

# **Research** Article

# Estimating the Railway Network Capacity Utilization with Mixed Train Routes and Stopping Patterns: A Multiobjective Optimization Approach

Zhengwen Liao (D,<sup>1,2</sup> Haiying Li (D,<sup>1</sup> Jianrui Miao (D,<sup>1,2</sup> and Lingyun Meng (D<sup>1</sup>

<sup>1</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China <sup>2</sup>State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Jianrui Miao; jrmiao@bjtu.edu.cn

Received 1 August 2022; Revised 27 October 2023; Accepted 24 January 2024; Published 15 February 2024

Academic Editor: Dilum Dissanayake

Copyright © 2024 Zhengwen Liao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Railway capacity estimation problem is typically defined as estimating the maximum number of trains that can be operated in a railway section within a given time interval. However, trains with different speeds, routes, and stopping patterns in a railway network will likely compete for the limited capacity of network nodes and sections. As these trains may provide different services, it is ambiguous to simply indicate the network capacity by a scalar number of trains. To comprehensively estimate and interpret the railway capacity considering the capacity competition between heterogeneous trains, we propose a multiobjective perspective for the capacity estimation problem to enrich the capacity theory while handling the competition among trains with different routes and stopping patterns. Based on a time-space network timetable saturated timetable under capacity estimation objectives with respect to different routes and stopping patterns. With the  $\varepsilon$ -constraint method, we can obtain the Pareto front of saturated timetables, i.e., a set of nondominated optimized timetables that no more candidate train can be additionally scheduled. The result is a more comprehensive capacity representation than a single absolute scalar number. A case study is conducted on a combined high-speed and intercity network of Zhengzhou Railway group in China. An extensive set of Pareto-optimal saturated timetables describing the effects on the capacity of the railway network is obtained. The results can help infrastructure managers select saturated timetables as the capacity utilization reference by considering the trade-off between time indexes from passengers' and operators' perspectives.

## 1. Introduction

1.1. Background. The intensive capacity utilization of railway infrastructure has become the bottleneck restriction for improving passenger mobility. Before implementing the capacity enhancement measurements [1] (e.g., updating the signal system and building some extra tracks), an important method for increasing the railway capacity is to optimize the railway traffic pattern to reduce the capacity loss as far as possible.

Railway capacity is subject to many criteria, which can be classified into technical and operational factors. Technical factors include track layouts, the performance and the configuration of the signaling system, and train speed, whereas operational factors concern timetables, implying the traffic on infrastructures. In this paper, we discuss the railway capacity only by considering the operation factors and regard the facility factor as given. The trains with mixed routes and with mixed stopping patterns share the same railway corridor, resulting in capacity competition. Although these complicated train routes and stopping patterns ensure the minimum transfer for passengers by the diversity of train routes and maximize the average speed of trains by diversifying the stopping patterns, the capacity is strongly influenced by the complicated traffic mix. Therefore, it would be very necessary to estimate the capacity loss of implementing these mixed routes and stopping patterns and analyze the relationship between the capacity performance and other timetable performance aspects from operators' and passengers' perspectives.

In those railways with abundant capacity, the railway company tends to minimize the operation cost to satisfy the passenger or freight demand and evaluate the timetable through economic performance indexes. In these cases, the maximization of capacity utilization is not necessarily a performance index. They do not push the number of trains to the technical maximum but rather limit it such that a reasonable level of service might be expected [2]. However, in some capacity-intensive timetabling scenarios (i.e., a railway corridor where the passenger or freight demand is far over the capacity), the train capacity utilization becomes the most concerning performance index. In these cases, the more trains are scheduled, the higher income the railway company gets in general, but the transportation service quality (e.g., the average speed and the possibility of delays) might decrease when too many trains are scheduled. In other words, the operation object of these railways is to maximize the carrying capacity while guaranteeing the worse acceptable service quality-related performance indexes. For these capacity-intensive railways, before the train scheduling procedure, the traffic managers need to estimate the railway capacity in advance to decide the upper bound of the candidate trains so as to decide the size of the candidate train pool before the timetabling stage.

For analyzing the impact of mixed traffic (i.e., mixed routes and mixed stopping patterns) on capacity, we first introduce a timetable saturation method for estimating the railway capacity under mixed train traffic based on a set of different objectives under the classification of competition train groups. The relations between railway capacity and timetable performances of operation-centric and passengercentric are analyzed from the Pareto front. The correlations between railway capacity and service quality-related timetable performances are helpful for railway schedulers to arrange the traffic at the bottleneck sections better to achieve maximum capacity utilization.

The remainder of the paper is organized as follows. Section 1.2 gives the background introduction and the literature review of the related research. Section 2 describes the railway capacity estimation problem with mixed routes and stopping patterns formally. Then, a time-space network representation followed by an associated integer programming model is proposed in Section 3. A Lagrangian relaxation-based heuristic algorithm is proposed to solve the  $\varepsilon$ -constraint model for obtaining the Pareto front to solve the model efficiently. Section 5 provides a case study employing the real-world instances of Zhengzhou Railway group to show the capacity estimation result and related comparison between the capacity and other timetable performance indexes.

1.2. Literature Review. Railway capacity is defined according to the different operation environments of railway systems in different countries, as the enterprise architectures,

transportation volumes, and traffic patterns are diverse. The most common definition describes practical capacity as "the total number of possible paths in a defined time window, considering the actual path mix or known developments respectively and the IM's own assumptions in nodes, individual lines or part of the network with market-oriented quality" [3]. This definition describes the railway capacity in a static and deterministic manner, assuming the corresponding timetables can be executed without disturbances. Besides, several studies apply an extending definition considering potential train delay, as well as the robustness of the corresponding timetables. These studies regard the capacity as a "resilient" value corresponding to the level of feasibility in practice (e.g., Yuan and Hansen [4] and De Kort et al. [5]), as the higher number of scheduled trains might result in more frequent consecutive delays and more severe delay propagations. In this paper, we neglect the possibility of train delay caused by high-capacity utilization and adopt the traditional static and deterministic capacity definition. Several categories of capacity estimation methods have been proposed based on different fundamental methodologies: analytical, timetable-based, and simulation methods [6].

The analytical method calculates the capacity by proposing capacity calculation formulations. The formulations consist of capacity-related items, such as the minimum headway and the capacity loss caused by the mixed traffic. These capacity calculation formulations are able to work without determining a specific timetable. The parameters can be calibrated by the railway traffic practice of similar railway lines. The simple prerequisite of the analytical method allows it to be successfully performed in cases where detailed information is not given (e.g., the performance of the signaling system, the layout of the railway lines, or the timetables). However, the results given by the analytical methods are not as detailed and concrete compared with timetable-based or simulation-based methods. Armstrong and Preston [7] regard the service quality performance of railway stations as impact factors of railway capacity and investigate the different capacity utilization strategies on different service levels. Lai et al. [8] propose a normalized value of base train equivalents, which can approximately estimate the capacity of railway sections with mixed traffic. Based on queueing theory, Weik et al. [9] determine the railway capacity with given service quality constraints implied by consecutive delays. Rotoli et al. [10] use the capacity calculation formulation to estimate the capacity occupation level of the nodes and corridors of railway networks. Goverde [11] describes a linear system description of a railway timetable in max-plus algebra, and it can be applied for timetable compression.

The timetable-based methods are often used in the operational phase when the timetable structure is predefined. According to the level of freedom when tackling the train schedule, the timetable-based methods can be further classified into timetable compression and timetable saturation methods. The well-known timetable compression method recommended by UIC 406 leaflet [3] and its updated version [12] are widely used to calculate the capacity occupation. This timetable-dependent method provides the capacity occupation level of specific timetables with fixed train orders. In contrast to analytical methods, the timetable compression method requires detailed data in terms of timetabling constraints, such as the blocking time data for determining the minimum headway between two successive trains. To increase the level of flexibility of scheduling trains, Landex [13] and Landex [14] randomly generate timetables with various combinations of trains and structures and estimate the overall capacity performances of these timetables. These methods can be used to evaluate the overall capacity consumption without knowing the exact timetables and can be applied in the planning and designing phases of railway systems. Jensen et al. [15] propose a mixed-traffic capacity estimation framework by searching timetable structures using branch-and-bound and tabu search heuristics, where a timetable compression method is embedded. Besides, many studies apply the timetable compression method to investigate the impact of operational or technical factors on railway capacity, which can be referred to Jensen et al. [16], Goverde et al. [17], Jamili [18], and Zhang et al. [19]. The UIC 406 method is able to express the capacity use of one specific timetable by a simple percentage value, but no complex capacity competition is reflected.

The timetable saturation method is to insert as many standardized virtual trains as possible into the original timetable or to schedule an entirely new saturated timetable to calculate the maximum number of feasible trains running on the railway. This method could benefit from established train timetabling research. One popular methodology set to saturate timetables is Job-shop-based MIP model, such as Pellegrini et al. [20], Kim et al. [21], and Zhang [22]. To estimate the railway's capacity by applying cyclic timetables, Petering et al. [23] and Zhang and Nie [24] propose optimization models to minimize the cyclic time horizon length instead of maximizing the train number. Besides, the time-space network-based model has got more and more attention in recent timetabling research. Reinhardt et al. [25] apply a time-space network model to build a freight train schedule by minimizing train cancellation penalties. Yaghini et al. [26] and Yaghini et al. [27] apply a rough granularity time-space network to model the railway traffic flow and saturate timetables to estimate the capacity in the planning and construction phases. The multicommodity flow model can be solved by local branching, Lagrangian relaxation, or column generation algorithm, showing great potential for solving the capacity estimation problem of large-scale instances. Yaghini et al. [27] conclude that the mixture of train types reduces the railway capacity by applying the method proposed by Yaghini et al. [27]. Harrod et al. [28] apply a multicommodity flow model proposed in Harrod [29] to conduct an experiment of generating optimal timetables with mixed trains, which needs less calculation demand compared with other simulation methods. Similar to the UIC 406 method, the timetable saturation method can only get the total number of trains that can be scheduled, but no further information reflecting the capacity competition can be obtained.

Microscopic simulation is a typical method for capacity utilization validation with given timetables and/or train dispatching rules. Several general simulation toolkits built

on microscopic and macroscopic level infrastructure are widely used for capacity estimation. Typically, the simulation method can easily generate disturbances or disruptions to evaluate the capacity performance under specific potential delays. For example, RailSys is used for microscopic simulation (Lindfeldt [30]), and Rail Traffic Controller (RTC), reported by Shih et al. [31] and Dingler et al. [32] is used for macroscopic simulation. Besides, a famous simulation software, OpenTrack, is applied to simulate the train operation procedure for capacity estimation purposes [33]. Some research combined different timetabling or simulation tools with building the capacity estimation framework. The timetable compression or timetable saturation modules are embedded into the framework, showing the roles of timetable and facility in the railway capacity, such as Nash and Huerlimann [33] and Pouryousef and Lautala [34].

These methods mentioned above mainly concern the capacity problem in given railway sections. However, many preconditions are needed for capacity estimation to avoid ambiguity as trains compete for capacity in one or more critical sections on a network scale. Different types of trains are not always comparable while competing capacity, so measuring the capacity of a railway network only by the total number of trains or the percentage of capacity used is ambiguous. Therefore, a train bundle method for different types of trains is established by Vieira et al. [35]. This paper proposes an optimization method to estimate the capacity with given train operating parameters. A train bundle consists of a fixed proportion of trains of different types and can only be scheduled simultaneously. Burdett and Kozan [36] focus on the multimodal transportation system and propose a linear programming model to determine the maximum flow of multimodal rolling stocks. Similarly, Bevrani et al. [37] propose an MIP model maximizing the number of trains considering the probability of interference. In order to model the capacity compromise between different (groups of) trains, Mussone and Calvo [38] proposes a multiobjective model and technique for analyzing the absolute capacity of railway networks. This paper proposes the idea that capacity analysis should consider the different mixes of trains (i.e., passenger versus freight, competitive network corridors, and particular train types). Still, this timetable-free method only considers the capacity at the strategic level and neglects a more detailed traffic mix.

1.3. Contribution Statements. This paper extends the scope of capacity estimation from simple linear lines to networks. We design a multiobjective capacity estimation approach in this paper, including a level of detail of optimized precise timetabling. Compared with Jensen et al. [15] where generating the timetable structure by branch-and-bound or tabu search, we generate saturated timetables that run a maximum number of trains based on a set of different objectives. The Pareto front of the optimized solutions corresponds to a compact representation of the competition of trains with different routes and stopping patterns on the railway network on the timetable level (compared with Mussone and Calvo [38], which describes the multiobjective capacity estimation problem on train flow level). The major contributions of our paper are based on a timespace network timetable saturation model, and we extend the multiobjective capacity estimation approach to the detailed timetable level. The multiobjective capacity estimation approach can estimate the railway capacity of various train path combinations without enumerating timetable structures. With  $\varepsilon$ -constraint method, we can obtain the Pareto front of saturated timetables, i.e., a set of nondominated optimized timetables that no more candidate train can be additionally scheduled. The result is a more comprehensive representation of capacity than a single absolute scalar number representing the simple amount of trains that would be able to run or the percentage of capacity used.

#### 2. Problem Description

2.1. Timetable Saturation for Capacity Estimation. Railway capacity is defined as the maximum number of trains that can run on a given railway network during a specific period. When the traffic pattern has the highest level of freedom, we can get the theoretical capacity, which reflects the maximum number of trains that can be scheduled under very homogeneous traffic. When the traffic pattern is predefined, the practical capacity can be estimated by saturating a timetable with a given candidate train set, in which the trains satisfy the predefined traffic pattern (i.e., the predefined routes and stopping patterns).

In order to describe clearly the impact of the mixed traffic on capacity, we give an illustrative example, as shown in Figure 1.

Figure 1(a) shows that trains with different routes on a railway network are likely to run through one or more shared sections. If these sections are the critical bottleneck sections where the capacity of the section is less than the requesting train numbers, the capacity competition of trains with different routes occurs. For example, trains with Route 1, 2, and 5 share section c-d where these trains have competition for the limited capacity. The three timetables in Figure 1(a) and other possible timetables not enumerated here are instances of the saturated timetable representing the full use of capacity. The capacity of a railway network is no longer appropriate to be represented by a single scalar value (i.e., the maximum number of trains) but as a set of saturated timetables with various combinations of trains with different routes.

Similar to the capacity competition between trains with different routes, Figure 1(b) shows that applying different stopping patterns might result in various capacities. The "direct" trains without intermediate stops guarantee the fastest running speed but are not friendly to the passengers visiting the intermediate stations. On the contrary, the "stopby-stop" trains can provide maximum accessibility but reduce the average speed of trains. Thus, it would be necessary to use mixed stopping patterns with various skip stops to maintain the average speed and accessibility. Thus, it is preferable to describe the capacity of the railway network with various stopping patterns by a set of saturated timetables with different combinations of various stopping patterns rather than a single maximum train number. With the saturated timetable, the practical capacity can approximate the number of successfully scheduled trains. It is worth mentioning that our work is not to build an optimal timetable in terms of the multidimensional performance indexes but to estimate the maximum number of trains under certain service quality by saturating timetable subject to the related constraints. This work often happens before the train timetabling procedure, even earlier than the train line plan is decided. The multiobjective capacity estimation aims to find a balance point between the service quality and the capacity utilization in a rough manner for those capacityintensive railway lines.

Applying the idea of timetable saturation, we can use an optimization model to maximize the total number of trains, with the elementary constraints of the train timetabling problem and the extra constraint for the predefined traffic pattern. Therefore, the traffic pattern is included in the constraint when we build the timetable saturation model. In the typical capacity estimation problem, the traffic pattern is given, such as the proportion of train routes, stopping patterns, as well as the timetable structure. The number of successfully scheduled trains can be referred to as the railway capacity under the given traffic pattern.

2.2. A Multiobjective Approach for Estimating the Mixed Train Traffic. The timetable saturation model can calculate the maximum number of trains under a certain traffic pattern constraint. However, this result can only reflect one capacity competition consequence under a particular traffic pattern. When estimating the practical capacity in different traffic patterns influenced by the various combinations of trains in categories, one available method is to change the constraint of the train traffic patterns in the timetable saturation model to get many saturated timetables with different traffic patterns. This approach is mainly used to study the impact of timetable parameters (e.g., running time and minimum headway) on the capacity. The other available method is to extend the objective function to multiobjective ones to maximize the number of trains in different groups simultaneously. The Pareto front of this multiobjective programming can reflect the competition relation between different groups of trains, thus describing the practical capacity globally in different but related traffic patterns (i.e., the gradient proportion of train routes). This approach mainly focuses on the characteristics of capacity performance under the competition of trains that share the same infrastructure elements.

In this study, we apply the latter solution approach, i.e., constructing a multiobjective train timetable saturation model to calculate the Pareto front of the maximum number of trains of different competition groups. With this multiobjective programming, we can obtain a Pareto front with many saturated timetable solutions for analyzing the practical capacity possibilities under the competition of different train groups. The multiobjective approach can figure out more saturated timetables with various traffic patterns, which can be referred to as the template for improving the capacity utilization in capacity-intensive railways.



FIGURE 1: An illustrative example of the impact of mixed traffic on capacity. (a) Different routes at sharing sections. (b) Different stopping patterns.

2.3. Saturated Timetable Performance Indexes. After obtaining the Pareto front (i.e., the possible saturated timetables), we need to select the most preferable saturated timetable by comprehensively analyzing the timetable performance according to the timetabling objectives from both operator- and passenger-centric points of view. According to Burdett [39] and Parbo et al. [40], the waiting time, travel time, delay uncertainty, and transfer maintenance should be considered in the passenger-centric train timetable performance evaluation. Thus, when building a timetable, one can directly deal with the passenger perspective by building the passenger-centric objective function and optimizing the structured timetable performance indexes to obtain an ideal timetable. However, in this paper, we estimate the timetable performance in a postevaluation manner, i.e., calculating the performance indexes of the timetables on the saturated timetable Pareto front. We propose two categories of timetable performance indexes, namely operation-centric and passenger-centric timetable performance indexes, which are listed in Table 1.

The calculation method for the above indexes is introduced in Section 4.2. According to the timetable performance indexes, the human-machine interactive preferable timetable decision method is discussed and introduced in Section 6.

# 3. The Multiobjective Timetable Saturation Model

3.1. Notations. We model the timetable saturation problem in the macroscopic level (stations, platform number, and segments) and neglect the detailed blocking section in stations and segments. The railway network applied in the paper can be referred to in Figure 2. The notations used in the paper are listed in Table 2.

3.2. Time-Space Network Model for Timetable Saturation. In this study, we propose a time-space-state network (with simplified platform assignment improvements compared to Parbo et al. [41] and Caprara et al. [42]) to describe the train movement for timetable saturation considering station track assignment. A continuous time-space path on the network represents each scheduled train path. The components of the time-space network are shown in Figure 3.

The time-space network shown in Figure 3 has the following nodes and arcs, describing the different procedures of train mobilities.

- (i) Origin or sink node: they are the origin and sink point of the train time-space flow, which are denoted by v<sup>O</sup><sub>f</sub> and v<sup>S</sup><sub>f</sub>, respectively.
- (ii) Arrival node: it represents the arrival event that happens at the moment t at station s for train f, denoted by  $v_f(s, t, A)$ .
- (iii) Departure node: it represents the departure event that happens at the moment *t* at station *s* for train *f*, denoted by  $v_f(s, t, D)$ .

Correspondingly, the time-space network has three types of arcs as follows.

(i) Virtual arcs: origin virtual arc is denoted by  $a_f(v_f^0, v)$  representing the train coming to the railway network. Sink virtual arc is denoted by  $a_f(v, v_f^S)$  representing the train missing from the railway network.

	1			
Timetable performance series	Operation-centric	Passenger-centric		
Performance indexes	<ul> <li>(i) Capacity</li> <li>(ii) Average train travel speed</li> <li>(iii) Heterogeneity</li> <li>(iv) Extra stopping time</li> <li>(v) Train departure shift</li> <li>(vi) Service frequency for stations</li> </ul>	<ul> <li>(i) OD coverage</li> <li>(ii) Average waiting time before boarding</li> <li>(iii) Travel time onboard</li> <li>(iv) Passenger needing transfer</li> <li>(v) Train loading factor</li> </ul>		

TABLE 1: Timetable performance indexes used in the paper.



FIGURE 2: The railway network structure.

- (ii) Train running arcs: representing that the train moves from node v (a departure node) to node v' (an arrival node), denoted by a<sub>f</sub> (v, v').
- (iii) Train dwelling arcs: denoted by  $a_f(v, v', k)$ , representing that the train dwelling from node v (an arrival node) to node v' (a departure node). Note that the train without stopping is denoted by a dwell arc with a length of 0.

The time-space-state network generation follows the following rules.

(1) Arrival and departure nodes: in order to reduce the number of time-space nodes, we only generate the arrival and departure nodes that fall within the possible scheduling time slot of the train as follows:

$$V_{f} = \left\{ v_{f}(s, t, \omega) \mid ET_{f} \le t \le LT_{f} \right\}, \quad \forall s \in S, \omega \in \{A, D\}.$$
(1)

(2) Train running arcs: the train running arc a<sub>f</sub>(v, v') can be generated between a pair of time-space nodes if v = v<sub>f</sub>(s,t,D) and v' = v<sub>f</sub>(s',t + τ, A) for any e(s,s') ∈ E and τ = RT<sub>f</sub>(e). Considering the acceleration and deceleration process, we have the following formula to calculate RT<sub>f</sub>(e) where e(s,s') ∈ E:

$$\mathrm{RT}_{f}(e) = R_{f}(e) + \mathrm{Acc}_{f}(s) \times \delta_{f}(s) + \mathrm{Dec}_{f}\left(s'\right) \times \delta_{f}\left(s'\right).$$
(2)

The stopping pattern remains unchanged in the train timetabling process. Thus, the variable  $\delta_f(s)$  remains constant when we generate the running arcs. The above formula considers the running time difference caused by train stopping in the time-space arc.

- (3) Train dwelling arcs: the train dwelling arc a<sub>f</sub> (v, v', k) can be generated between a pair of time-space nodes if v = v<sub>f</sub> (s, t, A) and v' = v<sub>f</sub> (s, t + τ, D) for any s ∈ S and DT<sup>min</sup><sub>f</sub> (s) ≤ τ ≤ DT<sup>max</sup><sub>f</sub> (s), and k ∈ K<sub>s</sub>. Note that if the train is designated to pass through station s without stopping, τ is set to 0. The platform assignment is considered in a dimension of the label of the dwelling arc.
- (4) Virtual arcs: we specify a departure time window for each train to limit the layout flexibility of the train paths. The train departure time window constraint is considered when generating the origin virtual arcs. The origin virtual arc  $a_f(v_f^0, v)$  is generated where  $v = v_f(s, t, A) \quad \forall \text{EST}_f \le t \le \text{LST}_f.$

3.3. *Time-Space Resources*. We use the concept of *time-space resource* (Meng and Zhou [43]) to model the block section occupation. The time-space resource representation can implicitly denote minimum arrival, departure, or platform occupation headway restriction between two successive trains. Every running arc and dwell arc occupies a certain set of time-space resources (associated with block sections in segments or platforms in stations).

Notation	Description			
Elements and sets for infrastructures and trains	-			
$s \in S$	Station element and station set			
$e \in E$	Segment element and segment set			
$k \in K_s$	Platform track element and platform track set of station s			
$f \in F$	Train element and candidate train set			
Š <sub>f</sub>	Station set that train f traverse by			
S <sup>O</sup> S <sup>D</sup>	Origin and destination station of train $f$			
E.	The segment set that train $f$ traverse by			
$\omega \in \Omega$	Event type, $\Omega = \{A, D\}$ , "A" denotes arrival, and "D" denotes departure			
Flowents and sets for time space network				
t $\in T$	Time index (discrete time stamn)			
V	Generic node of the time-space-state network			
v <sup>O</sup> v <sup>S</sup>	Origin and sink virtual node of train $f$			
$v_f, v_f$ $v_c(s \pm \omega)$	Time-space node of train $f$ at station s at the moment $t$ with event $\omega$			
$V_f(s,t,w)$	Node of the time-space-state network belongs to train f			
	Generic arc of the time-space-state network			
$a_{x}(v^{O}, v) = a_{x}(v, v^{S})$	Origin and sink virtual arc			
$a_f(v_f, v), a_f(v, v_f)$	Running arc			
$a_f(v,v')$	Dwelling arc			
$A^+$ $A^-$	Arc set entering/leaving node $v$			
$A_{\nu}$	The arc set belongs to train f			
$\Delta s$	The dwelling arc set belongs to train $f$			
a a a a a a a a a a a a a a a a a a a	Train group categorized by multiobjective function			
y C	Train group set			
F	Train set that belongs to group $a$			
$r_g$	Time-space resource			
A	Arc set that occupies time-space resource $r$			
c	s value of train group a			
	e value of train group g			
Parameters (numbers)				
$[EI_f, LI_f]$	The time range of train operation			
$\operatorname{RI}_{f}(e)$	Running time of train f through segment e			
$R_f(e)$	Free flow running time through segment e			
$\operatorname{Acc}_{f}(s), \operatorname{Dec}_{f}(s)$	Acceleration and deceleration time loss for train stopping at station s			
$O_f(s)$	Binary indicator denoting whether train $f$ stops at station s. 1 for stop, 0 otherwise			
$[DI_{\overline{f}}^{m}(s), DI_{\overline{f}}^{m}(s)]$	The time range of train f dwells at station s			
$[\text{EST}_f, \text{LST}_f]$	The departure time window for train $f$			
Decision variables				
$z_f$	Binary variable, $z_f = 1$ if train f is scheduled, 0 otherwise			
$x_a$	Binary variable, $x_a = 1$ if the time-space arc <i>a</i> is selected, 0 otherwise			
c <sub>g</sub>	Total number of scheduled trains in train group $g$			
$\lambda_r$	The Lagrangian multiplier for time-space resource r			
$ ho_g$	The Lagrangian multiplier for train group $g$			
Elements and sets for timetable performance indexes	calculation			
$F^*$	Successfully scheduled train set			
Р	Passenger set			
$P^*$	Assigned passenger set			
F	Train set that in which the passenger $p$ assigned			
$P_{\ell}^{e}$	Passenger set that travels with train $f$ at segment $e$			
$f_{p}^{\prime}(n)$	The <i>n</i> th train during passenger $p'$ journey			
FC <sub>p</sub>	Available train connection set for passenger $p$			
FCon <sup>*</sup>	Successive train pair set running at segment e			
Parameters (numbers) for timetable performance inde	xes calculation			
arr <sup>s</sup> , dep <sup>s</sup>	Arrival and departure time of train $f$ at station s			
<i>i</i>	Binary indicator 1 denotes that there exists a train service from c to c			
Js <sub>1</sub> ,s <sub>2</sub> 1	The length of segment $a$			
ESI <sub>p</sub>	The earliest departure time of passenger p			

TABLE 2: Notations.



FIGURE 3: Time-space network considering platform assignment.

For railway segments, a train running arcs occupies a series of time-space resources according to the layout of block sections in the segment, as shown in Figure 4. For a running arc, the starting and ending time for occupying a block section can be calculated according to the blocking time theory. Note that with the predefined train running time, many of the block sections are redundant (i.e., the trains will never have occupation conflict in these block sections if they have no occupation conflict in other bottleneck block sections) and can be neglected without losing the feasibility. The departure and arrival headways between two successive trains are guaranteed if the selected train running and dwell arcs are conflict-free with respect to the time-space resources occupation. For example, in Figure 4, train  $f_1$  and  $f_2$  running through the segment with block section  $b_1, b_2$ , and  $b_3$ . The train running arcs of the two trains occupy a series of predetermined time-space resources. The minimum departure and arrival headway (6 and 5, respectively) is guaranteed if the occupation does not overlap. The details of the representation can be referred to Liao [44].

For the block section associated with platforms, a train dwelling arc occupies a series of time-space resources. The time interval between the departure time of the former train and the arrival time of the latter train is described by the occupation of the time-space resources of the platform. For example, in Figure 5, as train  $f_1$  and  $f_2$  use the same platform, train  $f_2$  can only arrive at the platform 3 time instances after train  $f_1$  leave the platform. This headway can be guarantee with the restriction of occupation overlapping.

With the time-space network introduced above, the timetable saturation problem can be modeled as finding the maximum number of train paths selected in the time-space network without time-space resource occupation conflict (i.e., every time-space resource is nonoccupied or only occupied by one arc).

3.4. Multiobjective Optimization Model for Saturating the *Timetable*. For saturating the timetable to the maximum number of trains, a network flow-based integer programming model is proposed to maximize the number of trains that are scheduled. Based on the multiobjective



FIGURE 4: Block section-related time-space resource representation in segment.



FIGURE 5: Block section-related time-space resource representation for platform.

characteristic analysis in Section 2.2, the multiobjective function of the timetable saturation problem can be written as follows:

$$\text{maximize}_{g \in G} \left\{ \sum_{f \in F_g} z_f \right\}, \tag{3}$$

s.t.

a

$$\sum_{a_f \in A_{\nu}^+} x_{a_f} = \sum_{a_f \in A_{\nu}^-} x_{a_f}, \quad \forall f \in F, \nu \in V_f,$$
(4)

$$\sum_{\substack{f \in A_{v_f}^+ \\ v_f}} x_{a_f} = \sum_{\substack{a_f \in A_{v_f}^- \\ v_f}} x_{a_f} = z_f, \quad \forall f \in F,$$
(5)

$$\sum_{a_f \in A_r} x_{a_f} \le 1, \quad \forall r \in R,$$
(6)

 $z_f \in \{0, 1\}, \quad \forall f \in F, \tag{7}$ 

$$x_a \in \{0, 1\}, \quad \forall a \in A. \tag{8}$$

In the objective function (3), *G* is the train group set, which can be defined according to the train competition that needs to be studied (i.e., grouping the trains by routes or stopping patterns). The objective function is to maximize the number of trains for each train group simultaneously. Formulations (4) and (5) are the flow balance constraints. If a train is scheduled ( $z_f = 1$ ), the train must have a continuous time-space path from its origin node to its sink node.

#### Journal of Advanced Transportation

Formulation (6) is the time-space resource occupancy constraint described in 3.3. It ensures that each time-space resource can only serve one train. The type of headway it denoted depends on the type of block section that the time-space resource r is associated with. If the time-space resource is associated with a block section in segments, it denotes the arrival or departure headway. Otherwise, it denotes the headway of trains using the same platform. Formulations (7) and (8) indicate the domain of the variables.

3.5.  $\varepsilon$ -Constraint Reformulation. This integer programming model is a multiobjective programming model. We apply an  $\varepsilon$ -constraint reformulation to convert the multiobjective optimization model to a single-objective one. By this transformation, we can calculate the Pareto front of the model, showing all possibilities of the saturated timetable generated under the competition of trains in different groups. The model of  $\varepsilon$ -constraint reformulation can be written as follows:

maximize 
$$c_g = \sum_{f \in F_g} z_f$$
  
s.t.  
 $c_{g'} = \sum_{f \in F_{g'}} z_f \ge \varepsilon_{g'} \forall g' \in G - g.$   
Formulation (4)–(8).

By enumerating the possible value of  $\varepsilon_{g'}$ , we can compute the Pareto front of the multiobjective programming. The capacity of the railway line can be represented by a solution of the objective programming, namely,

$$\left(c_{g_1}, c_{g_2}, \dots, c_{g|G|}\right). \tag{10}$$

The Pareto front is the set of the dominant optimal solutions, namely,

$$PF = \left\{ \left( c_{g_1}, c_{g_2}, \dots, c_{g|G|} \right)^1, \dots, \left( c_{g_1}, c_{g_2}, \dots, c_{g|G|} \right)^n \right\},$$
(11)

in the Pareto front set, each element represents a saturated timetable. The entire Pareto front *PF* represents all possible saturated timetables considering the competition between the groups  $g \in G$ .

#### 4. Solution Approach

4.1. Solving the Timetable Saturation Problem. The proposed MIP model with  $\varepsilon$ -constraint can be solved by commercial solvers (e.g., CPLEX and Gurobi). However, obtaining a solution in a reasonable computational time for large-scale problems is difficult. Therefore, we apply a Lagrangian relaxation algorithm with the intensity-based heuristic proposed by Meng and Zhou [43] and make the following modification to the algorithm framework to adapt to the extra-added  $\varepsilon$ -constraint.

The Lagrangian relaxation reformulation of the  $\varepsilon$ -constraint programming can be written as follows:

$$\operatorname{minimize}_{\lambda,\rho}\left\{\operatorname{maximize}\sum_{f\in F_{g_1}} z_f - \sum_{r\in R} \lambda_r \left(\sum_{a_f\in A_r} x_{a_f} - 1\right) + \sum_{g'\in G-g_1} \rho_{g'} \left(\sum_{f\in F_{g'}} z_f - \varepsilon_{g'}\right)\right\}.$$
(12)

Subject to formulation (4) and (5), and (7) and (8), the objective function can be reformulated to train-based items as follows:

$$\operatorname{minimize}_{\lambda,\rho}\left\{\operatorname{maximize}\sum_{f\in F} \left[\phi_f z_f - \sum_{a_f \in A_f} \delta_a x_a\right] + \sum_{r\in R} \lambda_r - \sum_{g'\in G-g_1} \rho_{g'} \varepsilon_{g'}\right\},\tag{13}$$

where

$$\begin{split} \delta_{a} &= \sum_{r \in R: a \in A_{r}} \lambda_{r}, \\ \phi_{f} &= \begin{cases} 1, & f \in F_{g_{1}}, \\ \rho_{g'}, & f \in F_{g'}. \end{cases} \end{split}$$
(14)

With this reformulation, the Lagrangian relaxation problem can be referred to as a series of train-based shortest path subproblems for each train f, thus can be solved efficiently by directed graph shortest path algorithm (e.g., topological ordering). The Lagrangian relaxation programming can be solved according to Meng and Zhou [43]. The fundamental solution procedure is shown in Algorithm 1.

Output: a set of saturated timetables. Step 1. Initialization Generate the time-space network according to the conditions introduced in Section 3.2. Step 2. Calculate utopian point Calculate the utopian point: solving the following single-objective programming for  $q \in G$  by the Lagrangian relaxation algorithm introduced by Meng and Zhou [43]. maximize  $\varepsilon_g^{\max} = \sum_{f \in F_g} z_f$ Formulation (4)–(8) With the utopian point  $(\varepsilon_{g_1}^{\max}, \varepsilon_{g_2}^{\max}, \varepsilon_{g_3}^{\max}, \dots, \varepsilon_{|G|}^{\max})$ , enumerate all possible  $(\varepsilon_{g_2}, \varepsilon_{g_3}, \dots, \varepsilon_{|G|}) \in \mathcal{E}$ , for all  $g \in G$  and  $0 \le \varepsilon_g \le \varepsilon_g^{\max}$ . Step 3. Lagrangian relaxation for train shortest path subproblems Given  $(\varepsilon_{g_2}, \varepsilon_{g_3}, \ldots, \varepsilon_{|G|})$ , solve the Lagrangian relaxation problem iteratively by the shortest path algorithm introduced by Meng and Zhou [43]. The Lagrangian relaxation solution can be obtained from the Lagrangian relaxation dual problem. Step 4. Heuristic method for fixing the Lagrangian relaxation solution Execute the intensity-based train-by-train scheduling heuristic introduced in Meng and Zhou [43] to get a feasible solution from Step 3. During the train-by-train scheduling procedure, check whether satisfying the ε-constraint if the train is successfully scheduled before scheduling a train. If the  $\varepsilon$ -constraint is violated, abandon the train and turn to the next train. Step 5. Update  $\varepsilon$ Turn to the next  $(\varepsilon_{g_2}, \varepsilon_{g_3}, \dots, \varepsilon_{|G|}) \in \mathcal{E}$ , go to Step 3. Step 6. Output timetables Output the saturated timetables and the associated successfully scheduled train sets  $F^*$ .

ALGORITHM 1: Solving the epsilon constraint multiobjective programming.

4.2. Saturated Timetable Performance Indexes Calculation. After generating the saturated timetables, we can comprehensively estimate the timetable performance quality by the performance indexes listed in the following. The comparison between these indexes, especially the capacity, and the others can be very helpful for timetable schedulers to realize what level the capacity can realize and what level of timetable performance can be achieved with the competition of categorized trains. In order to estimate the performance of the timetable in both operational and passenger-centric aspects for all saturated timetables (i.e., the dominant optimal solutions in the Pareto front PF), we define the following indexes to evaluate the saturated timetable, helping the timetable scheduler to evaluate and balance the capacity utilization and the railway traffic quality. The timetable performance indexes of a given saturated timetable can be classified into two categories as follows.

Input: The candidate train set F, and the necessary timetabling parameters.

4.2.1. Operation-Centric Timetable Performance Indexes. The successfully scheduled trains form a train set  $F^*$ . The arrival and departure time of train f at station s can be parsed from the solution of the time-space network model as follows:

$$\operatorname{arr}_{f}^{s} = \sum_{a_{f}(v(s,t_{1},A),v(s,t_{2},D))\in A_{f}: x_{a_{f}}=1} x_{a_{f}} \times t_{1},$$
  
$$\operatorname{dep}_{f}^{s} = \sum_{a_{f}(v(s,t_{1},A),v(s,t_{2},D))\in A_{f}: x_{a_{f}}=1} x_{a_{f}} \times t_{2}.$$
(15)

The following operation-centric timetable performance indexes can be calculated with a given saturated timetable.

(1) Capacity. The railway capacity is the total number of trains on the saturated timetable as follows:

$$\sum_{f \in F^*} z_f,\tag{16}$$

where *C* is the railway capacity. Moreover, we can separately calculate the total number of different groups of trains to better reflect the detailed train combination on the saturated timetable.

(2) Average Train Travel Speed. The train travel speed is the average speed during its entire journey, including the running and dwelling time. Train travel speed reflects the train operation efficiency from the perspective of railway operators and can be calculated as follows:

$$\frac{\sum_{f \in F^*} \sum_{e \in E_f} l_e}{\sum_{f \in F^*} \operatorname{arr}_f^{\mathrm{D}} - \operatorname{dep}_f^{\mathrm{O}}},\tag{17}$$

where  $\bar{\nu}$  is the average train travel speed of the saturated timetable.

(3) Heterogeneity. The heterogeneity of the saturated timetable reflects the capability of recovering to its normal condition while facing unexpected disturbances or disruptions. In general, the higher heterogeneity results in more significant capacity loss. We apply the following definition of heterogeneity to indicate the potential stability and robustness of the timetable.

$$\frac{\sum_{(s,s')\in E} \left(\sum_{(f,f')\in FCon_e^*} \left| \left(dep_f^s - dep_{f'}^s\right) - \left(arr_f^{s'} - arr_{f'}^{s'}\right) \right| / |FCon_e^*| - 1 \right)}{|E|}.$$
(18)

In the formulation,  $FCon_e^*$  can be generated with a given timetable by ordering the trains according to their departure time at station *s* for each segment e(s, s'). In the ordered train list, a pair of successive trains can compose (f, f') and to be included in  $FCon_e^*$ . It is calculated according to the difference between the actual arrival and departure headway of two successive trains running through the same segment.

(4) *Extra Stopping Time*. Extra stopping time is the difference between the actual train dwell time and the planned train dwell time. This parameter reflects the extra time loss while maximizing the capacity. A train might allow a faster train to overtake by extending its dwell time to increase the overall capacity. The total extra stopping time can be calculated as follows:

$$\sum_{f \in F^*} \sum_{s \in S_f} d_f^s - a_f^s - DT_f^{\min}(s).$$
(19)

(5) The Time Shift between Actual and Desired Departure Time. This parameter reflects the difference between the desired scheduling timeslot and the actual one. While increasing the capacity, the train paths might be redistributed, thus deviating from their desired timeslot. The total time shift between actual and desired departure time can be calculated as follows:

$$\sum_{f \in F^*} \deg_f^{\mathcal{O}} - \mathrm{EST}_f.$$
(20)

(6) Service Frequency for Stations. This parameter indicates the number of trains that stop at the station. It would be useful to estimate the relationship between capacity and the workload of the station. The total service frequency for stations can be calculated as follows:

$$\sum_{f \in F^*} \sum_{s \in S_f} \left( z_f \times \delta_f^s \right).$$
(21)

4.2.2. Passenger-Centric Timetable Performance Indexes. The passenger-centric timetable performance indexes might need to be calculated with a saturated timetable and a passenger-to-train assignment result. Therefore, we design a simple agent-based passenger assignment algorithm to assign the passengers to trains before calculating the passenger-related timetable performance index. With a given saturated timetable and time-dependent passenger OD matrix, we apply a random-sequence passenger assignment algorithm, applying the first-come-first-serve principle with the following assumption for simplifying the passenger flow assignment procedures.

- (1) The passengers depart within their desired departure time window.
- (2) If more than one train is available in the departure time window, the passenger prefers the faster train (the shortest total travel time from his or her origin to destination).
- (3) We assume that passengers have at most one transfer during the entire journey.
- (4) The trains have maximum loading factors. The passenger cannot be assigned to a train that has already reached its loading limitation.

The detailed passenger-to-train assignment procedure is shown in Algorithm 2.

With this approach, we can approximately estimate passengers' satisfaction by calculating the passenger-centric timetable performance indexes with given saturated timetables. Note that other sophisticated passenger assignment methods can replace this approach (e.g., simulation approaches considering passenger choice behavior and seat reservation strategy), obtaining a more accurate passengerrelated performance index.

In this paper, we apply the following passenger-centric timetable performance indexes.

4.2.3. OD Coverage. OD coverage denotes the total amount of the OD pairs that the saturated timetable can serve. The saturated timetable might abandon some OD services to increase the capacity due to the neglecting of intermediate stops. This measurement might reduce the passenger utilities, especially for those travelers between intermediate stations.

$$\sum_{s_1 \in S} \sum_{s_2 \in S - s_1} j_{s_1, s_2}, \tag{22}$$

where

$$j_{s_1,s_2} = \begin{cases} 0, & \forall f \in F^*, \delta_f^{s_1} = 0 \lor \delta_f^{s_2} = 0, \\ 1, & \text{otherwise.} \end{cases}$$
(23)

4.2.4. Average Passenger Waiting Time before Boarding. This waiting time at its origin station is a very important index to passenger experience, as passengers are most sensitive to this part of time loss.

**Input:** A given saturated timetable, passenger set P (each passenger agent  $p \in P$  has his/her origin station  $s_p^0$ , destination  $s_p^0$ , and their desired departure time  $EST_p$ ). **Output:** passenger-to-train assignment result  $F_p$  and  $P_f^e$ . Step 1. Initialization Sort the passengers  $p \in P$  in random order. Step 2. Direct passenger-to-train assignment Foreach  $p \in P$ For each  $f \in F^*$ If  $\delta_f^{s_f^{\cup}} = \delta_f^{s_p^{\cup}} = 1$  (train f can serve the passenger OD) and  $\text{EST}_p \le \text{dep}_f^{s_p^{\cup}}$  (passenger p can get on train f) and the loading limitation of train f is satisfied  $F_p \coloneqq F_p \cap \{f\}$ If  $F_p = \emptyset$  Then  $P_0 := P_0 \cap \{p\}$ , and go to Step 3. Else select the  $f \in F_p$  with the earliest departure time at station  $s_p^0$ , and set  $F_p = \{f\}$ , and go to Step 4. Step 3. Transfer passenger-to-train assignment For each  $p \in P$ For each  $f_1 \in F^*$  and  $s_p^{O} \in S_f$ For each  $f_2 \in F^*$  and  $s_p^{D} \in S_f$ If passenger p can finish his/her journey by transferring from train  $f_1$  to train  $f_2$ , let train pair set  $FC_p := FC_p \cup (f_1, f_2)$ . If  $F_p = \emptyset$  Then Else select the  $(f_1, f_2) \in FC_p$  with the earliest departure time at station  $s_p^0$ , and set  $F_p = \{f_1, f_2\}$ . Step 4. Output passenger assignment result Output the passenger assignment result: For each passenger p, the assignment train set  $F_p = \{f_p(1), f_p(2), \dots, f(|F_p|)\}$ . For each train f and each segment e, the loading passenger set  $P_f^e$ .

ALGORITHM 2: FCFS passenger-to-train assignment algorithm.

$$\frac{\sum_{p \in P^*} de p_{f_p(1)}^{s_{f_p(1)}^0} - \text{EST}_p}{|P^*|},$$
(24)

where  $f_p(1)$  is the first train that passenger p takes.

4.2.5. Average Travel Time Onboard. The average travel time onboard can be calculated as follows:

$$\frac{\sum_{p \in P^*} \arg_{f_p}^{s^{D}}(|F_p|) - \deg_{f_p(1)}^{s^{D}}}{|P^*|}$$
(25)

4.2.6. Number of Lost Direct Passengers. Due to the loading limitation, some passengers must finish their journey by transfer. The number of lost direct passengers is a direct measurement of passenger satisfaction. This index can be calculated according to the passenger assignment result.

4.2.7. Average Loading Factor. The average loading factor is the index that reflects the crowdedness of the train, which describes the level of service onboard.

$$\frac{\sum_{f \in F^*} \left( \sum_{e \in E_F} \left| P_f^e \right| / \left| E_f \right| \right)}{\left| F^* \right|},\tag{26}$$

where  $P_f^e$  is the passenger set in which the passenger traverses through the segment *e* by train *f*.

#### 5. Case Study

5.1. Experiment Setup. The multiobjective analyses are conducted to study the competition of capacity based on different routes and stopping patterns. All of the instances are conducted on a personal computer with an AMD R9-5900X CPU, and 64 GB of internal memory. The instances of the mathematical programming are solved by Gurobi 9.5 with default settings invoked by a C# program (for implementing the  $\varepsilon$  constraint method). For each instance, the solving process is terminated when the optimal solution is obtained unless the solving time reaches the limit of 300 seconds. This means the entire Pareto front can be determined in a couple of hours maximum.

The data for the case study are extracted from the highspeed (HSR) and intercity (IC) railway network of Zhengzhou railway group of China Railway, as shown in Figure 6(a). The passenger flow data shown in Figure 6(b) are greater than the maximum carrying volume, which implies that some passengers might not be able to be transported due to capacity limitations. We design the following instances containing different amounts of trains to study the algorithm's performance, which are described with detailed configurations in Tables 3 and 4 for IC trains and HSR trains, respectively.

5.2. Comparison between Single- and Multiobjective Solution. In this section, we first compare the computational performance of Gurobi solver and the Lagrangian relaxation heuristic. Then, the comparison between single-objective and multiobjective capacity estimation results is shown to





FIGURE 6: The high-speed (HSR) and intercity (IC) railway network of Zhengzhou railway group. (a) Train lines. (b) Passenger flow from origins to destinations.

Instance (departure time horizon length, minute)	# Candidate trains						
	Categorized by lines			Categorized by stopping pattern			
	Zhengjiao IC	Zhengji IC	Zhengkai IC	Direct (0 inter. stop)	Skip-stop (≥1 inter. stop(s))	Stop-by stop	Total
60	10	10	12	12	12	8	32
120	30	40	38	40	42	26	108
180	53	64	60	65	75	37	177
240	62	90	86	92	101	49	238
300	69	106	96	103	115	53	271

TABLE 3: The instances for case study (IC).

TABLE 4: The instances for case study (HSR).

	# Candidate trains						
Instance (departure time horizon length, minute)	Cate	gorized by railway lines	Categorized by origins and destinations		T-4-1		
	Jingguang HSR (noncross line)	Xulan HSR (noncross line)	Cross line HSR	Terminating or originating	Passby	Total	
60	40	42	42	36	88	124	
120	80	74	84	66	172	238	
180	124	106	90	92	228	320	
240	146	124	110	124	256	380	
300	146	124	110	124	256	380	

display the capacity variant under different competition patterns.

As the commercial solver Gurobi can only solve the small instances, it is necessary to use the LR heuristic for solving the large-scale instances. Before applying the LR heuristic, the comparison of Gurobi and LR heuristic is conducted for small-scale instances to show the reliability and effectiveness of the LR heuristic. The results are shown in Table 5.

Table 5 shows the computational performance comparison between Gurobi solver and LR heuristic approach. As the train timetable saturation problem is difficult to solve, the Gurobi can only tackle small instances. However, the LR heuristic approach's computational performance shows that it can obtain high-quality solutions in small instances compared to the Gurobi optimal solution and can solve large-scale instances in an acceptable computational time.

5.3. The Capacity of the Mixed Route. In this section, we conduct a joint analysis between capacity and timetable performance indexes for the Pareto solutions under the competition between trains with different routes. We categorize the trains into three categories, namely *noncrossline* HSR train, crossline HSR train, and intercity train, by their routes.

For each instance, we get the Pareto front from the abovementioned solutions. In the instance of HSR-300 (competition of trains with different routes), we get 33 Pareto-optimal solutions (i.e., saturated timetables) to constitute the Pareto front, which is displayed in Figure 7 with blue dots. The grey lines in Figure 7 show the total number of trains. The related utopian point is also shown in Figure 7.

It can be concluded from Figure 7 that the Pareto front of the mixed train route shows the trade-off between crossline trains and crossline trains. When 97 crossline trains are scheduled, there is nearly no noncrossline train that can be scheduled. However, when 120 noncrossline trains are scheduled, there are still 34 crossline trains that can be scheduled. With the contour of the total train number, we can see that the total number of trains reaches the maximum (174 trains) when the number of crossline trains is 62, and the number of noncrossline trains is 112. The Pareto solution that is closest to the Utopian point is 62,112. It can be concluded from Figure 7 that some of the crossline trains seldom have capacity competition with noncrossline trains. However, nearly all noncrossline trains are likely to have capacity competition with crossline trains. Therefore, schedulers should fully use the noncompetitive capacity to schedule noncrossline trains.

Furthermore, we sample the timetables in the feasible domain (i.e., fix the numbers of crossline trains and noncrossline trains and optimize the passenger-related timetable performance indexes), located on the left-bottom side of the Pareto front. Figure 8 shows the timetable performance indexes (namely extra stop time, direct passenger loss, average loading factor, and average passenger waiting time) with a given number of crossline and noncrossline trains. The x-axis and the y-axis are the numbers of crossline trains and noncrossline trains, respectively, while the z-axis is timetable performance indexes.

From Figure 8, we can conclude that, in the capacityintensive railway (i.e., the passenger flow reaching the capacity limit of the railway), the performance indexes related to passenger accessibility have a positive correlation to the number of scheduled trains, such as the direct passenger loss (as Figure 8(b)). This is because the more trains are scheduled, the more passengers can be transported. If passenger accessibility is regarded as the most important evaluation metric of railway capacity utilization, the traffic manager prefers to build a timetable close to the Pareto front. However, for the performance indexes related to the service quality, such as the extra stop time (as Figure 8(a)), the average loading factor (as Figure 8(c)), and the average passenger waiting time (as Figure 8(d)), the phenomenon shows that full utilization of railway capacity might result in the deterioration of service quality. The findings of the results support the significance of the multiobjective capacity estimation study. This may help determine the appropriate train combination in capacity-intensive railways based on the evaluation of timetable performance indexes.

Besides, it can be observed in Figure 8 that, on the Pareto front, different combinations of crossline trains and noncrossline trains lead to various timetable performances. Therefore, we calculate the timetable performance indexes for all Pareto solutions based on the Pareto front. The timetable performance indexes, which show a strong correlation with the proportion of crossline trains, are reported in Figure 9. Each point represents a Pareto solution (saturated timetable). The size of the point is associated with the capacity (i.e., the total number of scheduled trains). The trendline and the corresponding confidence interval are also displayed in the figures.

From Figure 9, we can see that the total extra stop time has a linear negative correlation with the proportion of crossline train, while the total departure shift time and the average passenger waiting/onboard time shows a linear positive correlation with the proportion of crossline train. In terms of the passenger-centric timetable performance indexes, the OD coverage, the average loading factor, and the direct passenger loss show a complicated nonlinear correlation to the proportion of crossline trains.

Two conclusions can be drawn from the above data. First, considering several timetable performance indexes, such as extra stop time, there is an optimal crossline train proportion for OD coverage and indirect passenger. This can be a crossline and noncrossline train proportion reference while scheduling the train timetable. Besides, the passenger-related timetable performance indexes strongly depend on the distribution of the passenger OD matrix. As the non-crossline passenger is dominant to the crossline passenger in the given passenger OD matrix (63.27% of the passengers traveling by HSR are noncrossline passengers), the loading factor shows a monotone decreasing pattern with the increase of the crossline train proportion. Besides, the indirect passenger shows a "U" shape curve pattern, as a high

Instance	Init. obj. (# trains)		Best obj. (# trains)		Opt. gap (%)		Computational time (sec)	
	Gurobi	LR heuristic	Gurobi	LR heuristic	Gurobi	LR heuristic	Gurobi	LR heuristic
HSR-60	31	30	31	31	0.00	1.33	7	72
HSR-120	61	60	60	60	0.00	5.62	88	141
HSR-180	92	86	92	91	0.81	6.90	300	205
HSR-240	—	136	_	142	_	19.32	300	281
HSR-300	_	161	_	174	_	33.16	300	300
IC-60	48	40	48	46	0.00	3.32	13	84
IC-120	89	85	89	92	7.35	4.86	107	172
IC-180	—	131	—	138	_	13.88	300	257
IC-240	—	157	—	172	_	26.90	300	300
IC-300	—	216	—	230	—	38.52	300	300

TABLE 5: Computational performance comparison between Gurobi and LR heuristic (100 iterations) without  $\varepsilon$ -constraint.

-: cannot get any nonzero feasible solution within 300 seconds.



FIGURE 7: The Pareto front and the optimal solutions of the instances with different routes.





FIGURE 8: Timetable performance indexes with a given number of crossline train and noncrossline train. (a) Extra stop time. (b) Direct passenger loss. (c) Average loading factor. (d) Average passenger waiting time.





FIGURE 9: Timetable performance indexes and capacity by saturated timetables. (a) Total extra stop time. (b) Total departure shift time. (c) OD coverage. (d) Avg. passenger waiting/onboard time. (e) Average loading factor. (f) Direct passenger loss.

proportion of crossline trains might reduce the direct travel possibility of noncrossline passengers and vice versa. It can be surmised that when the noncrossline passenger and crossline passenger are unbalanced, these indexes might show a different distribution. These conclusions would be useful for timetable schedulers to schedule a reasonable proportion between crossline and noncrossline trains, considering the above timetable performance indexes when planning the train line service.

5.4. The Capacity of Mixed Stopping Pattern. This section also conducts a similar joint analysis for the Pareto solutions under the competition between trains with different



FIGURE 10: The Pareto front and the optimal solutions of the instances with different stop patterns.

stopping patterns. The trains can be classified into three categories according to their stopping pattern, namely *direct train, stop-by-stop train,* and *skip-stop train.* Based on this train categorization, we solve the multiobjective optimization to obtain the Pareto front of the competition of trains with different stop patterns, as shown in Figure 10.

Figure 10 shows the three-dimensional Pareto front of different categories of trains, which is like a "Y" shape. From Figure 10, we can see that when the number of the stopby-stop train is greater than 30, the shape of the Pareto front is the two branches of the "Y," as in this area, the major capacity competition is between direct trains and skip-stop trains. Each "branch" represents the dominant corresponding train type (the skip-stop train on the left and the direct train on the right). However, when the number of stop-by-stop trains is less than 30, the direct and skip-stop trains can almost reach their maximum value by sacrificing them. From the Pareto front, we can conclude that the railway network's capacity varies from mixed traffic. Lower stopping pattern heterogeneity results in higher capacity. The different mixed train traffic might result in different passenger satisfaction performances. The detailed performance is displayed in Figure 11.

Concerning the timetable performance indexes, the passenger-independent indexes, such as heterogeneity, departure shifts, and service frequency, have a remarkable correlation with the total number of trains that can be scheduled, namely the capacity. Specifically, the heterogeneity is shown to negatively correlate with the total number of trains, as achieving higher capacity requires a more compatible stopping pattern to make consecutive trains run closer. The correlation between heterogeneity and the total amount of trains is stronger in this case than in the route competition cases. This phenomenon implies that, in the capacity competition between trains with stopping patterns, the very important impact factor on capacity is the compatibility of train paths, which can be denoted by heterogeneity. Besides, we report the train departure shift time by the total number of trains and the proportion of direct trains (with no intermediate stop). Concerning the total amount of trains, the departure shift time shows a linear trend. However, concerning the proportion of direct trains, the departure shift time shows a trend of increasing then decreasing with the maximum departure shift time at the proportion of direct trains of 0.4.

Concerning the loading factor, there is a saddle when the proportion of direct trains reaches between 0.4 and 0.5. Specifically, in Figure 11(d), the loading factor of stopby-stop trains increases with the proportion of direct trains. However, the proportion of direct and skip-stop trains decreases and then increases as the corresponding proportion increases, as the loading factor depends on the number of trains available, the number of onboard passengers, and stopping pattern combinations. The passenger average waiting time reaches its maximum when the proportion of direct trains is about 0.4, as the high-capacity performance reduces the number of trains chosen for the passengers visiting intermediate stations.

These conclusions would be useful for balancing the number of direct trains (with fewer stops) and local trains to satisfy the passenger demand and increase the capacity.

#### 6. Discussions

Reviewing our research background, the timetable saturation approach is applied to figure out the possible maximum train combinations and the timetable performance for capacity-intensive railways. Thus, we only focus on the Pareto front of saturated timetables, and obtaining an optimized timetable in terms of operation cost and passengers' utility is not our first goal. In this paper, our research focus is on revealing the correlation between the maximum number of scheduled trains and the corresponding performance indexes. From the analysis, we can help traffic managers to find a better balance point between the scheduled train number and the timetable performance under the condition of fully utilizing the railway capacity.

However, the multiobjective timetable saturation method has the following limitations that need to be further investigated. Firstly, the number of objective functions is





FIGURE 11: Timetable performance indexes and capacity by saturated timetables. (a) Heterogeneity. (b) Departure shift time. (c) Average loading factor. (d) Avg. loading factor by train proportion. (e) Average passenger waiting time. (f) Service frequency.

limited. It would be very difficult to conduct the  $\varepsilon$ -constraint method to obtain the Pareto front with many objectives (i.e., more than three objectives). Besides, the saturated timetable cannot be used directly as the objective functions only consider maximizing the train number rather than the comprehensive timetabling evaluation metrics. How to use the saturation timetable solutions is a challenge and becomes an open question for future research.

To our best understanding of the railway capacity estimation, there are several potential directions to utilize the capacity evaluation results and the corresponding saturated timetables. The capacity estimation results can provide the upper bound of railway capacity for the train service planning stage. By reviewing the train service proposal from the perspective of capacity, the timetable scheduler can give feedback to the railway operators to make further amendments to their train service proposal. Besides, the capacity can be used in the train pool generation for the train timetabling procedure, as generalizing the candidate train pool is an essential measure to the practical train timetabling problem.

### 7. Conclusions

A multiobjective measurement of train capacity is proposed to extend the original railway capacity definition. This measurement overcomes the difficulty of railway capacity estimation when facing competition between different types of trains. In mixed traffic with different train routes and stopping patterns, railway capacity estimation is regarded as getting a set of saturated timetables where timetables with different train combinations are included. By analyzing this saturated timetable set, we can quantify realistic timetables and compare different capacity utilization strategies under dense railway traffic to better understand the capacity of railway networks. In the methodology aspect, we use a timespace network and a corresponding integer programming model for saturating the train timetable. An  $\in$  constraint method is used to obtain the Pareto solutions.

In the route competition case study, the analysis shows that the different combinations of trains with various routes might result in different capacity performances, and there are trade-offs between the maximum total amount of trains and other timetable quality indexes on both operators' and passengers' views. For the stopping pattern case study, we are able to quantify different Pareto-optimal saturated timetables from multiple points of view, addressing operator's and passengers' wishes. The Pareto front describes a large trade-off between the total amount of trains that can be scheduled and the service quality indexes that strongly depend on the stopping pattern design.

Such an analysis is very useful for determining service intentions, which can be related to a specific Pareto-optimal saturated timetable. This latter would be a reference for determining the capacity utilization of a production timetable. Resolving the trade-off between the various objectives is able to determine the most desirable heterogeneity level for the timetable.

In policy, the infrastructure manager can reasonably distribute the railway capacity (i.e., by arranging the time slots for train paths) to different users (passenger and freight operator companies) based on the capacity estimation results to balance the capacity competition among them. The railway passenger and freight operators can refer to the capacity estimation results to develop their product structures to better use the remaining unused capacity. The infrastructure managers can also make targeted infrastructure investments to increase the capacity of the bottlenecks in accordance with the capacity estimation results.

In practice, timetable schedulers can obtain the upper bound of the train numbers that satisfy the given quality performance indexes by this method and refer to these values to decide the candidate train pool (i.e., the train set that is ready to be scheduled) in the train line planning stage under certain service quality index requirements, focusing on using the railway capacity to the utmost extent. Besides, the corresponding timetable structures of the saturated timetable are also very useful for the traffic manager when scheduling the train paths on the dense railway. The traffic manager can refer to the structure of the timetable (i.e., the scheduling sequence and the corresponding overtaking arrangement) as a template to achieve such a high level of capacity utilization.

For further research, larger and more networked test cases can be studied. Moreover, the interactive multiobjective analysis would enable railway infrastructure managers to select and investigate attractive solutions on the Pareto front interactively and dynamically.

#### **Data Availability**

The data used to support the findings of this study are available from the corresponding author upon request.

#### **Conflicts of Interest**

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. U236820018) and the Fundamental Research Funds for the Central Universities (2023JBCZ005).

#### References

- R. L. Burdett, "Optimisation models for expanding a railway's theoretical capacity," *European Journal of Operational Research*, vol. 251, no. 3, pp. 783–797, 2016.
- [2] A. Landex, A. H. Kaas, B. Schittenhelm, and J. Schneider-Tilli, "Evaluation of railway capacity proceedings from the annual transport conference at," *Aalborg University*, vol. 13, no. 1, 2006.
- [3] Leaflet UIC, 406-Capacity, International Union of Railways (UIC), Paris, France, 2004.
- [4] J. Yuan and I. A. Hansen, "Optimizing capacity utilization of stations by estimating knock-on train delays," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 202–217, 2007.

- [5] A. F. De Kort, B. Heidergott, and H. Ayhan, "A probabilistic (max,+) approach for determining railway infrastructure capacity," *European Journal of Operational Research*, vol. 148, no. 3, pp. 644–661, 2003.
- [6] M. Abril, F. Barber, L. Ingolotti, M. A. Salido, P. Tormos, and A. Lova, "An assessment of railway capacity," *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 5, pp. 774–806, 2008.
- [7] J. Armstrong and J. Preston, "Capacity utilisation and performance at railway stations," *Journal of rail transport planning & management*, vol. 7, no. 3, pp. 187–205, 2017.
- [8] Y. C. Lai, Y. H. Liu, and Y. J. Lin, "Standardization of capacity unit for headway-based rail capacity analysis," *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 68–84, 2015.
- [9] N. Weik, N. Niebel, and N. Nießen, "Capacity analysis of railway lines in Germany-A rigorous discussion of the queueing based approach," *Journal of Rail Transport Planning & Management*, vol. 6, no. 2, pp. 99–115, 2016.
- [10] F. Rotoli, G. Malavasi, and S. Ricci, "Complex railway systems: capacity and utilisation of interconnected networks," *European transport research review*, vol. 8, no. 4, pp. 1–21, 2016.
- [11] R. M. Goverde, "Railway timetable stability analysis using max-plus system theory," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 179–201, 2007.
- [12] U. I. C. Leaflet, 406-Capacity (2<sup>nd</sup> Version), International Union of Railways (UIC), Paris, France, 2013.
- [13] A. Landex, Methods to Estimate Railway Capacity and Passenger Delays, Technical University of Denmark (DTU), Kongens Lyngby, Denmark, 2008.
- [14] A. Landex, "Evaluation of railway networks with single track operation using the UIC 406 capacity method," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 7–23, 2009.
- [15] L. W. Jensen, M. Schmidt, and O. A. Nielsen, "Determination of infrastructure capacity in railway networks without the need for a fixed timetable," *Transportation Research Part C: Emerging Technologies*, vol. 119, Article ID 102751, 2020.
- [16] L. W. Jensen, A. Landex, O. A. Nielsen, L. G. Kroon, and M. Schmidt, "Strategic assessment of capacity consumption in railway networks: framework and model," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 126–149, 2017.
- [17] R. M. Goverde, F. Corman, and A. D'Ariano, "Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions," *Journal of rail* transport planning & management, vol. 3, no. 3, pp. 78–94, 2013.
- [18] A. Jamili, "Computation of practical capacity in single-track railway lines based on computing the minimum buffer times," *Journal of Rail Transport Planning & Management*, vol. 8, no. 2, pp. 91–102, 2018.
- [19] J. Zhang, B. Han, and L. Nie, "Research on capacity calculation and assessment framework for Chinese high speed railway based on UIC406," *Journal of System and Management Sciences*, vol. 1, no. 6, pp. 59–75, 2011.
- [20] P. Pellegrini, G. Marliere, and J. Rodriguez, "Recife-sat: a milp-based algorithm for saturating railway timetables," *HAL*, vol. 2017, 2017.
- [21] H. Kim, I. J. Jeong, and D. Park, "Railway capacity allocation modeling using a genetic algorithm," *Transportation Research Record*, vol. 2608, no. 1, pp. 115–124, 2017.
- [22] J. Zhang, "Analysis on line capacity usage for China high speed railway with optimization approach," *Transportation*

Research Part A: Policy and Practice, vol. 77, pp. 336–349, 2015.

- [23] M. E. Petering, M. Heydar, and D. R. Bergmann, "Mixedinteger programming for railway capacity analysis and cyclic, combined train timetabling and platforming," *Transportation Science*, vol. 50, no. 3, pp. 892–909, 2016.
- [24] X. Zhang and L. Nie, "Integrating capacity analysis with highspeed railway timetabling: a minimum cycle time calculation model with flexible overtaking constraints and intelligent enumeration," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 509–531, 2016.
- [25] L. B. Reinhardt, D. Pisinger, and R. Lusby, "Railway capacity and expansion analysis using time discretized paths," *Flexible Services and Manufacturing Journal*, vol. 30, no. 4, pp. 712– 739, 2018.
- [26] M. Yaghini, M. Sarmadi, N. Nikoo, and M. Momeni, "Capacity consumption analysis using heuristic solution method for under construction railway routes," *Networks and Spatial Economics*, vol. 14, no. 3-4, pp. 317–333, 2014.
- [27] M. Yaghini, N. Nikoo, and H. R. Ahadi, "An integer programming model for analysing impacts of different train types on railway line capacity," *Transport*, vol. 29, no. 1, pp. 28–35, 2014.
- [28] S. Harrod, F. Cerreto, and O. A. Nielsen, "A closed form railway line delay propagation model," *Transportation Research Part C: Emerging Technologies*, vol. 102, pp. 189–209, 2019.
- [29] S. Harrod, "Capacity factors of a mixed speed railway network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 5, pp. 830–841, 2009.
- [30] A. Lindfeldt, Railway Capacity Analysis: Methods For Simulation and Evaluation of Timetables, Delays and Infrastructure Doctoral Dissertation, KTH Royal Institute of Technology, Stockholm, Sweden, 2015.
- [31] M. C. Shih, C. T. Dick, and C. P. Barkan, "Impact of passenger train capacity and level of service on shared rail corridors with multiple types of freight trains," *Transportation Research Record*, vol. 2475, no. 1, pp. 63–71, 2015.
- [32] M. H. Dingler, Y. C. Lai, and C. P. Barkan, "Impact of train type heterogeneity on single-track railway capacity," *Transportation Research Record*, vol. 2117, no. 1, pp. 41–49, 2009.
- [33] A. Nash and D. Huerlimann, "Railroad simulation using OpenTrack," WIT Transactions on The Built Environment, vol. 74, 2004.
- [34] H. Pouryousef and P. Lautala, "Hybrid simulation approach for improving railway capacity and train schedules," *Journal* of Rail Transport Planning & Management, vol. 5, no. 4, pp. 211–224, 2015.
- [35] A. P. Vieira, L. M. Christofoletti, and P. R. Vilela, "Analyzing railway capacity using a planning tool," ASME/IEEE Joint Rail Conference, vol. 50978, 2018.
- [36] R. L. Burdett and E. Kozan, "Techniques for absolute capacity determination in railways," *Transportation Research Part B: Methodological*, vol. 40, no. 8, pp. 616–632, 2006.
- [37] B. Bevrani, R. L. Burdett, A. Bhaskar, and P. K. Yarlagadda, "A capacity assessment approach for multi-modal transportation systems," *European Journal of Operational Research*, vol. 263, no. 3, pp. 864–878, 2017.
- [38] L. Mussone and R. W. Calvo, "An analytical approach to calculate the capacity of a railway system," *European Journal* of Operational Research, vol. 228, no. 1, pp. 11–23, 2013.
- [39] R. L. Burdett, "Multi-objective models and techniques for analysing the absolute capacity of railway networks,"

European Journal of Operational Research, vol. 245, no. 2, pp. 489–505, 2015.

- [40] J. Parbo, O. A. Nielsen, and C. G. Prato, "Passenger perspectives in railway timetabling: a literature review," *Transport Reviews*, vol. 36, no. 4, pp. 500–526, 2016.
- [41] J. Parbo, O. A. Nielsen, and C. G. Prato, "User perspectives in public transport timetable optimisation," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 269–284, 2014.
- [42] A. Caprara, M. Fischetti, and P. Toth, "Modeling and solving the train timetabling problem," *Operations Research*, vol. 50, no. 5, pp. 851–861, 2002.
- [43] L. Meng and X. Zhou, "Simultaneous train rerouting and rescheduling on an N-track network: a model reformulation with network-based cumulative flow variables," *Transportation Research Part B: Methodological*, vol. 67, pp. 208– 234, 2014.
- [44] Z. Liao, "Rescheduling out-of-gauge trains with speed restrictions and temporal blockades on the opposite-direction track," *Mathematics*, vol. 11, no. 12, p. 2659, 2023.