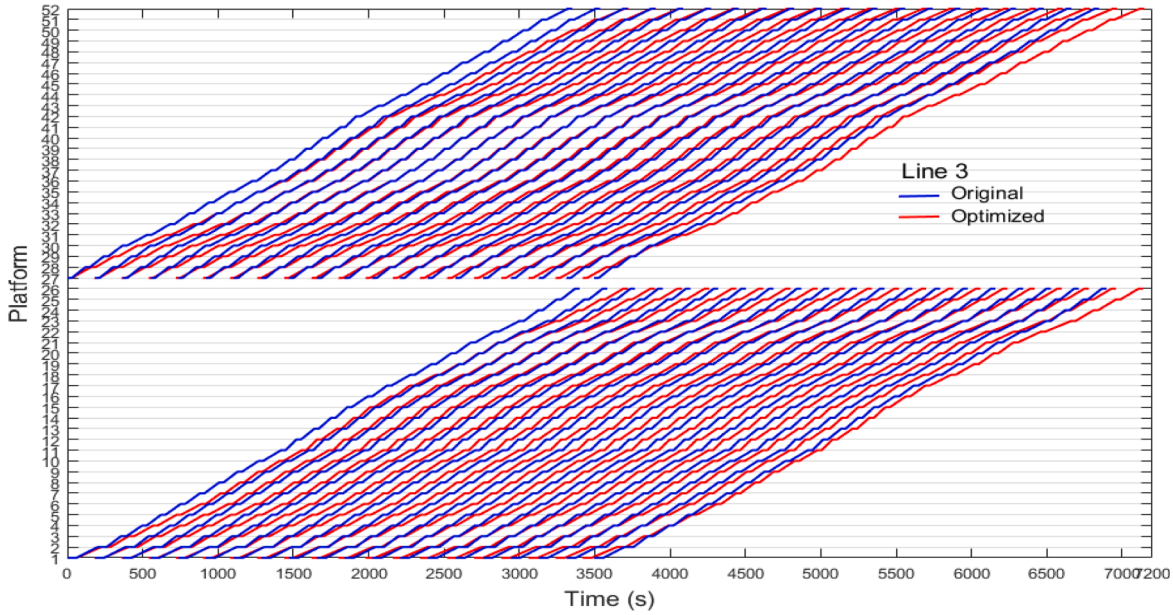
(a) Original and optimized timetables of Line 1.



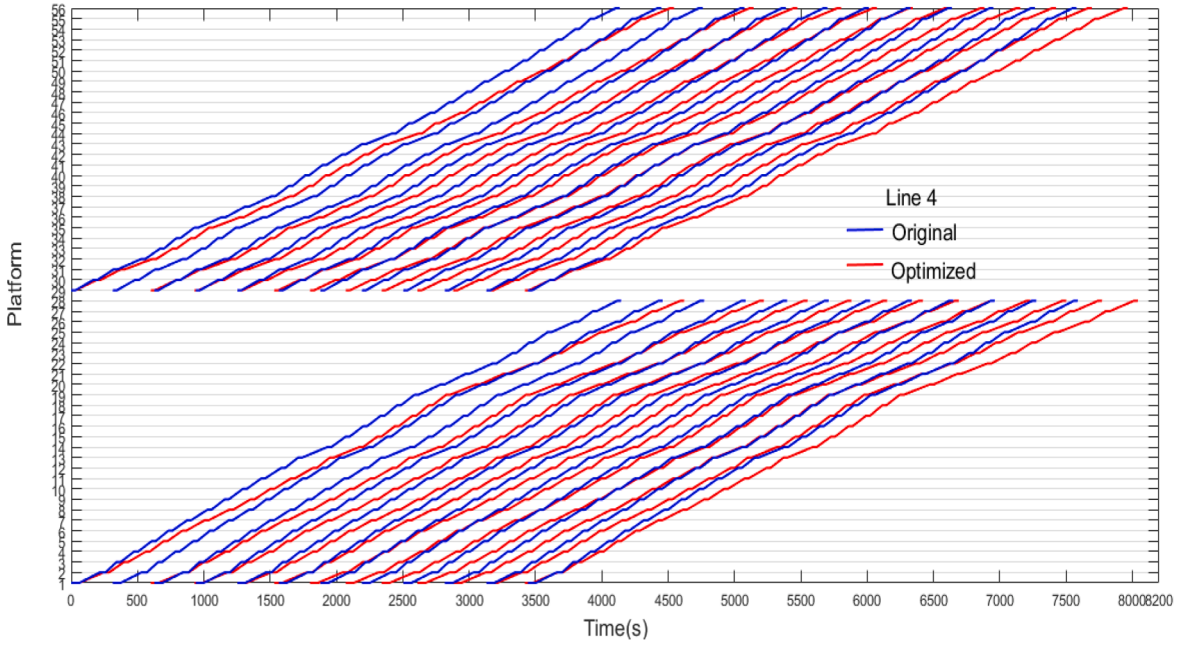
(b) Original and optimized timetables of Line 2.

Fig. 11. Timetable comparisons after optimization.

when the passenger demand is small (e.g., less than 10% of the original), the travel time is small but the consumed energy is near the value that corresponds to the empty train. When the passenger demand becomes smaller, the effect of the allowed 10% travel time increment is weaker. Therefore, the non-linear shape is apparent when the passenger demand approaches 0. The results emphasize the importance of the supply and demand relationship when designing an energy-efficient timetable. The total passenger travel time is only slightly influenced by the limited fleet size variation as shown in Fig. 15 (b). The two high-end curves refer to the decrease of two trains, while the two low-end curves refer to the increase of two trains. The optimal result of different passenger demands and train fleets with $\Delta c = 10\%$ are listed in Table S4 (in the supplemented document).



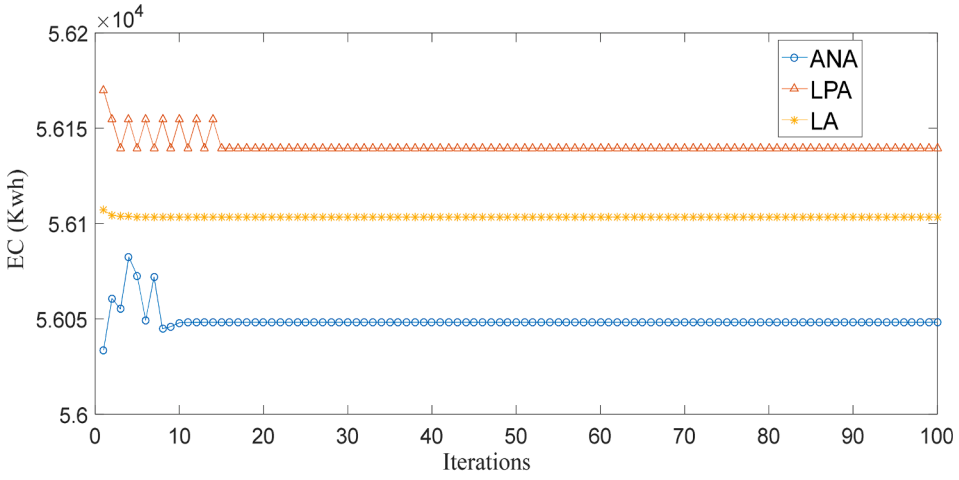
(c) Original and optimized timetables of Line 3.



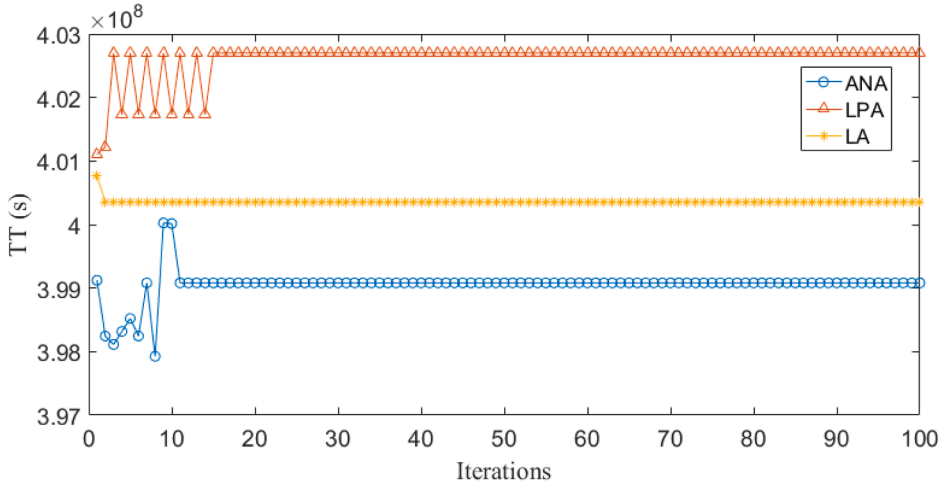
(d) Original and optimized timetables of Line 4.

Fig. 11. (*continued*).

Based on the above analyses, we have demonstrated that the suggested model framework and solution algorithm are efficient and effective for generating parallel and nonuniform energy-efficient timetables under different passenger path choice behaviors, passenger demands, and fleet sizes. The conjecture that travel time is stable given running time and train fleet size is verified by the results. Based on the conjecture, the path-based constraint of travel time is transformed into a track-based constraint of running time. Hence, the energy-efficient timetabling problem can be efficiently addressed by the proposed algorithmic framework. The results also show that the influences of passenger demand and train fleet are crucial to energy-efficient timetabling. Based on these findings, the operator can make a well-informed compromise between train energy consumption reduction and passenger travel time increment.

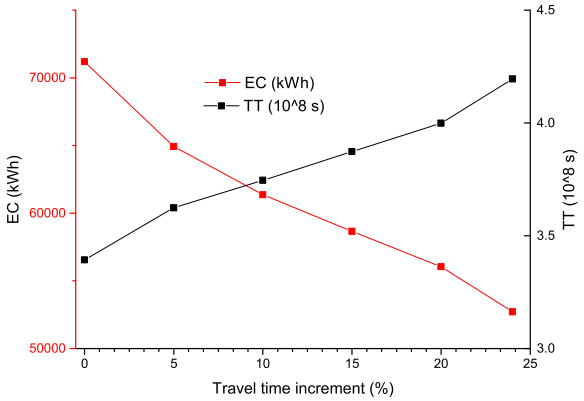


(a) The convergence of energy consumption.

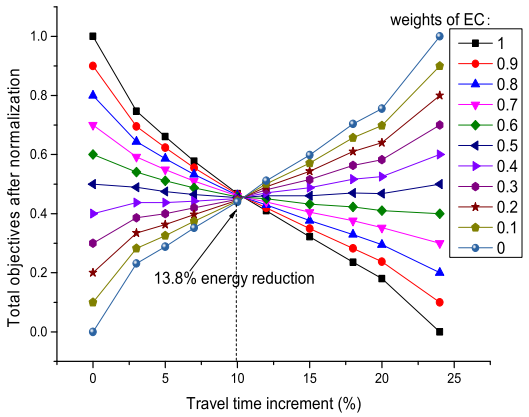


(b) The convergence of total passenger travel time.

Fig. 12. The convergence of different passenger behavioral mechanisms.

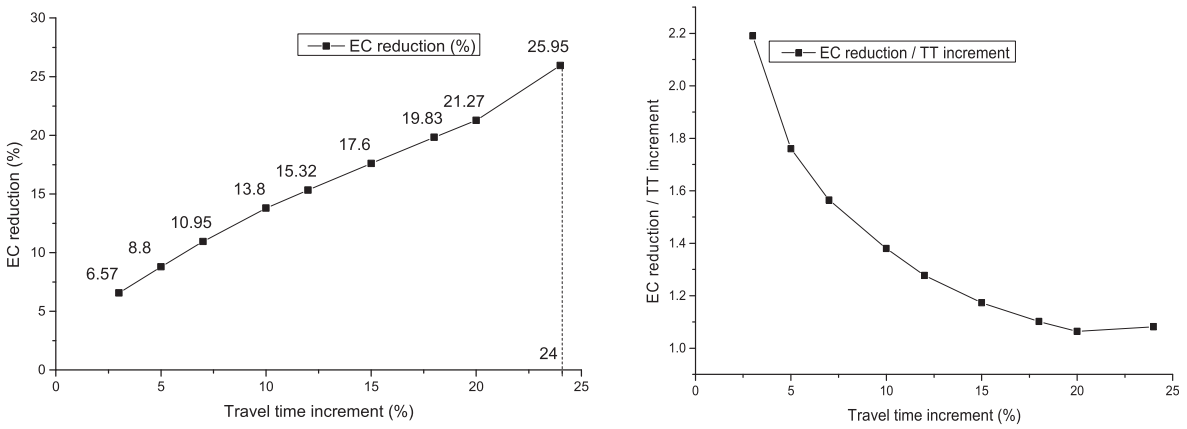


(a) Energy consumption and travel time.



(b) Normalized energy consumption and travel time

Fig. 13. Relationship between train energy consumption and passenger travel time with different Δc.

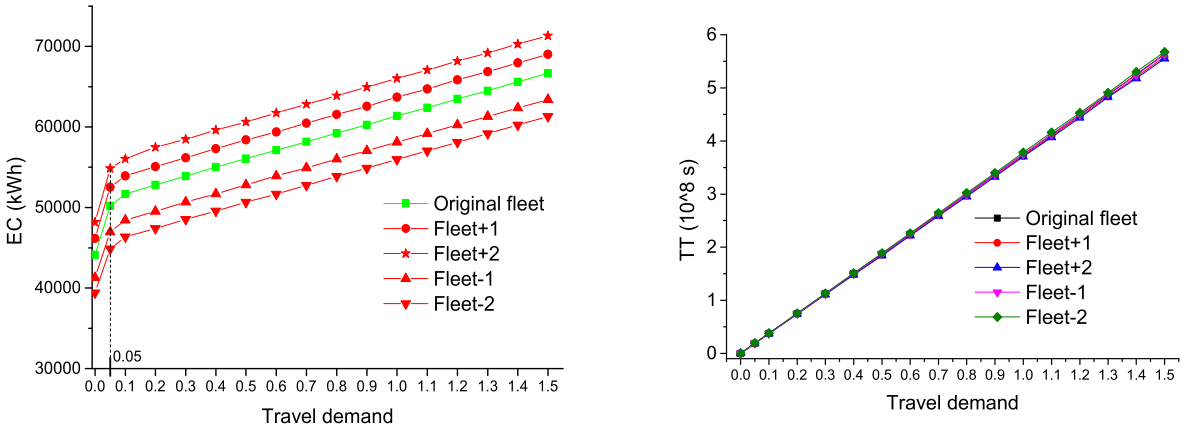


(a) Energy consumption reduction. (b) The ratio of energy consumption reduction.

Fig. 14. Relationship between energy consumption reduction and Δc.

Table 4
The proportions of different times for different passenger behavioral mechanisms after optimization (%).

Mechanisms	RT	TST	WT	DWT
ANA	55.85	9.42	18.45	16.28
LPA	56.1	9.13	18.42	16.35
LA	56.03	9.21	18.42	16.33



(a) Effects on energy consumption. (b) Effects on travel time.

Fig. 15. Effects of passenger demand and train fleet on energy consumption and total travel time.

6. Conclusions and future research

The majority studies in energy-efficient timetabling focus on a single URT line, and those few studies that extended this increasingly important problem to an entire URT network have notable limitations. In this study, we developed a general model framework including timetabling and passenger assignment considering passenger travel time constraints. Three passenger loading mechanisms were considered to capture path choice behaviors in non-peak hours. Based on a novel model reformulation, we proposed a DP algorithm and a heuristic method for determining track running times, headways, and dwell times, together constituting a satisfactory timetable solution. The performance of the suggested model framework and solution algorithm was validated in the numerical experiments using a real-world URT network. The model framework can obtain a significant reduction in energy consumption given certain allowed travel time increments. Moreover, we explored the relationship between train energy consumption reduction and passenger travel time increment and evaluated the influences of passenger demand and train fleet size. The results generated insights for URT operators to for making satisfactory compromised timetables.

A few research topics can be further investigated based on the proposed model framework of energy-efficient timetabling. First, the assumptions stated in Section 2 should be relaxed to enhance realism. Second, the model should consider the overcrowding scenarios where the passengers may be stranded on the platform. Third, passenger flow management and control can be added up to the model framework to achieve energy-efficient timetabling. Fourth, since the fleet size is influential over energy consumption, fleet size management should be incorporated to adapt to passenger demands at different levels. Fifth, to further demonstrate the applicability of the proposed model framework and two-stage approach, more benchmark problems should be tested, which however has been quite rare in this line of research due to the lack of data. Sixth, the case study should be extended with flexible time intervals of discrete passenger demands and validated travel preferences and chains in multimodal transport systems (e.g., Liao et al., 2012, 2013, 2020). Finally, the influence of the network topology on energy consumption should also be explored for designing energy-efficient and robust timetables.

CRediT authorship contribution statement

Kang Huang: Conceptualization, Methodology, Validation, Writing – original draft. **Feixiong Liao:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work is jointly supported by the National Natural Science Foundation of China (72288101, 72101018, 72271127, 71801134) and the Dutch Research Council (NWO). The first author is grateful to the financial support by the China Scholarship Council (CSC).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tre.2023.103212.

References

Albrecht, A.R., Howlett, P.G., Pudney, P.J., Vu, X., 2013. Energy-efficient train control: from local convexity to global optimization and uniqueness. Automatica 49, 3072–3078.
Albrecht, A., Howlett, P., Pudney, P., Vu, X., Zhou, P., 2016a. The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points. Transp. Res. B 94, 482–508.
Albrecht, A., Howlett, P., Pudney, P., Vu, X., Zhou, P., 2016b. The key principles of optimal train control—Part 2: Existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques. Transp. Res. B 94, 509–538.
Berggren, U., Kjær-Rasmussen, T., Thorhauge, M., Svensson, H., Brundell-Freij, K., 2021. Public transport path choice estimation based on trip data from dedicated smartphone app survey. Transportmetrica A 1–34.
Caccetta, L., Kulanoot, A., 2001. Computational aspects of hard knapsack problems. Nonlinear Anal. 47 (8), 5547–5558.
Cai, X., Goh, C.J., 1994. A fast heuristic for the train scheduling problem. Comput. Oper. Res. 21 (5), 499–510.
Canca, D., Zarzo, A., 2017. Design of energy-efficient timetables in two-way railway rapid transit lines. Transp. Res. B 102, 142–161.
Ceder, A., Tal, O., 2001. Designing synchronization into bus timetables. Transp. Res. Record: J. Transp. Res. Board 1760 (1), 28–33.
Chen, Y., Mao, B., Bai, Y., Ho, T.K., Li, Z., 2019. Timetable synchronization of last trains for urban rail networks with maximum accessibility. Transp. Res. C 99, 110–129.
Chen, B., Si, B., Jiang, M., Yang, X., 2015. Scheduled- based equilibrium assignment model and algorithm for urban subway network. Sci. China Technol. Sci. 45 (12), 1269–1278.
Howlett, P., 1996. Optimal strategies for the control of a train. Automatica 32 (4), 519–532.
Howlett, P., Pudney, P., 1998. An optimal driving strategy for a solar powered car on an undulating road. Dyn. Continuous Discrete Impulsive Syst. 4, 553–567.
Howlett, P., Pudney, P., Vu, X., 2009. Local energy minimization in optimal train control. Automatica 45 (11), 2692–2698.
Huang, K., Wu, J., Yang, X., Gao, Z., Liu, F., Zhu, Y., 2019. Discrete train speed profile optimization for urban rail transit: a data-driven model and integrated algorithms based on machine learning. J. Adv. Transp. 4, 1–17.
Huang, K., Liao, F., Gao, Z., 2021. An integrated model of energy-efficient timetabling of the urban rail transit system with multiple interconnected lines. Transp. Res. C 129, 103171.
Hui, C., 2014. Application study of all-or-nothing assignment method for determination of logistic transport route in urban planning. Comput. Modelling New Technol. 18, 932–937.
Ibarra-Rojas, O.J., Giesen, R., Rios-Solis, Y.A., 2012. Synchronization of bus timetabling. Transp. Res. B 46 (5), 599–614.
Kang, L., Sun, H., Wu, J., Gao, Z., 2020. Last train station-skipping, transfer-accessible and energy-efficient scheduling in subway networks. Energy 206, 118127.
Khmelnitsky, E., 2000. On an optimal control problem of train operation. IEEE Trans. Autom. Control 45 (7), 1257–1266.

Kuppusamy, P., Venkatraman, S., Rishikeshan, C.A., Reddy, Y.P., 2020. Deep learning based energy efficient optimal timetable rescheduling model for intelligent metro transportation systems. Phys. Commun. 42, 101131.
Lai, Q., Liu, J., Haghani, A., Meng, L., Wang, Y., 2020. Energy-efficient speed profile optimization for medium-speed maglev trains. Transp. Res. E 102007.
Li, X., Lo, H.K., 2014a. An energy-efficient scheduling and speed control approach for metro rail operations. Transp. Res. B 64, 73–89.
Li, X., Lo, H.K., 2014b. Energy minimization in dynamic train scheduling and control for metro rail operations. Transp. Res. B 70, 269–284.
Li, X., Zhang, B., Liu, Y., 2021. A little bit flexibility on headway distribution is enough: data-driven optimization of subway regenerative energy. Inf. Sci. 554, 276–296.
Liao, F., 2016. Modeling duration choice in space–time multi-state supernetworks for individual activity-travel scheduling. Transp. Res. C 69, 16–35.
Liao, F., 2019. Joint travel problem in space–time multi-state supernetworks. Transportation 46 (4), 1319–1343.
Liao, F., Arentze, T., Timmermans, H., 2012. Supernetwork approach for modeling traveler response to park-and-ride. Transp. Res. Rec. 2323 (1), 10–17.
Liao, F., Arentze, T., Timmermans, H., 2013. Incorporating space–time constraints and activity-travel time profiles in a multi-state supernetwork approach to individual activity-travel scheduling. Transp. Res. B 55, 41–58.
Liao, F., Tian, Q., Arentze, T., Huang, H.J., Timmermans, H., 2020. Travel preferences of multimodal transport systems in emerging markets: the case of Beijing. Transp. Res. A 138, 250–266.
Liu, Y., Bunker, J., Ferreira, L., 2010. Transit users' route-choice modelling in transit assignment: a review. Transp. Rev. 30 (6), 753–769.
Liu, P., Schmidt, M., Kong, Q., Wagenaar, J.C., Yang, L., Gao, Z., Zhou, H., 2020. A robust and energy-efficient train timetable for the subway system. Transp. Res. C 121, 102822.
Lv, H., Zhang, Y., Huang, K., Yu, X., Wu, J., 2019. An energy-efficient timetable optimization approach in a bi-direction urban rail transit line: a mixed-integer linear programming model. Energies 12 (14), 2686.
Martello, S., Toth, P., 1990. Knapsack Problems: Algorithms and Computer Implementations. John Wiley and Sons Inc.
Mo, P., Yang, L., D'Ariano, A., Yin, J., Yao, Y., Gao, Z., 2019a. Energy-efficient train scheduling and rolling stock circulation planning in a metro line: a linear programming approach. IEEE Trans. Intell. Transp. Syst. 21 (9), 3621–3633.
Mo, P., Yang, L., Wang, Y., Qi, J., 2019b. A flexible metro train scheduling approach to minimize energy cost and passenger waiting time. Comput. Ind. Eng. 132, 412–432.
Mo, P., D'Ariano, A., Yang, L., Veelenturf, L.P., Gao, Z., 2021. An exact method for the integrated optimization of subway lines operation strategies with asymmetric passenger demand and operating costs. Transp. Res. B Methodol. 149, 283–321.
Parkinson, T., Fisher, I., 1996. Rail transit capacity. Transp. Res. Board, 13, 119–122.
Qi, J., Yang, L., Di, Z., Li, S., Yang, K., Gao, Y., 2018. Integrated optimization for train operation zone and stop plan with passenger distributions. Transp. Res. E 109, 151–173.
Qu, Y., Wang, H., Wu, J., Yang, X., Yin, H., Zhou, L., 2020. Robust optimization of train timetable and energy efficiency in urban rail transit: a two-stage approach. Comput. Ind. Eng. 146, 106594.
Scheepmaker, G.M., Goverde, R.M., Kroon, L.G., 2017. Review of energy-efficient train control and timetabling. Eur. J. Oper. Res. 257 (2), 355–376.
Shang, P., Li, R., Liu, Z., Yang, L., Wang, Y., 2018. Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network. Transp. Res. C 89, 321–343.
Su, S., Wang, X., Tang, T., Wang, G., Cao, Y., 2021. Energy-efficient operation by cooperative control among trains: A multi-agent reinforcement learning approach. Control Eng. Pract. 116, 104901.
Wang, P., Goverde, R.M., 2017. Multi-train trajectory optimization for energy efficiency and delay recovery on single-track railway lines. Transp. Res. B 105, 340–361.
Wang, P., Goverde, R.M., 2019. Multi-train trajectory optimization for energy-efficient timetabling. Eur. J. Oper. Res. 272 (2), 621–635.
Wang, C., Meng, X., Guo, M., Li, H., Hou, Z., 2022. An integrated energy-efficient and transfer-accessible model for the last train timetabling problem. Phys. A 588, 126575.
Wang, X., Tang, T., Su, S., Yin, J., Gao, Z., Lv, N., 2021. An integrated energy-efficient train operation approach based on the space-time-speed network methodology. Transp. Res. E 150, 102323.
Xie, J., Zhang, J., Sun, K., Ni, S., Chen, D., 2021. Passenger and energy-saving oriented train timetable and stop plan synchronization optimization model. Transp. Res. D 98, 102975.
Xu, Y., Jia, B., Li, X., Li, M., Ghiasi, A., 2020. An integrated micro-macro approach for high-speed railway energy-efficient timetabling problem. Transp. Res. C 112, 88–115.
Yang, X., Chen, A., Ning, B., Tang, T., 2017. Bi-objective programming approach for solving the metro timetable optimization problem with dwell time uncertainty. Transp. Res. E 97, 22–37.
Yang, S., Wu, J., Yang, X., Liao, F., Li, D., Wei, Y., 2019. Analysis of energy consumption reduction in metro systems using rolling stop-skipping patterns. Comput. Ind. Eng. 127, 129–142.
Yang, S., Liao, F., Wu, J., Chen, Y., 2022. An efficient train timetable scheduling approach with regenerative-energy supplementation strategy responding to potential power interruptions. IEEE Trans. Intell. Transp. 23, 14267–14282.
Yang, S., Liao, F., Wu, J., Timmermans, H.J., Sun, H., Gao, Z., 2020. A bi-objective timetable optimization model incorporating energy allocation and passenger assignment in an energy-regenerative metro system. Transp. Res. B 133, 85–113.
Yang, X., Wu, J., Sun, H., Gao, Z., Yin, H., Qu, Y., 2019. Performance improvement of energy consumption, passenger time and robustness in metro systems: a multi-objective timetable optimization approach. Comput. Ind. Eng. 137, 106076.
Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: Mixed-integer linear programming approaches. Transp. Res. B 97, 182–213.
Yin, J., Su, S., Xun, J., Tang, T., Liu, R., 2020. Data-driven approaches for modeling train control models: comparison and case studies. ISA Trans. 98, 349–363.
Zhan, S., Wang, P., Wong, S.C., Lo, S.M., 2022. Energy-efficient high-speed train rescheduling during a major disruption. Transp. Res. E 157, 102492.
Zhang, Y., Peng, Q., Yao, Y., Zhang, X., Zhou, X., 2019. Solving cyclic train timetabling problem through model reformulation: extended time-space network construct and alternating direction method of multipliers methods. Transp. Res. B 128, 344–379.
Zhou, W., Huang, Y., Deng, L., Qin, J., 2023. Collaborative optimization of energy-efficient train schedule and train circulation plan for urban rail. Energy 263, 125599.