Article

A Stop-Probability Approach for O-D Service Frequency on High-Speed Railway Lines

Shuo Zhao [1,*], **Xiwei Mi** [2] **and Zhenyi Li** [3]

1   School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, China
2   School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; mixiwei@bjtu.edu.cn
3   Institute of Computing Technologies, China Academy of Railway Sciences, Beijing 100081, China;
    bimlizhenyi@rails.cn
*   Correspondence: zhaoshuo@csu.edu.cn

check for
updates

Abstract: Train stop planning provides appropriate service for travel demand and stations and plays a significant role in railway operation. This paper formulates stop planning from the point of view of direct travel between origin-destination (O-D) stations and proposes an analytical method to theoretically derive optimal service frequencies for O-D demand on different levels. Considering different O-D demand characteristics and train service types, we introduce the concept of stop probability to present the mathematical formulation for stop planning with the objective of minimizing per capita travel time, which is solved by an iterative algorithm combined with local search. The resulting optimal stop probabilities can be used to calculate the required service frequency for each train type serving different demand categories. Numerical examples, based on three real-life high-speed railway lines, demonstrate the validity of the proposed method. The proposed approach provides a more flexible and practical way for stop planning that explicitly takes into account the importance of different stations and passenger travel characteristics.

Keywords: High-speed Railway; Stop planning; Stop Probability; Per Capita Travel Time; Average Number of Train Services

1. Introduction

Recent years have seen an accelerated development of high-speed railway (HSR) in China and many other countries to meet the increasing demand for travel and the socio-economic developments in these countries. This rapid development poses huge challenges to railway infrastructure managers and operators to best manage large HSR networks and to produce the most effective operating plans. Railway planning generally follows a hierarchical top-down approach, starting from network planning, to line planning, timetable generation, train routing, rolling stock scheduling, and crew scheduling [1].

Within a given network, line planning becomes the first stage of the railway planning process, which is the process that determines the train routes, frequencies, and stop patterns in the network. Generally, train routes are determined first. Then, train service frequencies can either be optimized with the pre-determined stops or be simultaneously optimized with stops [2]. Alongside train route and frequency, train stop planning directly represents the service level to stations and for passengers.

An effective stop planning is to balance the trade-off between railway supply and passenger demand; the latter is usually characterized as passenger flow levels through a station, or between an origin-destination station pair. A stop plan with a large number of stops will clearly be able to serve more passengers more conveniently but at the expense of increased train travel times and energy consumption caused by acceleration and deceleration. On the other hand, a stop plan with fewer

stops will lead to longer waiting times for passengers but shorter train running times. Such practical supply-demand trade-offs increase the complexity of line planning or stop planning.

Most of the studies on line/stop planning have attributed travel demand to stations and categorize the station importance (or level) by the volume of passengers through the station—stations with higher degrees of importance correspond to higher passenger volumes. Then, the stop planning problem is to address the different service requirements for different stations. A commonly adopted approach in line/stop planning for station categories is to select from a set of lines with pre-determined routes and stop patterns that satisfies certain objectives [2–5]. Fu et al. [2] classified stations and trains into two levels and applied the concept of "possible line set" to define train routes and stop stations for line planning, where the intermediate stops were pre-set. Kaspi and Raviv [5] presented a model that combined line planning, timetabling, and routing decisions of the passengers. Again, they only selected lines from a predefined set of lines, each of which had its route and stopping stations pre-determined. A key drawback of the line set approach is that the candidate lines have fixed stopping stations. This makes it inflexible in line planning, and it may not always find the desired service frequencies for origin-destination (O-D) pairs.

To ensure the required level of service to a station, a minimum service frequency can be set based on the degree of importance of the station [6–12]. Yang et al. [6] proposed a mix-integer linear programming model to integrate train stop planning and train scheduling problems. The constraint that each station must be served by a given number of trains was considered. With given numbers and types of trains, Chen et al. [7] proposed a stop-schedule optimization model that jointly minimized the train stop cost and maximized passenger travel convenience. They considered the reasonable range of train stop number and the minimum service frequency for each station. In addition, there has also been extensive literature on the design of stop patterns to cater to different demand at stations [12–22]. Patterns include all-stop pattern, skip-stop pattern, zonal pattern, and express/local pattern, which cater to different demand characteristics (see a comparison in reference [13]). Each stop pattern has the specific applied area, so they are not suitable to be adopted for a general stop planning problem. Thus, some researchers also study a combination of several patterns [23–28]. However, the complexity and computation time required to solve a combined stop-pattern problem become significant as the number of variables to optimize increases, so this combined approach is not suitable to be applied to large networks unless an efficient algorithm can be found.

Overall, the above literatures in stop planning are concerned with the service levels at stations for simplification, i.e., the passenger volumes through stations. They do not specifically deal with passengers' O-D travel.

A good stop plan should consider travel characters and satisfy the passengers' direct travel requirements between their O-D stations. Different O-D demands need different levels of service, and a certain number of train services need to simultaneously stop at the origin and destination stations of each O-D pair, not single stations. Therefore, the ultimate purpose of stop planning is to provide a sufficient level of direct services for different O-D demand by determining the stop stations for each train on railway lines to reduce the transfer costs and travel times of passengers.

There have been studies, though limited, in the literature considering the direct choices of passengers in stop planning. Lin and Ku [29] constructed an integer program model to determine stop patterns and stop intervals for trains without passenger transfer. The objective was to maximize the profit earned by the train company, and the model was solved by two genetic algorithms. Lai et al. [30] designed an innovative stop planning network to formulate the stop sequences of trains and developed two binary integer programming models to tackle scenarios without and with passenger transfer. However, these two studies only considered a single O-D demand level and did not present a specific level of service (service frequency) for different O-D pairs. To ensure the defined level of service for O-D demand, Yue et al. [31] incorporated the service for major O-D demand in train stop schedule optimization, in which the major O-D demand were defined by the importance of their O-D stations. They set a constraint to ensure minimum percentages of direct services for these major O-D

pairs. The minimum service requirements for major O-D pair were given as input to the model, and the authors suggested that the values could be generated from empirical studies.

The decisions made in the above studies are in terms of whether a train is to stop at a station or not, of which the decision variables are discrete 0–1 variables from the point of view of service for single stations, and the required service level (i.e., the service frequency) is determined by empirical methods. Thus, there has not been a theoretical approach to analytically derive the service frequencies for different O-D demand levels. In practice, trains would not stop at each passing station. Stations on a higher level of importance would generally have more stopping trains; in other words, trains would have a higher probability to stop at a station with higher level of importance. Thus, we can formulate stop planning from another perspective. Different from the binary integer decision variables in the mentioned studies, we propose the concept of stop probability, i.e., the probability of a train stopping at a station that is approximately equal to the ratio of the number of stop trains to that of passing trains, which is a continuous variable to represent the relative importance of a station in stop planning. This service probability for stations or O-Ds is usually adopted in practical planning stage by empirical methods, when operators set required train percentages for major stations or O-D pairs [31].

This paper formulates stop planning from the point of view of direct travel between O-D stations. Considering different O-D demand levels and train service types, we introduce the concept of stop probability to present the general principle and mathematical formulation for stop planning with the objective of minimizing the per capita travel time, and determine the service frequencies for each O-D pair according to demand characteristics. The comparison to existing literature is shown in Table 1, and the key contribution of this paper is summarized below.

Table 1. Comparisons of modeling scenarios in stop planning.

Service Frequency	Decision Variables	Train Type	Key References
Station based	Integer	Single	Chen et al. [7]; Fu et al. [11]
Station based	Integer	Multiple	Fu et al. [2]; Yang et al. [6]; Ulusoy et al. [23]
O-D based	Integer	Single	Lai et al. [30]; Yue et al. [31]
O-D based	Integer	Multiple	Lin and Ku [29]
O-D based	Continuous	Multiple	This paper

First, we characterize railway stations in the HSR network by their geo-political importance, which also represents the attraction effects to travel demand. We consider spatial O-D movements that cross different, and movements within the same geo-political areas to present the travel characters of different O-D demand, and then define the O-D pairs jointly by the respective O-D station characteristics and by their spatial movements. Section 2 presents the detailed description of the geo-political classification of stations and the dual-level characterization of demand.

Second, we introduce the concept of stop probability of trains by station level and formulate the stop planning problem as to optimize trains' stop probabilities at stations on different levels. We show that the service frequency for a specific O-D pair can be derived directly from the stop probabilities at the respective O-D stations. We also consider two types of trains with different service availabilities—type 1 trains can stop at any station, while type 2 trains would only stop at higher level stations.

Third, we consider two additional constraints in the optimization to obtain feasible solutions—equilibrium of train load and equilibrium of stop density. Equilibrium of train load means that the proportions of train kilometers of the two train types should be similar to those of their transported passenger kilometers to keep a balance between the loads of the two types. To ensure the well-distributed stop stations within a geo-political area for the two train types, we adopt the concept of stop density, i.e., the average stop number in a province/district, to describe the distribution uniform degree of the stop stations of the train types, and keep it in a reasonable range.

We present the analytical formulations for the derivation of the optimal service frequencies for different O-D demand categories on an HSR line, which are important inputs for line planning

and stop planning. From the point of view of direct travel between O-D stations, we formulate the stop-probability based stop planning problem to minimize the per capita travel time. The resulting optimal stop probabilities are then used to calculate the number of trains required for each train type serving the different demand categories. A case study on three HSR lines in China validates the effectiveness and correctness of the approach, and the solved results show a more reasonable load division between the two train types.

The remainder of this paper is organized as follows. The stop probability optimization problem is formulated in Section 2, and the classification of stations and demand is also introduced in this section. Section 3 constructs a stop-probability based stop planning optimization model, and the solving method is also designed. The case studies of three HSR Lines are presented in Section 4, and conclusions follow in Section 5.

2. Problem Formulation

In this section, we first introduce the classification of stations, demand and trains, and then formulate the train stop probability optimization problem. Tables 2–4 show the input parameters and variables in this paper.

Table 2. Indexes.

Name	Definition
t	Index of train type
q	Index of representing O-D categories
i, j, h	Index of station level

Table 3. Input parameters.

Name	Definition
d	Average station distance, km
od_q	The qth O-D demand category
Q_c	Set of single-service demand categories
$Q_{\bar{c}}$	Set of double-service demand categories
l_c	Per capita travel distance of single-service demand, km
l_q	Per capita travel distance of od_q km
m_q	Average number of passing train services for the O-D pairs in od_q, train
S_h	Set of stations on level h on the HSR line
T	Operation period in a day, h
β_q	Proportion of demand who only choose trains of type 2 in the double-service demand category od_q, $q \in Q_{\bar{c}}$
λ_q	Proportion ofdemand in od_q
λ_c	Proportion of single-service demand
$\lambda_{\bar{c}}$	Proportion of double-service demand
θ_h	Proportion of level h stations
τ_{stop}	Average stop time at a station, including the acceleration and deceleration times, h
M	A large constant
ε_y	Threshold of the relative error between the train kilometers and transported passenger kilometers of train type 1
$n^1_{\text{stop},1}$	Upper limit of the difference between average numbers of stop stations in a province
$n^1_{\text{stop},2}$	Upper limit of the difference between average numbers of stop stations in a district

Table 4. Variables.

Name	Definition
p_q^t	Probability of trains of type t serving O-D pairs in od_q
$p_{h,1}^t$	Conditional probability of stopping at a station on level $h > 1$ in a province if the train stops at the provincial station
$p_{h,0}^t$	Conditional probability of stopping at a station on level $h > 1$ in a province if the train does not stop at the provincial station
n_q	Average number of train services for the O-D pairs in od_q, train
$T_{\text{stop},q}^t$	Per capita dwell time of passengers in od_q who choose trains of type t, h
$T_{\text{diff},q}^t$	Per capita boarding deviation time of passengers in od_q who choose trains of type t, h
$T_{\text{stop},c}$	Per capita dwell time of single-service demand, h
$T_{\text{stop},\bar{c}}$	Per capita dwell time of double-service demand, h
$T_{\text{diff},c}$	Per capita boarding deviation time of single-service demand, h
$T_{\text{diff},\bar{c}}$	Per capita boarding deviation time of double-service demand, h
α_q^t	Division proportion of passengers in od_q, $q \in Q_{\bar{c}}$ who choose trains of type t
H	Train headway
Decision variables	
x_h^t	Stop probabilities of trains of type t at stations on level h
$x_{h,1}^t$	Conditional stop probabilities of trains of type t at stations on level $h > 1$ if the train stops at the corresponding administrative station on level $h - 1$
$x_{h,0}^t$	Conditional stop probabilities of trains of type t at stations on level $h > 1$ if the train does not stop at the corresponding administrative station on level $h - 1$
y^t	Proportion of the train kilometers of train type t

2.1. Station Categorization

In this paper, the concept of stop probability represents the importance of stations on the network, which can also be presented by station levels. Thus, in order to present the station characters and reduce the scale of decision variables, we next characterize railway stations and categorize them into different levels, which also influence the attraction effects to travel demand.

Stations on an HSR network can be classified into different levels according to different criteria [17,27–29], including administrative level, location on the network, daily dispatched passenger number, the number of passing trains, etc. In China, an important classification criterion of stations is the geo-political (or geo-administrative) level of the locations of stations, which also represents the attraction to passenger demand. The government's administrative division follows a four-level structure—national, provincial (including municipalities), district, and county. Generally, a province includes the provincial city and several districts, which are administrated by the provincial capital. Each district has its own district capital city and its administrated counties. Figure 1 illustrates the standard structure of the station location within a province.

　　○ Provincial station　　□ District station　　△ County station
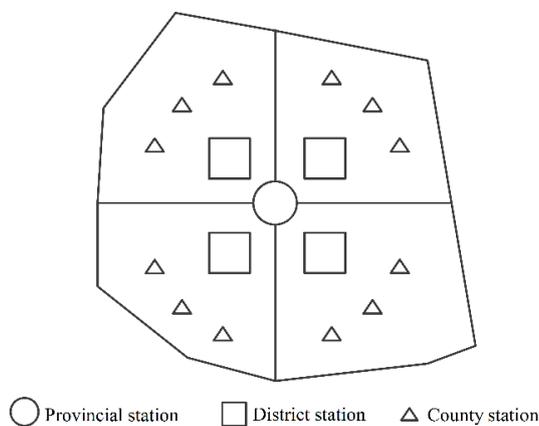
Figure 1. Illustration of the standard structure of the station location within a province.

In this paper, we classify stations into three levels according to their administrative division: provincial, district, and county level, and denote them as level $h = 1$, 2, and 3, respectively. The proportions of stations on the three levels are respectively denoted by $\theta_1, \theta_2, \theta_3$ satisfying $\theta_1 + \theta_2 + \theta_3 = 1$. The provincial station set includes the stations in the provincial capitals (different HSR stations in one city can be viewed as a combined single station), which is denoted by S_1, and the district and county station sets are denoted by S_2, S_3 respectively. As for a provincial station $s \in S_1$, the set of the district stations located in the corresponding province is denoted by $S_2(s) \subset S_2$; similarly, the set of the county stations located in the corresponding district to station $s \in S_2$ is denoted by $S_3(s) \subset S_3$.

2.2. O-D Demand Categorization

The above administrative structure influences the travel of demand and results in the attraction characters of provincial/district cities to travel demand. The attraction effects can be demonstrated in two aspects:

(1) High demand for travel across provinces or districts. Due to the high level of service for provincial stations, more trains would stop at these stations and a large number of passengers across provinces are generally attracted to get on/off at provincial stations to board direct trains to the destinations or find transfer travel schemes. Similarly, if there are a large number of county stations on the railway line, district stations would also attract a number of passengers across different districts to gather. Railway companies are required to provide sufficient train services to satisfy these direct travel requirements.

(2) Travel demand within a province or within a district. When a HSR line has been operated in a certain region such as a province or district, passengers traveling within that region often desire more direct train services to travel to/from the center city, i.e., the provincial or district city, for some political or economic reasons, even if the proportions may not be large enough. In order to better serve demand within a certain region, railway companies usually need to design the stop plan by taking provincial capitals and district cities as centers and improving the level of service for these passengers, i.e., increasing the service frequencies within a province or district to provide more convenience.

Therefore, stop planning should consider the attraction effects of stations on different levels and the travel characters of different O-D demand. The level of service for demand between the following O-D pairs should be mainly focused on: O-Ds between provincial stations and O-Ds traveling within a province or district. In this paper, we defined three movement types according to the O-D station location of passengers—(i) cross-province travel: passengers' origin and destination stations belong to different provinces; (ii) same-district travel: the O-D stations belong to the same district, including travel to/from the provincial station; and (iii) cross-district travel: the O-D stations belong to different districts of the same province.

Based on the above O-D movement types and station levels, we classify the travel demand into 13 categories of bi-directional travel demand. The 13 demand categories, denoted as $\{od_q, q = 1, 2, \cdots, 13\}$, can be expressed by a spatial-level matrix shown in Table 5, where i, j represent the levels of the O-D stations. The proportions of passenger volumes in the 13 demand categories are respectively denoted by $\lambda_q, q = 1, 2, \cdots, 13$, which satisfies $\lambda_q \geq 0$ and $\sum_{q=1}^{13} \lambda_q = 1$. These proportions can be collected in the set $\lambda = \lambda_1, \lambda_2, \cdots, \lambda_{13}$.

Table 5. The spatial-level matrix of 13 demand categories.

Movement Type	Cross-Province $(i,j)_{cp}$			Same-District $(i,j)_{sd}$			Cross-District $(i,j)_{cd}$		
Station level	1	2	3	1	2	3	1	2	3
1	od_1	od_2	od_3	–	od_7	od_8	–	–	–
2	–	od_4	od_5	–	–	od_9	–	od_{11}	od_{12}
3	–	–	od_6	–	–	od_{10}	–	–	od_{13}

2.3. Train Types

As the demand to/from a county stations are relatively low, in practice, not all trains need to stop at county-level stations. In this paper, we consider two types of trains: trains of type 1 can stop at any station, while those of type 2 would not stop at county stations.

Then the 13 demand categories can be further grouped in terms of the train service availability. Eight of the 13 demand categories are relevant to county stations and can therefore only be served by train type 1. We term this group of demand as the "single-service" demand category, and denote the set of these eight categories by $Q_c = \{3, 5, 6, 8, 9, 10, 12, 13\}$. The remainder five demand categories are grouped into set $Q_{\bar{c}} = \{1, 2, 4, 7, 11\}$, which is termed as "double-service" demand category. The proportions of the two kinds of demand are denoted by $\lambda_c = \sum_{q \in Q_c} \lambda_q, \lambda_{\bar{c}} = \sum_{q \in Q_{\bar{c}}} \lambda_q$.

Double-service demand have a choice of which train type to take. The proportions of passengers who only choose train type 2, denoted as β_q, $q \in Q_{\bar{c}}$ and satisfying $0 \leq \beta_q < 1$, can be obtained by observations and are considered as inputs to the model.

The proportions of the train kilometers of the two train types are denoted by $y^t, t = 1, 2$ satisfying $\sum_{t=1}^{2} y^t = 1$. Since the single-service demand can only choose train type 1, y^1 should be a positive value, and the two proportions need to satisfy $0 < y^1 \leq 1, 0 \leq y^2 < 1$. From the view of average, the proportions of the numbers of passing trains of type t at any station can also be viewed as y^t.

2.4. Train Stop Probability Optimization Problem

Generally, the service-oriented HSR line planning can be defined as follows.

Definition 1. *With the given HSR network allocation limit, passenger travel demand, and a predefined level of service, an optimal line plan should minimize the total travel time of passengers.*

If routes, frequencies and train types are provided, the only part to optimize in line planning is the stop plan. To solve this sub-problem of line planning, we propose the following assumptions for the train stop probability optimization:

(A1) In this paper, the train stop probability optimization only refers to a single HSR line.

(A2) As for any county or district station on the HSR line, its corresponding administrative district or provincial station is also on this line, which is illustrated in Figure 2. This assumption may not be absolute in reality, but it can be satisfied by merging stations, or upgrading/degrading the station level by the dispatched passenger volumes at each station.

○ Provincial station　　□ District station　　△ County station

▭ Single province　　══ High-speed railway line
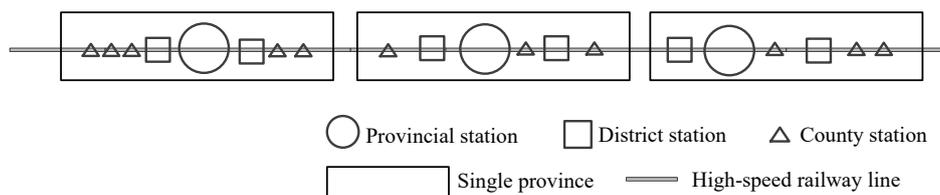
Figure 2. Illustration of assumption (A2).

(A3) All trains running on the line have the same capacity which can satisfy the predefined level of service.

(A4) The passenger demand of all O-D pairs follows the uniform distribution pattern during the whole operation period T.

(A5) The headway between any two adjacent trains is uniform.

(A6) In the planning stage, the ultimate purpose of stop planning is to provide sufficient direct services for different O-D pairs. Thus, the direct travel scheme is the only considered scheme in this paper.

(A7) Trains have independent stop probabilities in any two provinces, and the stop probabilities at provincial stations are independent and identically distributed at any two provincial stations. When the

stop at a provincial station is determined, the stop probabilities at any two administrated district stations in the corresponding province are independent and identically distributed. Similarly, after the stop at a district station is determined, the stop probabilities at any two county stations in the corresponding district are independent and identically distributed.

Thus, the train stop probability optimization problem can be defined as follows:

Definition 2. *Given the HSR line, routes and frequencies of trains, passenger demand, and assumptions (A1)–(A7), the stop planning can be described by the stop probabilities of the two train types at stations on different levels. The train stop probability optimization is to determine the train stop probabilities at stations on different levels and the reasonable proportion of the train kilometers of each type with the objective of minimizing per capita travel time. The outcome of the solved stop probabilities can be used to obtain the required average number of train services for O-D pairs in each demand category.*

The stop-probability approach follows a "stop probability → service probability → service frequency" process. As illustrated in Figure 3, given the input data, we adopt the stop probabilities to formulate stop planning with the objective of minimizing the per capita travel time, subject to constraints such as equilibrium of train load and stop density, and obtain the optimized stop probabilities of train types at different station levels. For an O-D pair, a train service should simultaneously stop at the origin station and destination station. The service probability for a demand category is deducted by the stop probabilities at the O-D stations that are determined by their station levels. Given the average number of passing trains of the demand category, we can further obtain the required average frequency for the O-D pairs in this category.
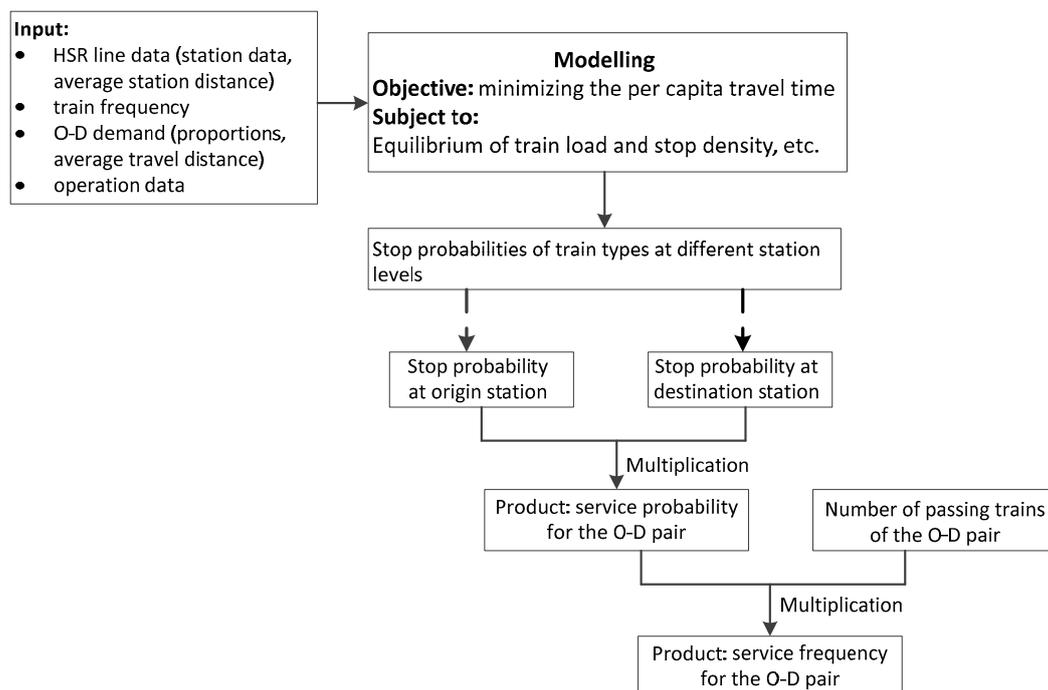
Figure 3. Illustration of the intuitional example.

3. A Stop Probability Based Stop Planning Optimization

3.1. Probability of Train Services by O-D Pair Categories

Let x_h^t denote the stop probabilities of train type t at stations on level h. We derive the probability of train services for each O-D category.

Due to the attraction effects and high levels of center stations within a region, the stop planning of stations in a district (or county) in a province (or district) is dependent on the train stopping at the corresponding provincial (district) station, and we adopt the conditional probabilities to describe the stop planning process between each two neighboring station levels. As for a station r on a lower level, which satisfies $r \in S_h(s), s \in S_{h-1}, h > 1$, the stop probability at r can be deducted in two situations—if a train of type t stops at the corresponding administrative station s, the conditional stop probability at r is denoted by $x_{h,1}^t$; otherwise, the conditional stop probability at r is denoted by $x_{h,0}^t$. Since trains of type 2 do not stop at county stations, $x_{3,1}^2 = x_{3,0}^2 = 0$. Thus, the stop probability of trains of type t at a station on a lower level $h > 1$ can be

$$x_h^t = x_{h-1}^t x_{h,1}^t + (1 - x_{h-1}^t) x_{h,0}^t \tag{1}$$

The average stop probability of trains of type t is

$$x^t = \sum\nolimits_{h=1}^{3} \theta_h x_h^t \tag{2}$$

In the range of a province, after the stop at the provincial station is determined, the stops at the district/county stations in that province have specific conditional probabilities, which can be deducted by the above conditional probabilities. If a train of type t has stopped at the provincial station, the conditional probability of stopping at a station on level $h > 1$ in that province is denoted by $p_{h,1}^t$; otherwise, it is $p_{h,0}^t$.

$$p_{h,1}^t = \begin{cases} x_{h,1}^t, h = 2 \\ x_{h-1,1}^t x_{h,1}^t + (1 - x_{h-1,1}^t) x_{h,0}^t, & h = 3 \end{cases} \tag{3}$$

$$p_{h,0}^t = \begin{cases} x_{h,0}^t, h = 2 \\ x_{h-1,0}^t x_{h,1}^t + (1 - x_{h-1,0}^t) x_{h,0}^t, & h = 3 \end{cases} \tag{4}$$

Similarly, if a train of type t has stopped at a district station, the conditional probability of stopping at a county station in that district is just $x_{3,1}^t$; otherwise, it is $x_{3,0}^t$.

We denote p_q^t as the probability of train type $t = 1, 2$, serving the demand category od_q, $q = 1, 2, \cdots, 13$. Next, we present the detailed deduction by the travel movement types of each demand category.

(1) Cross-province O-D movements

When $q = 1, 2, \cdots, 6$, the demand categories correspond to cross-province O-D movements, i.e., $od_q = (i, j)_{cp}$. Since the stop probabilities at the O-D stations are independent of each other, the probability of serving O-D pairs in each category is the product of the stop probabilities at the O-D stations.

$$p_q^t = x_i^t x_j^t, \qquad q = 1, 2, \cdots, 6, t = 1, 2 \tag{5}$$

(2) Same-district O-D movements

When $q = 7, 8, 9, 10$, the demand categories correspond to same-district O-D movements, i.e., $od_q = (i, j)_{sd}$. The stop probabilities at the O-D stations are not independent, and the stop probability at a station on a lower level in the O-D pair is related to the stop at the other station, we should consider the corresponding conditional probabilities of stopping at a station on level $h > 1$ in the range of a province or district.

$$\begin{cases} p_q^t = x_1^t p_{j,1}^t, & q = 7, 8, \ t = 1, 2 \\ p_9^t = x_2^t x_{3,1}^t, & t = 1, 2 \\ p_{10}^t = x_2^t x_{3,1}^t x_{3,1}^t + (1 - x_2^t) x_{3,0}^t x_{3,0}^t, & t = 1, 2 \end{cases} \tag{6}$$

(3) Cross-district O-D movements

When $q = 11, 12, 13$, the demand categories correspond to cross-district O-D movements, i.e., $od_q = (i, j)_{cd}$. The stop probabilities at the O-D stations are independent, but their stop probabilities both depend on the stop at the corresponding provincial station, so we should consider the corresponding conditional probabilities of stopping at a station on level $h > 1$ in the range of a province.

$$p_q^t = x_1^t p_{i,1}^t p_{j,1}^t + (1 - x_1^t) p_{i,0}^t p_{j,0}^t, \qquad q = 11, 12, 13, t = 1, 2 \tag{7}$$

With given train routes and frequencies, we can obtain the average numbers of passing train services for the O-D pairs in each demand category, and denote them by m_q, $q = 1, 2, \cdots, 13$, which can be collected in the set $\boldsymbol{m} = (m_1, m_2, \cdots, m_{13})$. For each demand category, the average numbers of train services of the two types serving each O-D pair is $y^t, t = 1, 2$, times of these numbers. Using Equations (5)–(7), the average number of train services for the O-D pairs in od_q, $q = 1, 2, \cdots, 13$, can be obtained by the following formula, and these service frequencies can be collected in the set $\boldsymbol{n} = (n_1, n_2, \cdots, n_{13})$

$$n_q = \sum_{t=1}^{2} y^t p_q^t m_q \tag{8}$$

3.2. Per Capita Dwell Time

Generally, more stop stations would result in longer passenger travel time, so the concept of per capita dwell time, i.e., the average passenger dwell time at intermediate stop stations, is a significant index of stop planning. We denote l_q as the per capita travel distance of od_q, $q = 1, 2, \cdots, 13$. Figure 4 illustrates the route of an O-D pair in od_q including the O-D stations and intermediate passing stations. If the demand in this category choose trains of type t, the corresponding per capita dwell time can be denoted by $T_{\text{stop},q}^t$, which is the product of the number of intermediate stop stations and the average stop time at a station τ_{stop}. In the view of average, using the average station distance d, we can obtain the number of intermediate passing stations is $l_q / d - 1$. Then, the average number of intermediate stop stations of this O-D pair is the product of the number of intermediate passing stations and the average stop probability at any station.
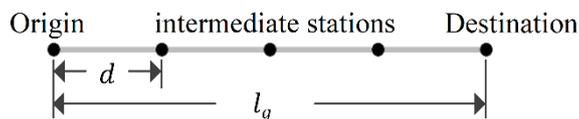
Figure 4. Illustration of the route of an O-D pair in W_q.

Using Equation (2), we can obtain the average stop probability of trains of type t at any station is $\sum_{h=1}^{3} \theta_h x_h^t$. Thus, the per capita dwell time of passengers in od_q choosing trains of type t is

$$T_{\text{stop},q}^t = \tau_{\text{stop}} (l_q / d - 1) \sum_{h=1}^{3} \theta_h x_h^t, q = 1, 2, \cdots, 13, t = 1, 2. \tag{9}$$

Since the single-service demand can only choose train type 1, the average dwell time of the single-service demand $T_{\text{stop},c}$ can be obtained if we replace l_q in Equation (9) by the per capita travel distance of the single-service demand l_c, i.e.,

$$T_{\text{stop},c} = \tau_{\text{stop}} (l_c / d - 1) \sum_{h=1}^{3} \theta_h x_h^1. \tag{10}$$

In each double-service demand category $od_q, q \in Q_{\bar{c}}$, except the passengers who only choose train type 2 (accounting for β_q), the remaining demand can choose either train type, so we need to calculate the division proportions of the two train types for these passengers. We denote the division proportions of passengers choosing the two train types in each double-service demand category by $\alpha_q^t, q \in Q_{\bar{c}}$, which satisfies $\sum_{t=1}^{2} \alpha_q^t = 1$.

From assumption (A4), passenger demand of all O-D pairs in each category follow the uniform distribution pattern during the whole operation period T, so the division proportions of double-service demand choosing the two train types are equal to those of attraction time lengths of the two train types in the whole operation period T.

Generally, passengers prefer the train type with less per capita dwell time, and a larger number of trains will also attract more passengers. Initial attraction time is the average attraction times of the two train types to each O-D pair in od_q, which are determined by the per capita dwell time, and denoted by $T^1_{attr,q}, T^2_{attr,q}$ respectively.

If $T^1_{stop,q} = T^2_{stop,q}$, then the attraction times of the two train types are both 0, i.e., $T^1_{attr,q} = T^2_{attr,q} = 0$.

If $T^2_{stop,q} < T^1_{stop,q}$, then the attraction time radius of each train of type 2 to each O-D pair in od_q is $(T^1_{stop,q} - T^2_{stop,q})$. In view of situations of departing earlier or later, the attraction time of a train of type 2 is $2(T^1_{stop,q} - T^2_{stop,q})$. The average number of train services of type 2 for O-D pairs in od_q is $y^2 p^2_q m_q$. Assumption (A5) indicates that trains run with a uniform headway, so we can obtain the initial attraction times $T^2_{attr,1} = \min\{T, 2(T^1_{stop,q} - T^2_{stop,q})y^2 p^2_q m_q\}$, and $T^1_{attr,q} = 0$.

If $T^1_{stop,q} < T^2_{stop,q}$, then we can obtain the attraction times similarly: $T^1_{attr,q} = \min\{T, 2(T^2_{stop,q} - T^1_{stop,q})y^1 p^1_q m_q\}, T^2_{attr,q} = 0$.

Except from the total attraction time, the surplus time length is $T_{rem,q} = T - T^1_{attr,q} - T^2_{attr,q}$. We can roughly divide the surplus time by the proportions of the train services of the two types for each O-D pair in od_q. The surplus attraction times of the two train types can be shown as $T_{rem,q} y^t p^t_q / \sum_{t=1}^{2} y^t p^t_q, t = 1, 2$.

According to the initial attraction times and surplus attraction times of the two types of trains, we can obtain the division proportions of train type

$$\alpha^t_q = (T^t_{attr,q} + T_{rem,q} y^t p^t_q / \sum_{t=1}^{2} y^t p^t_q)/T, \ t = 1, 2, \tag{11}$$

satisfying $\alpha^t_q \geq 0, \sum_{t=1}^{2} \alpha^t_q = 1$.

Combined with the division proportions, we can obtain the per capita dwell time of the double-service demand.

$$T_{stop,\bar{c}} = \sum_{q \in Q_{\bar{c}}} [\beta_q \lambda_q T^2_{stop,q} + (1 - \beta_q)\lambda_q \sum_{t=1}^{2} \alpha^t_q T^t_{stop,q}]/\lambda_{\bar{c}} \tag{12}$$

3.3. Per Capita Boarding Deviation Time

3.3.1. Per Capita Boarding Deviation Time of Single-Service Demand

Since single-service demand can only choose type 1, and the whole O-D demand follows the uniform distribution, we assume the average headway between the train services for the corresponding O-D demand is $H = T/(y^1 p^1_q m_q), q \in Q_c$. As shown in Figure 5, the attraction time radius of a train of type 1 is $H/2$, and the boarding deviation time follows a linear distribution in the attraction time radius, so the per capita boarding deviation time of single-service demand is equal to the integral of the boarding deviation time of the demand departing during the attraction time, i.e.,

$$T^1_{diff,q} = 0.25T/(y^1 p^1_q m_q), q \in Q_c. \tag{13}$$

Thus, the per capita boarding deviation time of single-service demand is

$$T_{diff,c} = \sum_{q \in Q_c} \lambda_q T^1_{diff,q}/\lambda_c. \tag{14}$$
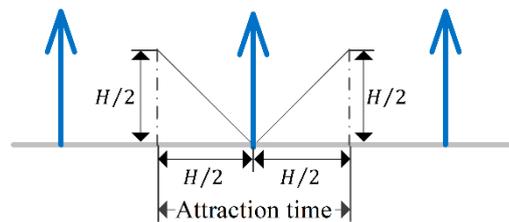
Figure 5. Illustration of attraction to boarding deviation time.

3.3.2. Per Capita Boarding Deviation Time of Double-Service Demand

As for the passengers who only choose trains of type 2, we obtain the per capita boarding deviation times

$$T^2_{\text{diff},q} = 0.25T/(y^2 p_q^2 m_q), q \in Q_{\bar{c}}. \tag{15}$$

As for the passengers who can choose either train type, we can determine the service time of the two train types $\alpha_q^1 T, \alpha_q^2 T$. Then we can obtain the per capita boarding deviation times of passengers who choose the two train types

$$T^{\bar{c},t}_{\text{diff},q} = 0.25\alpha_q^t T/(y^t p_q^t m_q), q \in Q_{\bar{c}}, t = 1, 2. \tag{16}$$

Thus, the per capita boarding deviation time of double-service demand is

$$T_{\text{diff},\bar{c}} = \sum_{q \in Q_{\bar{c}}} [\beta_q \lambda_q T^2_{\text{diff},q} + (1 - \beta_q)\lambda_q \sum_{t=1}^{2} \alpha_q^t T^{\bar{c},t}_{\text{diff},q}]/\lambda_{\bar{c}}. \tag{17}$$

If the denominators in formulas (15) and (16) are 0, then the per capita boarding deviation times can be replaced by a sufficiently large constant M.

3.4. Train Load and Stop Density Equilibrium

3.4.1. Equilibrium of Train Load

The equilibrium of train load is that the proportions of the train kilometers of the two train types are close to those of their transported passenger kilometers, so as to keep a balance between the loads of the two train types. The average transported passenger kilometers of the total demand by the two types are $V_1 = \sum_{q \in Q_c} \lambda_q l_q + \sum_{q \in Q_{\bar{c}}} \alpha_q^1 (1 - \beta_q)\lambda_q l_q$ and $V_2 = \sum_{q \in Q_{\bar{c}}} [\beta_q + \alpha_q^2 (1 - \beta_q)]\lambda_q l_q$, respectively. Thus, the proportion of the transported passenger kilometers by type 1 is $\rho = V_1/(V_1 + V_2)$. The equilibrium of train load can be expressed by the relative error between y^1 and ρ, i.e.,

$$|\rho - y^1|/y^1 \le \varepsilon_y, \tag{18}$$

in which $0 < \varepsilon_y \le 0.05$, and a larger line capacity corresponds to a larger ε_y.

3.4.2. Equilibrium of Stop Density

If a train stops at a provincial/district station, then the stop probability in the corresponding province/district will reach maximum; otherwise, it will be minimal. We define the concept of stop density to describe the stop distribution in a province/district, which is the average stop number in that region. The imbalanced stop probabilities would result in the deviation among stop densities of trains at one province or district, in other words, the deviation among travel speeds (the travel time of a train includes the stop time at intermediate stations). With given headways at the starting station, the utilization ratio of the line capacity will be improved if trains have approximate travel speeds because in this way the headways between trains in each section are close to the defined headways at the starting station.

Next, we discuss how to ensure the equilibrium of stop density by stop probabilities.

(1) In a province

If a train stops at a provincial station, then the average stop probability in that province is $\theta_1 + \theta_2 x^1_{2,1} + \theta_3[x^1_{2,1}x^1_{3,1} + (1 - x^1_{2,1})x^1_{3,0}]$; otherwise, it is $\theta_2 x^1_{2,0} + \theta_3[x^1_{2,0}x^1_{3,1} + (1 - x^1_{2,0})x^1_{3,0}]$. The difference between them, i.e., e^1_1, is equal to $\theta_1 + [\theta_2 + \theta_3(x^1_{3,1} - x^1_{3,0})](x^1_{2,1} - x^1_{2,0})$. If the stop probability of trains of type 1 at provincial stations is 1, then $e^1_1 = 0$.

Generally, there is only one provincial station in a province, so the average number of stations in a province is $1 + \theta_2/\theta_1 + \theta_3/\theta_1$, and the difference between average numbers of stop stations is $e^1_1(1 + \theta_2/\theta_1 + \theta_3/\theta_1)$, of which the upper limit is $n^1_{\text{stop},1} \geq 0$. $n^1_{\text{stop},1}$ is determined by the utilization ratio of the line capacity: as for a busy line, $n^1_{\text{stop},1}$ is usually no larger than 2.5. Thus, the difference between the average stop probabilities of trains of type 1 must satisfy

$$\left|e^1_1\right| \leq n^1_{\text{stop},1}/(1 + \theta_2/\theta_1 + \theta_3/\theta_1). \tag{19}$$

(2) In a district

If a train stops at a district station, then the average stop probability in that district is $\theta_2 + \theta_3 x^1_{3,1}$; otherwise, it is $\theta_3 x^1_{3,0}$. The difference e^1_2 is $\theta_2 + \theta_3(x^1_{3,1} - x^1_{3,0})$. If the stop probability of train of type 1 at district stations is 1, then $e^1_2 = 0$.

If there is only one district station in a district, then the average number of stations in a district is $1 + \theta_3/\theta_2$, and the difference between average numbers of stop stations is $e^1_2(1 + \theta_3/\theta_2)$, of which the upper limit is $n^1_{\text{stop},2} \geq 0$, which is usually no larger than 1 on a busy line. Thus, the difference between the average stop probabilities of trains of type 1 must satisfy

$$\left|e^1_2\right| \leq n^1_{\text{stop},2}/(1 + \theta_3/\theta_2). \tag{20}$$

(3) The constraint of the stop probabilities of trains of type 2

Since trains of type 2 do not stop at county stations, the stop probabilities of trains of type 2 are far less than those of type 1. In order to ensure reasonable stop probabilities of trains of type 2, we propose a constraint that the stop probabilities of trains of type 2 at provincial or district stations should be no less than those of type 1, i.e.,

$$x^2_h \geq x^1_h, h = 1, 2. \tag{21}$$

3.5. Optimization Formulation and Solution Algorithm

3.5.1. Model Formulation

As introduced earlier, the stop planning problem can be described as to optimize stop probabilities to minimize passengers' direct O-D travel time. The travel time of a passenger consists of the boarding deviation time, running time and stop time at intermediate stations. Since the train routes on a line are predefined, we do not consider the running times between stations, and the per capita travel time is equal to the weighted average of the travel times of single/double-service demand, of which the weights are λ_c and $\lambda_{\bar{c}}$ respectively. Thus, the optimization model of stop probability is as follows:

$$\min Z = \lambda_c(T_{\text{stop},c} + T_{\text{diff},c}) + \lambda_{\bar{c}}(T_{\text{stop},\bar{c}} + T_{\text{diff},\bar{c}}). \tag{22}$$

s.t.
Equations (18)–(21)
$0 < x^t_h \leq 1, t = 1, 2, h = 1, 2, 3$
$0 \leq x^t_{h,1}, x^t_{h,0} \leq 1, t = 1, 2, h =, 2, 3$

The objective function (22) is to minimize the per capita travel time; constraint (18) aims to ensure a balance between the train loads of the two types; constraints (19) and (20) are aimed to improve the

utilization ratio of the line capacity; constraint (21) is to ensure the reasonable demand allocation on the two train types.

3.5.2. Solution Algorithm

Since the optimization model is a non-convexity problem, we adopt the penalty function method and introduce large penalty factors to add the constraints (19)–(21) to the objective function as penalty functions. The proportion of transported passenger kilometer by type 1, i.e., ρ, is not a continuous variable, so we keep the formula (18) in the constraint to avoid shocks of the objective function in the solving process.

After eliminating the intermediate variables, the remaining variables include $X = (x_1^1, x_{2,1}^1, x_{2,0}^1, x_{3,1}^1, x_{3,0}^1, x_1^2, x_{2,1}^2, x_{2,0}^2)$ and y^1 (y^2 is equal to $1 - y^1$). We set $x_1^1, x_{2,1}^1, x_{3,1}^1, y^1 \neq 0$ to ensure the feasibility of α_q^t and Equation (18). Due to the categorization of stations and O-D demand, the number of decision variables is not influenced by the structure and scale of the studied HSR line, and the solving scale of decision variables is fixed.

Given that all of the variables in X have continuous ranges, we randomly generate several initial solutions, search for the corresponding local optimal solutions from each initial solution, and select an optimal solution from these local optimal solutions. The detailed procedure is described below.

Algorithm 1 Procedure for solving stop probability optimization problem

---

Step 0: Initialization
Deduct the following parameters from the given data, i.e., train routes and frequencies, O-D demand, and the line operation data:
(i) HSR line parameters: d, θ_h ($h = 1, 2, 3$)
(ii) demand parameters: m_q, λ_q ($q = 1, 2, \cdots, 13$), l_c, l_q, β_q ($q \in Q_{\bar{c}}$)
(iii) operation parameters: T, τ_{stop}, ε_y, $n_{\text{stop},1}^1$, $n_{\text{stop},2}^1$

Step1: Random generations of solutions
Discretize the range of $(0, 1]$ by the step length 0.001 to generate different values of y_k^1, $k = 1, 2, \cdots, 1000$, and randomly generate an initial feasible solution X_k^0 from the uniform distribution for each y_k^1. Calculate the objective (Z_k^0 with determined y_k^1 and X_k^0 according to the following procedure. $\underline{k} \leftarrow 1$, $\bar{k} \leftarrow 1000$.
Solve $T_{\text{stop},q}^t$, $q \in Q_{\bar{c}}$ by Equation (9);
Obtain $T_{\text{stop},c}$, $T_{\text{stop},\bar{c}}$ by solving Equations (10)–(12);
Obtain $T_{\text{diff},c}$ by solving Equations (13), (14), and $T_{\text{diff},\bar{c}}$ by solving Equations. (15)–(17);
Solve the objective by Equation (22).

Step 2: Local search
As for each group of (y_k^1, X_k^0), $k \in [\underline{k}, \bar{k}]$, set the local optimal solution $X_k^* \leftarrow X_k^0$, $Z_k^* \leftarrow Z_k^0$.
Local search process
Set the initial search step length $\gamma = 0.999$.
While $\gamma > 0.001$ **do**
Search for the neighbor solution X_k using the Gradient Descent Method and calculate the objective Z_k.
If $Z_k < Z_k^*$ and the solution satisfies the constraint (18), update the solution $X_k^* \leftarrow X_k$, $Z_k^* \leftarrow Z_k$, $\gamma \leftarrow 0.5\,\gamma$.
End

Step 3: Local search repetition
As for the vth group of local optimal solution (y_v^1, X_v^*), $v = 1, 2, \cdots, 1000$, update the values of y_k^1 in the solutions of (y_k^1, X_k^*), $k \in [\underline{k}, \bar{k}]$, $\underline{k} = \max\{1, v - 5\}, \bar{k} = \min\{1000, v + 5\}$ by y_v^1, and repeat **Step 2** to search a better local optimal solution for y_v^1.

Step 4: Select the optimal solution
Compare the objectives of the local optimal solutions and select the solution with minimal objective as the optimal solution, and obtain n_q, $q = 1, 2, \cdots, 13$ by Equations (1)–(8).

---

4. Case Studies

4.1. Case Study on Beijing-Guangzhou HSR Line

4.1.1. Basic Input Parameters

In this section, we first make analyses on the stop probability optimization of Beijing-Guangzhou HSR Line, which is illustrated in Figure 6. The proportions of the numbers of stations on the three levels are respectively $\theta_1 = 0.1667, \theta_2 = 0.5278, \theta_3 = 0.3055$, the average station distance is $d = 65.17$ km, the operation time in a day is $T = 17$ h, and the average stop time at a station is $\tau_{\text{stop}} = 8/60$ h.
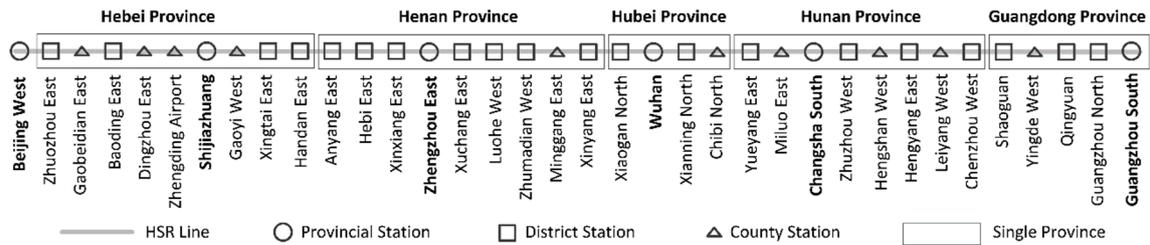
Figure 6. Illustration of Beijing-Guangzhou HSR Line.

According to the statistical results of the 392 trains running on the Beijing-Guangzhou HSR Line on 1 December 2015 (including some cross-line trains), we obtain λ_q and $m_q, q = 1, 2, \cdots, 13$, shown in Table 6. Table 7 shows the proportions of passengers who only choose trains of type 2 and the average travel distances in the double-service demand categories, i.e., $\beta_q, l_q, q \in Q_{\bar{c}}$, as well as the average travel distance of single-service demand, i.e., l_c. The corresponding parameters in Equations (18)–(20) are $\varepsilon_y = 0.05, n^1_{\text{stop},1} = 2, n^1_{\text{stop},2} = 0.8$.

Table 6. Proportions and average numbers of passing trains in each demand category.

q	1	2	3	4	5	6	7
λ_q	0.3043	0.2141	0.0639	0.0146	0.0149	0.0013	0.275
m_q	41.1	35.36	36.21	30.77	29.09	24.83	72.32

q	8	9	10	11	12	13	
λ_q	0.0679	0.0029	0.0003	0.0292	0.011	0.0006	
m_q	80.27	83.28	81.75	53.88	61.68	62.39	

Table 7. Distribution of l_c/l_q and β_q in each demand category.

O-D Index	Q_c	$Q_{\bar{c}}$				
		1	2	4	7	11
l_c/l_q (km)	265.56	687.77	547.48	455.56	175.78	216.47
β_q	–	0.1	0	0	0	0

4.1.2. Optimization Results

We conduct the proposed model on an Intel i7 2.4 GHZ with 16 GB RAM in the environment of Microsoft Windows 10. The computation time for the case study network is around 30 min. The comparison with the results in practice is shown in Table 8.

Table 8. Comparison between solved results and practical results.

	Solved results	Practical results
Objective (h)	$Z^* = 0.6071$ h	$Z^0 = 0.6090$ h
Stop probabilities	$X^* =$ $(1, 0.388, 0, 0.341, 0.411, 1, 0.682, 0)$ $x_2^{1,*} = 0.388, x_3^{1,*} = 0.384, x_2^{2,*} = 0.682$	$X^0 =$ $(1, 0.476, 0, 0.229, 0.4, 0.998, 0.442, 0)$ $x_2^{1,0} = 0.476, x_3^{1,0} = 0.321, x_2^{2,0} = 0.441$
Proportion of train kilometers of train type 1	$y^{1,*} = 0.689$	$y^{1,0} = 0.576$
Proportion of transported passenger kilometers by train type 1	$\rho^* = 0.655$	$\rho^0 = 0.605$
Relative error between y_1 and ρ	$\left\lvert y^{1,*} - \rho^* \right\rvert / y^{1,*} = 0.050$	$\left\lvert y^{1,0} - \rho^0 \right\rvert / y^{1,0} = 0.050$
Average number of train services for each demand category	$n^* = (41.10, 16.95, 9.58, 7.64, 2.99,$ $2.52, 34.67, 21.23, 7.59, 8.36, 13.38,$ $6.33, 6.33)$	$n^0 = (41.03, 14.70, 7.36, 5.32,$ $2.69, 1.07, 30.66, 19.14, 6.22, 3.38,$ $11.26, 7.08, 3.67)$

It can be indicated from Table 8 that the travel time with the solved stop probabilities is slightly less than that with practical stop probabilities, and the solved average stop probability is a little larger than that in practice. We simplify the elements in $X = (x_1, x_2, \cdots, x_N)$, then the relative error of the values of decision variables, i.e., $(y^{1,*}, X^*)$ and $(y^{1,0}, X^0)$, is 0.1640 which can be calculated by the following formula:

$$\text{relative\_err} = \sqrt{(y^{1,*} - y^{1,0})^2 + \sum_{k=1}^{N} (x_k^* - x_k^0)^2} \Big/ \sqrt{y^{1,*2} + \sum_{k=1}^{N} x_k^{*2}}. \tag{23}$$

The small relative error demonstrates the correctness of the solved results, and the obtained service frequencies for each demand category are reasonable.

From the comparisons between the solved results and practical results, we can obtain some findings:

(a) It is shown from optimization and practical operation that the stop probabilities of the two train types at provincial stations are approximate to 1, i.e., $x_1^{1,*} = 1, x_1^{2,*} = 1$, so $x_{2,0}^{1,*} = 0, x_{2,0}^{2,*} = 0$, that is, every train must stop at each provincial station it passes. This result demonstrates the importance of provincial stations.

(b) In the solved results, the stop probability of train type 1 at county stations increases, while that at district stations declines; the stop probability of train type 2 at district stations increases. In this way, the single-service demand can be served by train type 1 with a larger probability, and more double-service demand can choose train type 2. It shows in the solved results that the two train types can have a more definite load division of transported passengers, which represents the specific services of the two train types.

(c) Most of the solved numbers of train services for O-D pairs in each demand category are slightly larger than those in practice with a minor error, which results from the increased service probabilities for O-D demand. From the solved results, the demand categories with a large number of direct services include od_1, od_2, od_7 due to the large proportions, and od_8 due to a high level of service for passengers within a province.

(d) The solved values of y^1, ρ are both larger than the practical values. This is because we only consider the direct schemes for passengers, and the single-service demand can only choose train type 1, while single-service demand can choose trains of type 2 to transfer in the practical operation. Thus, in the solved results, y^1, ρ both increase to show a more definite allocation of different passengers on the two train types. In addition, the relative errors between y^1 and ρ in the two results both satisfy the constraint.

4.1.3. Analysis on Proportions of Train Kilometers of Train Type 1

We solve the minimal travel times Z with different proportions $y^1 \in (0, 1]$; the results are illustrated in Figure 7 and the following characteristics can be observed:
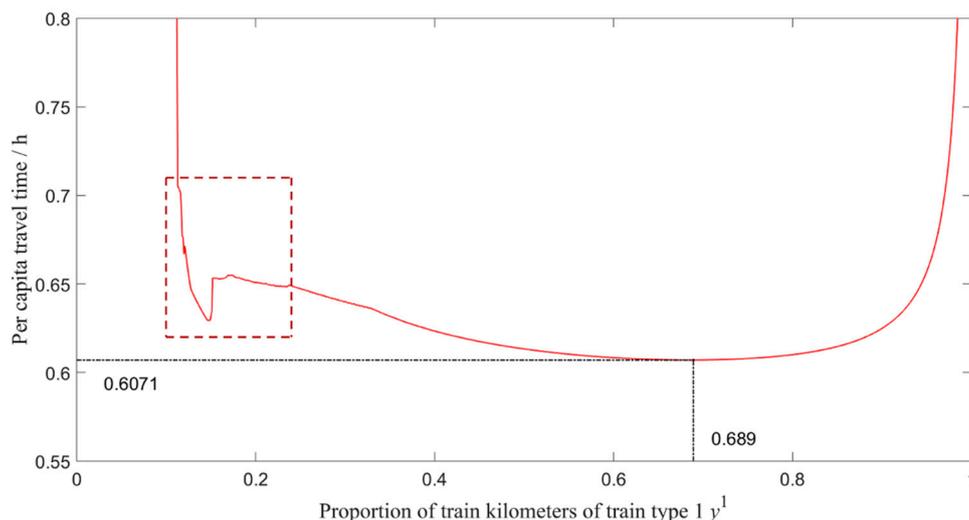
Figure 7. Illustration of the change of per capita travel time on the Beijing-Guangzhou HSR Line.

(a) In the range of $y^1 \leq 0.104$, the objective is infinity. This is because the values in that range cannot satisfy constraint (18). Constraint (18) is a discrete constraint, which requires y^1 should be close to ρ. The first train type can load the single-service demand and a part of double-service demand. Note that the proportion of passenger kilometers of single-service demand is a given value, i.e., $\eta = \lambda_c l_c / (\lambda_c l_c + \sum_{q \in Q_{\bar{c}}} \lambda_q l_q)$, and these passengers can only choose type 1, η can be viewed as the lower bound of ρ, accounting for a large percentage of ρ. Since this proportion on the Beijing-Guangzhou HSR Line is $\eta = 0.1003$, the minimal feasible y^1 should be slightly larger than η; otherwise, some single-service passengers would be detained resulting in an infeasible solution. In this case, the minimal value of y^1 is 0.104.

(b) In the range of $y^1 \in [0.104, 0.240]$, the objective shocks sharply (as marked by the square in Figure 7). This also results from constraint (18), which restricts the transported passenger kilometers of the double-service demand by train type 1. Thus, some double-service demand cannot choose the proper train type, and this imbalanced passenger assignment in each iteration leads to the shocks of the objective value in this range.

(c) In the remaining range $y^1 > 0.240$, the objective first decreases before rising with y^1. When y^1 reaches a range that can keep the balance of passenger assignment, passengers can freely choose proper train types, the objective value would decrease smoothly. When $y^1 = 0.689$, the objective value is minimal. In the neighborhood of $y^{1,*} = 0.689$, the objective changes in a small range, so the value of the practical proportion of the train kilometers of type 1 can be found in the neighborhood of $y^{1,*}$ in the adding direction. After that, if y^1 still keeps increasing, the stop time of the two train types as well as the boarding deviation time of type 2 will increase, and the choices of passengers who only choose type 2 would also be influenced, so the objective value rises.

4.2. Comparative Analysis with Two Other HSR Lines

In order to make further conclusions, we calculate the solved results for other two HSR Lines, i.e., the Beijing-Shanghai HSR Line and the Hangzhou-Fuzhou-Shenzhen HSR Line, of which the computation times are also around 30min. Some basic parameters about the distributions of stations and demand on the two lines are listed in Table 9; the stop probabilities at stations on different levels on the two lines in solved and practical results are listed in Table 10. The relative errors between the solved and practical results of the two lines are respectively 0.1217 and 0.1255. The curves of the solved per capita travel time with y^1 of the two lines are illustrated in Figures 8 and 9, respectively, which illustrate similar change trends with the Beijing-Guangzhou HSR Line.

Table 9. Basic parameters of the other two HSR Lines.

HSR Line	θ_1	θ_2	θ_3	λ_c	λ_c^-	η
Beijing-Shanghai	0.1739	0.6522	0.1739	0.1002	0.8998	0.0435
Hangzhou-Fuzhou-Shenzhen	0.0755	0.2075	0.7170	0.5011	0.4989	0.4281

Table 10. Solved and practical results of the other two HSR Lines.

HSR Line		x_1^1	x_2^1	x_3^1	x_1^2	x_2^2	y^1	Z (**h**)
Beijing-Shanghai	Solved Results	0.981	0.2678	0.4911	0.981	0.4738	0.527	0.6272
	Practical results	1	0.4314	0.4102	0.983	0.3991	0.479	0.8493
Hangzhou-Fuzhou-Shenzhen	Solved Results	1	0.743	0.3982	1	0.767	0.853	0.8298
	Practical results	0.995	0.8557	0.3182	0.997	0.9741	0.889	0.8504
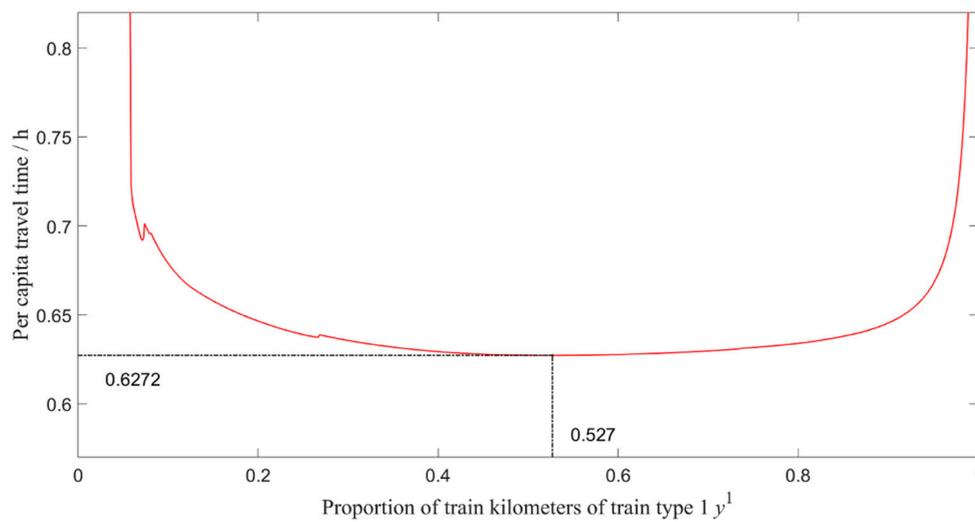
Figure 8. Illustration of the change of per capita travel time on the Beijing-Shanghai HSR Line.
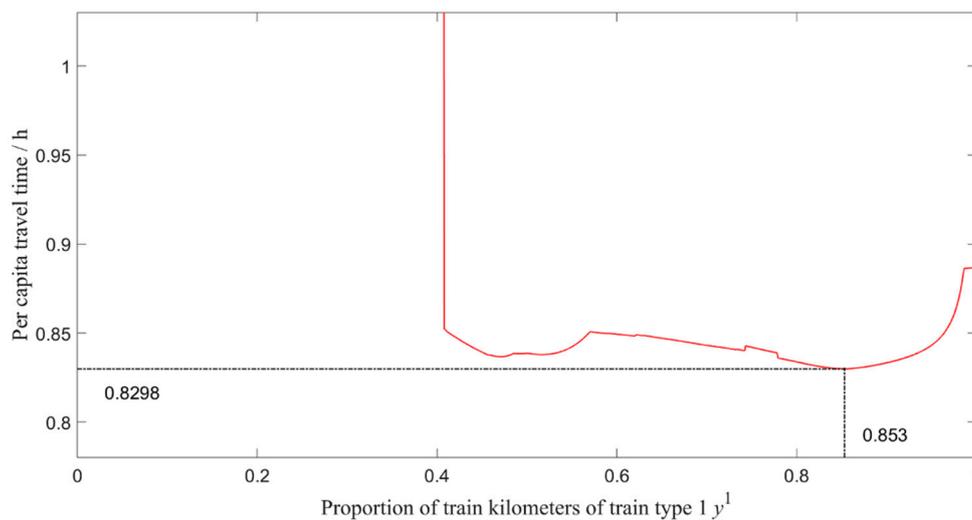
Figure 9. Illustration of the change of per capita travel time on the Hangzhou-Fuzhou-Shenzhen HSR Line.

Combined with Beijing-Guangzhou HSR Line, some common conclusions can be drawn as follows:
(a) Stop probabilities of the two train types at stations

From the results of the three HSR Lines, the stop probabilities of the two train types at provincial stations are close to 1 which demonstrates the attraction effect of provincial stations.

Since the single-service demand can only be served by trains of type 1, the stop probability of train type 1 at county stations should be increased to decrease the boarding deviation time of single-service demand; the stop probability of train type 1 at district stations should be slightly decreased to lead double-service demand to choose train type 2. As for train type 2, the stop probability at district stations should be properly increased to transport more double-service demand. In this way, it can reach a more reasonable load division between the two train types.

From the view of proportions of different stations, stations with a smaller proportion (such as provincial stations) usually need a larger stop probability to serve the travel demand from/to them.

(b) The proportion division of the two train types

It can be indicated from Figures 7–9: since η is the lower bound of ρ which is close to y^1, a smaller η would narrow the infeasible range. Generally, a smaller proportion of the number of county stations θ_3 corresponds to a smaller η (not a strict correspondence). Thus, we can roughly deduct that a smaller θ_3 corresponds to a smaller minimal feasible y^1.

A regression relationship between the proportions of train kilometers and passenger kilometers of single-service demand of the three railway lines can be found: $y^1 = -0.1067 \times \eta^{-0.4842} + 1.014$; the result is also illustrated in Figure 10. This provides a practical guide as to plan the train kilometers with a given passenger kilometers to county stations.
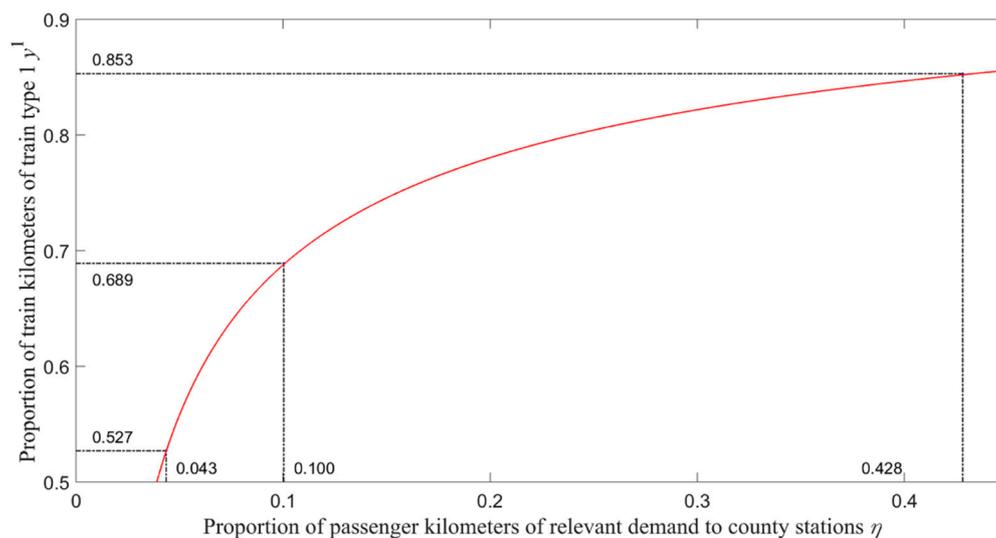
Figure 10. Fitting curve between η and y^1.

5. Conclusions

In this paper, we propose an analytical method that provides the theoretical deduction of required service frequencies for passenger O-D demand as specified by geo-political station levels and passenger demand levels. Considering the attraction effects of provincial and district stations, we classify demand into several categories that can reduce the scale of the number of O-D pairs and also present the travel characters of each demand category. Given the level of service, trains are classified into two types by service availabilities. We propose a concept of stop probability to describe the stop planning problem and formulate the optimization of stop probabilities of HSR trains to obtain the average numbers of train services for O-D pairs in each demand category. The proposed model can determine the train stop probabilities at stations on different levels and the reasonable proportion of the train kilometers of each type with the objective of minimizing per capita travel time.

In the model formulation, we consider two equilibrium conditions—(i) equilibrium of train load to keep a balance between the train kilometers and transported passenger kilometers of the two train types and (ii) equilibrium of stop density to improve the utilization ratio of the line capacity.

The proposed method is validated on three real-life HSR lines. The solved stop probabilities are shown to match closely to those in practice. The solved stop probabilities of the two train types at provincial stations are close to 1, which complies with the attraction effect of provincial stations. The stop probability of train type 1 at county stations should be increased to serve more single-service demand, while that at district stations is decreased to lead more double-service demand to choose the other train type, which can reach a more reasonable load division between the two train types. The station set with a smaller proportion (the provincial station set) generally needs a higher stop probability, which can ensure more serving trains and reduce the added travel time.

The proportion of passenger kilometers of single-service demand is the lower bound of the proportion of transported passenger kilometers of train type 1, so it has an influence on the minimal feasible proportion of train kilometers of train type 1. We obtain a regression relationship between the proportions of train kilometers and passenger kilometers of single-service demand; the result has good practical significance.

Our future research can be extended in three aspects: the first is to optimize the service frequency considering more O-D demand characteristics, such as passengers' time value, different distribution patterns during a day, the influence of O-D distances, and so on; the second is to adopt the solved stop probabilities to design a more reasonable stop plan for different O-D levels; and the third is to extend the studied structure to HSR networks considering characteristics of different train services, such as in-line and cross-line trains.

References

1. Lusby, R.; Larsen, J.; Ehrgott, M.; Ryan, D. Railway track allocation: Models and methods. *OR Spectr.* **2011**, *33*, 843–883. [CrossRef]
2. Fu, H.; Nie, L.; Meng, L.; Sperry, B.; He, Z. A hierarchical line planning approach for a large-scale high-speed rail network: The China case. *Transp. Res. Part A Policy Pract.* **2015**, *75*, 61–83. [CrossRef]
3. Su, H.; Shi, F.; Deng, L.; Shan, X. Time-dependent demand oriented line planning optimization for the HSR. *J. Transp. Syst. Eng. Inf. Technol.* **2016**, *16*, 110–116.
4. Guan, J.; Yang, H.; Wirasinghe, S. Simultaneous optimization of transit line configuration and passenger line assignment. *Transp. Res. Part B Methodol.* **2006**, *40*, 885–902. [CrossRef]
5. Kaspi, M.; Raviv, T. Service-oriented line planning and timetabling for passenger trains. *Transp. Sci.* **2013**, *47*, 295–311. [CrossRef]
6. Yang, L.; Qi, J.; Li, S.; Gao, Y. Collaborative optimization for train scheduling and train stop planning on HSRs. *Omega* **2016**, *64*, 57–76. [CrossRef]
7. Chen, D.; Ni, S.; Xu, C.; Lv, H.; Wang, S. High-speed train stop-schedule optimization based on passenger travel convenience. *Math. Probl. Eng.* **2016**. [CrossRef]
8. Deng, L.; Shi, F.; Zhou, W. Stop schedule plan optimization for passenger train. *China Railw. Sci.* **2009**, *30*, 102–107.

9. Niu, F.; Qi, J.; Qin, J. Optimization model for train stopping plan on HSR corridor with uncertain passenger demands. *J. China Railw. Soc.* **2016**, *38*, 1–6.

10. Li, D.; Han, B.; Li, X.; Zhang, H. HSR stopping schedule optimization model based on node service. *J. China Railw. Soc.* **2013**, *35*, 1–5.

11. Fu, H.; Sperry, B.; Nie, L. Operational impacts of using restricted passenger flow assignment in high-speed train stop scheduling problem. *Math. Probl. Eng.* **2013**. [CrossRef]

12. Jamili, A.; Aghaee, M. Robust stop-skipping patterns in urban railway operations under traffic alteration situation. *Transp. Res. Part C Emerg. Technol.* **2015**, *61*, 63–74. [CrossRef]

13. Vuchic, V. *Urban Transit: Operation, Planning and Economics*; John Wiley & Sons: Hoboken, NJ, USA, 2005.

14. Niu, H.; Zhou, X. Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. *Transp. Res. Part C Emerg. Technol.* **2013**, *36*, 212–230. [CrossRef]

15. Niu, H.; Zhou, X.; Gao, R. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transp. Res. Part B Methodol.* **2015**, *76*, 117–135. [CrossRef]

16. Shi, F.; Zhao, S.; Zhou, Z.; Wang, P.; Bell, M. Optimizing the train operational plan in an urban rail corridor based on the maximum headway function. *Transp. Res. Part C Emerg. Technol.* **2017**, *74*, 51–80. [CrossRef]

17. Freyss, M.; Giesen, R.; Muñoz, J. Continuous approximation for skip-stop operation in rail transit. *Transp. Res. Part C Emerg. Technol.* **2013**, *36*, 419–433. [CrossRef]

18. Suh, W.; Chon, K.; Rhee, S. Effect of skip-stop policy on a Korean subway system. *Transp. Res. Board* **2007**, *1793*, 33–39. [CrossRef]

19. Zolfaghari, S.; Azizi, N.; Jaber, M. A model for holding strategy in public transit systems with real-time information. *Int. J. Transp. Manag.* **2004**, *2*, 99–110. [CrossRef]

20. Cao, Z.; Yuan, Z.; Zhang, S. Performance analysis of stop-skipping scheduling plans in rail transit under time-dependent demand. *Int. J. Environ. Res. Public Health* **2016**, *13*, 707. [CrossRef]

21. Ghoseiri, N.; Wirasinghe, S. Optimum zone structure during peak periods for existing urban rail lines. *Transp. Res. Part B Methodol.* **1986**, *20*, 7–18.

22. Nemhauser, G. Scheduling local and express service. *Transp. Sci.* **1969**, *3*, 164–175. [CrossRef]

23. Ulusoy, Y.; Chien, S.; Wei, C. Optimal all-stop, short-turn and express transit services under heterogeneous demand. *Transp. Res. Rec.* **2010**, *2197*, 8–18. [CrossRef]

24. Jong, J.; Suen, C.; Chang, C. Decision support system to optimize railway stopping patterns. *Transp. Res. Rec. J. Transp. Res. Board* **2012**, *2289*, 24–33. [CrossRef]

25. Cheng, J.; Peng, Q. Combine stop optimal schedule for urban rail transit with elastic demand. *Appl. Res. Comput.* **2014**, *31*, 3361–3364.

26. Chang, Y.; Yeh, C.; Shen, C. A multi-objective model for passenger train services planning: Application to Taiwan's high-speed rail line. *Transp. Res. Part B Methodol.* **2000**, *34*, 91–106. [CrossRef]

27. Goossens, J.; van Hoesel, S.; Kroon, L. A branch-and-cut approach for solving railway line-planning problems. *Transp. Sci.* **2004**, *38*, 379–393. [CrossRef]

28. Goossens, J.; van Hoesel, S.; Kroon, L. On solving multi-type railway line planning problems. *Eur. J. Oper. Res.* **2006**, *168*, 403–424. [CrossRef]

29. Lin, D.; Ku, Y. Using genetic algorithms to optimize stopping patterns for passenger rail transportation. *Comput.-Aided Civ. Infrastruct. Eng.* **2014**, *29*, 264–278. [CrossRef]

30. Lai, Y.; Shih, M.; Chen, G. Development of efficient stop planning optimization process for high-speed rail systems. *J. Adv. Transp.* **2016**, *50*, 1802–1819. [CrossRef]

31. Yue, Y.; Wang, S.; Zhou, L.; Tong, L.; Saat, M. Optimizing train stopping patterns and schedules for high-speed rail corridors. *Transp. Res. Part C Emerg. Technol.* **2016**, *63*, 126–146. [CrossRef]