Journal of Advanced Transportation

Methods and Technologies for Next-Generation Public Transport Planning and Operations

Lead Guest Editor: Tao Liu
Guest Editors: Prakash Ranjitkar and Yu Jiang

Methods and Technologies for Next-Generation Public Transport Planning and Operations

Journal of Advanced Transportation

Methods and Technologies for Next-Generation Public Transport Planning and Operations

Lead Guest Editor: Tao Liu
Guest Editors: Prakash Ranjitkar and Yu Jiang

Contents

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5947802, 13 pages
https://doi.org/10.1155/2020/5947802

Research Article

A Short-Turn Dispatching Strategy to Improve the Reliability of Bus Operation

Runkun Liu,[1] **Haiyang Yu**,[1,2] **Pengfei Wang**,[3,4] **and Hai Yan**[5]

[1]*School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems and Safety Control, Beihang University, Beijing 100191, China*
[2]*Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China*
[3]*Key Laboratory of Urban Security and Disaster Engineering of Ministry of Education, Beijing University of Technology, Beijing 100124, China*
[4]*College of Urban Construction, Hebei Normal University of Science & Technology, Qinhuangdao, Hebei 066004, China*
[5]*Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Hai Yan; yhai@bjut.edu.cn

Received 16 October 2019; Revised 26 November 2020; Accepted 4 December 2020; Published 15 December 2020

Academic Editor: Tao Liu

Bus dwelling time at stops is affected by the number of boarding and alighting passengers. Due to the uncertainty of traffic state, dwelling time will become unstable, resulting in unreliable bus operations. A short-turn strategy shares passenger demand in certain segments of the full-length vehicles; therefore, this strategy can reduce the dwell time of the full-length vehicles and have potential to improve reliability of bus operations. Firstly, this paper builds a bus operation state model considering short-turn vehicles. Secondly, the process of bus operation is analyzed to investigate the impact of passenger demand on reliability. Then, we propose a short-turn dispatching strategy, which can accurately determine the operation area and departure time. Finally, a numerical simulation is used to evaluate different strategies. As the result, the analysis reveals the following: (1) the bus operational reliability of full-length vehicles under high passenger demand could be improved by a well-designed short-turn strategy; (2) it is beneficial to improve the reliability of bus operation when short-turn vehicles departure at the change point; and (3) an undesigned short-turn dispatching strategy has a negative effect on full-length vehicles operational reliability. This study can help design a short-turn strategy and make valuable contributions to the bus management system.

1. Introduction

Bus priority development policy is an important transportation strategy to solve traffic problems in large cities and achieves sustainable urban development. Moreover, the reliability of operation greatly affects the level of service on public transport and affects passengers' choices in public transport. Different researchers have different understandings of reliability [1]. Bates et al. [2] associated reliability with public transportation according to timetable operation. Turnquist and Bowman [3] believed that the maintenance of regular service is an important indicator of reliable public transport. In a large city like Beijing, there is no fixed timetable for bus departure. The departure time is based on the real-time traffic state, passenger demand, as well as other factors. For instance, in the morning and evening peak periods, the bus company increases the frequency of the departure. Therefore, reliability in Beijing refers to the stability of bus headway, which can be expressed as the mean absolute error between bus headway and the arrival time interval.

Multiple factors affect the bus operation reliability, such as traffic congestion, weather conditions, time of passengers' boarding and departure, and drivers' operations, which may result in late bus arrival or irregular arrival [4]. In the previous studies, Sorratini et al. [1] and Mark et al. [5] found that the number of passengers and the time they spend on boarding and alighting a bus were the main reasons for unreliability.

Especially when buses and cars are mixed and traffic flow is saturated, the speed of the buses drops sharply and buses cannot be operated in accordance with established timetables. Large interval and bus bunching will happen. These phenomena make the main contribution to unreliability.

Several methods have been proposed to ease the instability and excessive demand of passengers at some stations. Tétreault and El-Geneidy [6] adopted the express bus service, and Furth [7] adopted short-turn service, and also the bus holding and stop-skipping methods were used [8, 9]. Express bus service is to skip some stations with lower passenger demand. Once the departure frequency cannot be guaranteed, this method will waste some passengers' time for short trips and reduce the attractiveness of public transport. In terms of the bus holding method, this method also wastes the time of passengers to some extent. Liang et al. [10] proposed a coordinated control method combining two bus holding and stop-skipping methods. By contrast, short-turn services can meet the requirements of the passengers on the higher demand section of the road and effectively alleviate the situations that more passengers waiting on the bus line at some stations. Therefore, short-turn services can greatly affect buses dwell time and then affect the operation state of full-length vehicles.

Prior work has documented short-turn services. But, most of the existing research studies focused on the problem of combined dispatching with multiple services such as short-turn service, full-length service, and express bus service. Most studies focus on building an optimization model to reduce the waiting time of passengers or bus operation costs [11–14]. For example, Tirachini et al. [14] established a scheduling model considering passenger travel cost and the operation cost, and the genetic algorithm was used to solve departure frequency. The operation area of short-turn vehicles and the first and last stops are determined through OD (origin-destination) data with consideration of passenger demand and vehicle capacity in Furth's studies [7, 15]. Canca et al. [16] developed an optimization model which integrates short-turning shuttle operations. Because short-turn vehicles only travel between some certain stations of the line with a limited number of passengers, it is possible to consider using different models to reduce operation cost. For example, Ceder et al. [17] took different sized vehicles to optimize the bus dispatching problem. Within our knowledge, less work has been done to analyze the impact of the short-turn strategy on the public transport system from the perspective of reliability.

In this paper, we focus on designing a simple and effective short-turn strategy to improve bus operational reliability. There are two key factors in designing the short-turn strategy: where to operate and when to depart. For the first factor, it is generally believed that the reasonable operating area of short-turn vehicles should be in the stations with larger passenger demand [18]. Turn-back points should be located with the objective of diminishing the passengers' waiting time while preserving a certain level of service quality. The operation area of short-turn vehicles should satisfy the needs of passengers as much as possible. For the second factor, although the studies listed above showed that most of the scheduling methods of the short-turn vehicles are optimized under the targets of the departure frequency and the minimum passenger travel cost [19, 20], it is impossible to accurately determine the reasonable departure time of the short-turn vehicles. We believe the departure time should be in rush hours of the day to release the pressure of full-length vehicles. That means, the departure time of the short-turn vehicles is set based on the surge in passenger demand or the time at which the road reaches congestion (rush hour). That timing is named as change point. Taylor [21] proposed a simple and effective method to analyze the change point, which can accurately locate the change point of a sequence. The study of change point is widely used in economics, meteorology, and other fields. For example, studying economic changes helps discover potential financial risks [22]; observing climate change leads to corresponding countermeasures [23] and so on. Variables in the transportation system such as the speed and the number of passengers are similar to the sequences of economy and climate. The same change point test method can be used to detect the time for traffic congestion and passenger demand surge.

The main contributions of this paper are summarized as the following three points. (1) A simulation-based model of the bus operation state is established, which can describe the real operation process of buses. The causes of bus unreliability are analyzed in detail. (2) A short-turn dispatching strategy is proposed. Through this strategy, the departure time and the operating area of the short-turn vehicles can be set. (3) Different strategies are analyzed by numerical simulation based on the bus operation state model considering short-turn vehicles. The data come from the smart card data of a bus line in Beijing. Results believe that a well-designed short-turn strategy could improve the reliability of the bus operation.

The reminder of this paper is summarized as follows: in Section 2, the operation process of buses is modeled; in Section 3, we explored the impact of passenger number on the operation state of buses; then, in Section 4, the method of establishing the operating area and departure time of short-turn vehicles is discussed; in Section 5, a simulation experiment is executed with different strategies and the results are analyzed. Finally, Section 6 includes a discussion of the results and concluding thoughts.

2. Modeling of Bus Operation State

2.1. Notations. According to the operation process of buses, the operational time of buses consists of two parts: travel time and dwell time (see Figure 1).

We build a mathematical model to describe operation process of full-length buses and short-turn buses, including the bus dwell time, bus arrival time, and departure time in each station. The notations are summarized in Table 1.

This study is based on the following hypotheses:

(i) The assumption of bus dwell time.

The dwell time of buses at each station is assumed only to be related to the number of passengers

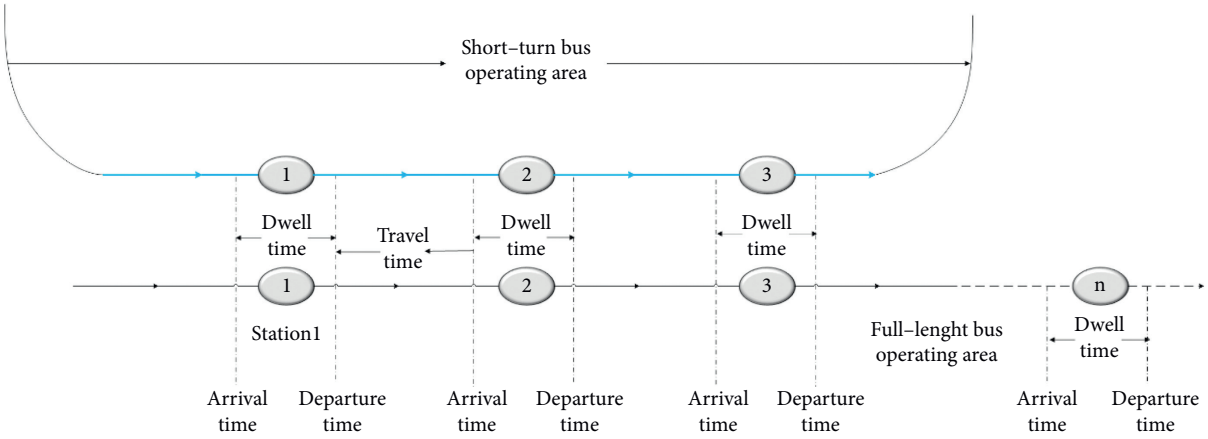


Figure 1: Sectional sketch map of bus travel time.

Table 1: Notations.

Notations	Definitions
N	Bus index, n $\in \forall N+$, n $\neq 1$
I	Station number, $i \in (1, 2, 3, \ldots, s)$ $i \in (1, 2, 3, \ldots, s)$
J	Interstation index, $j \in (1, 2, 3, \ldots, s-1)$ $j \in 1, 2, 3, \ldots, s-1$
ρ_i^m	Passenger arrival rate at station i in time period m
M	Number of periods between to buses
t'	The time it takes for each passenger to get on/off the bus
TS_i^n	Dwell time of bus n at station i
Q_i^n	Boarding number of bus n at the station i
V_j^n	The average speed of bus n travel at station interval j
L_j	The distance of the station intervals j
TT_j^n	Travel time of bus n at station intervals j
TL_k^n	The departure time of bus n at station k
ε	A constant for the other time loss required at each station
$H_i^{n-1,n}$	Headway of bus n and bus n-1 at station i (the difference between the departure time of the two buses at the station i) $(H_i^{n-1,n} = \mathrm{TA}_i^n - \mathrm{TA}_i^{n-1})$
U	Station set of short-turn vehicles operating $U \subseteq (1, 2, 3, \ldots, s)$
Q	Short-turn bus between full-length buses $n-1$ and n
OD_{ab}	The number of passengers who board at station a and get off at station b, $a, b \in (1, 2, 3, \ldots, s)$

getting on/off, and the maximum time for passengers to get on/off is adopted. The other time loss required for the bus to stop at each station is set as a constant.

(ii) The assumption that all passengers waiting for buses at a station do not need to wait for the bus a second time.

When the short-turn strategy is not applied, all passengers waiting at station i can board the bus [24].

Next, the main parameters of the full-length vehicles' operation state are modeled.

2.2. The Main Parameters of Full-Length Vehicles

2.2.1. Bus Dwell Time. When a bus departs from station i, if the arrival rate of passengers does not change during period m and they gather at the station i with a rate of ρ_i^m, the number of passengers gathered in station i is $\rho_i^m \cdot H_i^{n-1,n}$ before the next bus arrives. But actually, the arrival rate of passengers changes in real time. If the arrival rate has changed several times within $H_i^{n-1,n}$, for example, the arrival rate changes 3 times within 500 s, during 0–200 s is 2 pax/min, 200–300 s is 1 pax/min, and 300–500 s is 2 pax/min. Then, the total number of passengers arriving should be $(200^*2 + 100^*1 + 200^*2)/60 = 15$. The dwell time is related to the maximum number of passengers getting on and off at station i, which can be expressed as follows:

$$\mathrm{TS}_i^n = \max\left\{\sum_{a=1}^{i} \mathrm{OD}_{ai}, \sum_{m=1}^{M} \rho_i^m \cdot H_i^{n-1,n,m}\right\} \cdot t' + \varepsilon, \quad (1)$$

where M represents the number of arrival rate changes during $H_i^{n-1,n}$ and $\rho_i^1 \cdot H_i^{n-1,n,1}$ indicates the cumulative number of passengers before the first arrival rate changed in $H_i^{n-1,n}$.

2.2.2. Travel Time between Two Stations. The travel time of buses between adjacent stations is expressed as the distance been divided by speed:

$$\mathrm{TT}_j^n = \frac{L_j}{V_j^n}. \tag{2}$$

2.2.3. Arrival Time. We assume buses arrival at the first station in a fix time interval T and $\mathrm{TA}_1^1 = 0$. The arrival time of the bus n at station k is the dwell time plus all the travel time before station k:

$$\mathrm{TA}_k^n = \begin{cases} (n-1)T, & \text{if } k = 1, \\ \mathrm{TA}_1^n + \sum_{i=1}^{k-1} \mathrm{TS}_i^n + \sum_{j=1}^{k-1} \mathrm{TT}_j^n, & \text{if } 2 \le k \le s. \end{cases} \tag{3}$$

2.2.4. Departure Time. The departure time of the bus can be expressed as the summation of dwell time and travel time of all stations in the previous operation of the bus:

$$\mathrm{TL}_k^n = \begin{cases} \mathrm{TA}_k^n + \mathrm{TS}_k^n, & \text{if } k = 1, \\ \mathrm{TL}_1^n + \sum_{i=1}^{k} \mathrm{TS}_i^n + \sum_{j=1}^{k-1} \mathrm{TT}_j^n, & \text{if } 2 \le k \le s. \end{cases} \tag{4}$$

2.3. The Main Parameters of Short-Turn Vehicles. Based on the notations and two hypotheses in Section 2.1, a state model of short-turn vehicles operation is established. In this study, it is assumed that passengers waiting at the station will give priority to the first arriving vehicle. We define the operating area of a short-turn vehicle as U where $U = (u_1, u_2, K, u_q), U \subseteq (1, 2, 3, \ldots, s)$. When the short-turn vehicle arrives at one of the stations, all the passengers whose travel destinations are only in U while the passengers whose destination is not in U will wait for the next full-length vehicle. That means, the passenger whose travel destination is in U will not continue to wait.

2.3.1. Number of Passengers Boarding on Short-Turn Vehicle. We can calculate the probability that the passengers waiting at station i will have a destination within the U range from OD data. The probability model is expressed as follows:

$$P_i = \frac{\sum_{j=i+1}^{u_q} \mathrm{OD}_{ij}}{\sum_{j=i+1}^{s} \mathrm{OD}_{ij}}, \quad i \in U. \tag{5}$$

Then, how many passengers in station i ride the short-turn vehicles can be calculated, and the formula is expressed as follows:

$$Q_i^q = P_i \cdot \sum_{m=1}^{M} \rho_i^m \cdot H_i^{n-1,q,m}. \tag{6}$$

2.3.2. The Number of Passengers on the Full-Length Vehicle next to a Short-Turn Vehicle. In the study, only alternate departure between the short-turn and full-length vehicle is discussed. The scenario of the continuous short-turn vehicle departure will not be considered. nq stands for the number of short-turn vehicles that departs after $n-1$. Because not all passengers will take the short-turn vehicle, after nq departs from station i, some passengers will keep waiting, until bus n arrival. The number of passengers who continue to wait for n is calculated using the following equation:

$$Q^{q'} = (1 - P_i) \cdot \sum_{m=1}^{M} \rho_i^m H_i^{n-1,q,m} + \sum_{m'=1}^{M'} \rho_i^m H_i^{q,n,m'}, \tag{7}$$

where M' represents the number of arrival rate changes during $H_i^{q,n}$.

2.3.3. Dwell Time of Short-Turn Vehicles. The dwell time of a short-turn vehicle at the certain station is expressed as follows:

$$\mathrm{TS}_i^q = \max\left\{ \sum_{a=u_1}^{i} \mathrm{OD}_{ai}, Q_i^q \right\} \cdot t' + \varepsilon, \quad i \in U. \tag{8}$$

2.3.4. Dwell Time of Full-Length Vehicle next to Short-Turn Vehicles. After a short-turn strategy is applied, the next full-length vehicle will take fewer passengers. As a result, the dwell time of the full-length vehicles in U will change. When full-length vehicles travel beyond U, its dwell time can be calculated by (1). We combine (1) and (8) together, and the dwell time can be expressed as follows:

$$\mathrm{TS}_i^n = \begin{cases} \max\left\{ \sum_{a=1}^{i} \mathrm{OD}_{ai}, Q_i^{q'} \right\} \cdot t' + \varepsilon, & \text{if } i \in U, \\ \max\left\{ \sum_{a=1}^{i} \mathrm{OD}_{ai}, \sum_{m=1}^{M} \rho_{\mathrm{i}}^m \cdot H_{\mathrm{i}}^{n-1,n,m} \right\} \cdot t' + \varepsilon, & \text{if } i \notin U. \end{cases} \tag{9}$$

It can be found once the parameters of speed, OD data, and the passenger arrival rate are given in the model. The operational state of full-length vehicles and short-turn vehicles can be calculated. Obviously, since the IC card data can provide information about the three parameters above, it is most rational to use IC card data. In Section 3, we introduced how to extract the data required by the model and analyzed the impact of different factors on the public transportation system.

3. Bus Operation Analysis without Short-Turn Dispatching Strategy

3.1. Data Processing. The smart card data of a bus line (No. 973) in Beijing are used to study the impact of the short-turn strategy on the stability of the headway. There are 50 stations on this line. The data are collected from five working days between April 11, 2017 and April 15, 2017, and the weather is sunny. The bus route map is shown in Figure 2.

Original data mainly contain the following fields, and detailed explanation is given in Table 2.

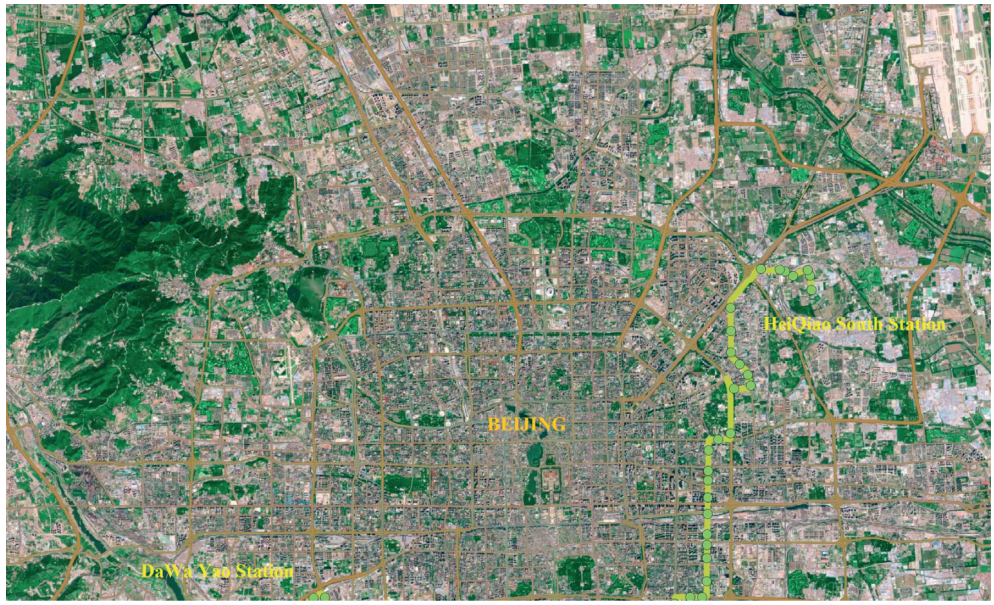


Figure 2: Bus line (sourcing from Amap: https://ditu.amap.com/).

Table 2: Data field description.

Field	Type	Comment
GRANT_CARD_CODE	String	Card ID
LINE_CODE	String	Line code
VEHICLE_CODE	String	Vehicle code
BANCIID	Int	Shift code
ON_STATIONID	Int	Get-on station ID
ON_STATIONNAME	String	Get-on station name
OFF_STATIONID	String	Get-off station ID
OFF_STATIONNAME	String	Get-off station name
OFF_STATION_TIME	String	Get-off card time
LOC_TREND	Int	Bus operation direction

We can find that there is no record of the time in the table when a passenger boarded. Since December 28, 2014, a passenger in Beijing has been charged by mileage in public transport. Passengers need to swipe their cards when getting on and off the bus. In order to save space for information storage, boarding time has not been recorded, and it is useless for charging [25].

The data required for the simulation are extracted.

3.1.1. The Distance of Station Interval. The distance can be calculated with the latitude and longitude of each station, and the result is calibrated by an investigation, and data are shown in Table 3.

3.1.2. Speed Data. The short-turn strategy is usually applied during rush hours. Therefore, we extracted the data of speed and passenger arrival rate from 7 to 10 am, including the morning peak.

Firstly, the arrival time of a bus at a station can be regarded as the average time of all card-wiping records when getting off the bus [25]. Based on the arrival time, travel time between the two stations can be calculated by distance and travel time. We divide the daily data by the unit of 10 minutes. Then, we calculate travel speed separately and take the average travel speed of the same group each day. In this way, we can get the average speed of each station interval every 10 minutes. As shown in Table 4, the unit is m/s and speed_1 indicates the speed of a bus traveling between each station within the first 10 mins. The data of 4 stations and 5 periods are only listed.

3.1.3. Passenger Arrival Rate Data. The smart card data record the information of the boarding and disembarking sites for each passenger, and the passenger OD data of bus line can be counted. The result is shown in Figure 3.

According to the statistical yearbook published by the Beijing Transportation Operations Coordination Center (TOCC), 85% of passengers choose to pay by card and 15% of passengers choose to pay cash. Therefore, the OD data obtained from the card data cannot present a complete map of the number of passengers. We multiplied the OD data by a coefficient of 1.17 to get the final passenger OD data.

Since only get-off time is recorded, passenger's boarding time needs to be supplemented to calculate the arrival rate in each station. Bus arrival time can be regarded as passenger's boarding time, which has been obtained when calculating the speed data. Then, the data are grouped every 10 minutes by boarding time, and the number of passengers in each group can be counted. The number of passengers transferred to the arrival rate data is shown in Table 5, the unit is pax/min, and time_1 indicates the passenger arrival rate at each station in the first 10 minutes. The data of 4 stations and 5 periods are only listed.

3.2. Numerical Simulation. Three scenarios are designed to discuss the impact of different parameters (bus travel speed and passenger arrival rate) on reliability. In each scenario, 15

Table 3: Distance of station interval.

Interstation index	Distance (m)
1	118
2	348
3	827
4	846
5	509
6	539
7	356
8	827
9	270
10	1010
11	432
12	384
13	3075
14	1055
15	1887
16	2004
17	865
18	996
19	1083
20	1506
21	826
22	1669
23	817
24	522
25	792
26	808
27	659
28	643
29	1002
30	353
31	880
32	477
33	868
34	868
35	342
36	2088
37	680
38	462
39	755
40	495
41	886
42	1177
43	2109
44	787
45	155
46	524
47	523
48	338
49	688

Table 4: Speed data.

Station interval ID	speed_1	speed_2	speed_3	speed_4	speed_5
1	5.52	5.52	5.52	5.52	5.52
2	6.90	6.90	6.90	6.90	6.90
3	6.90	6.90	6.90	6.90	6.90
4	5.99	5.99	5.99	5.99	5.99

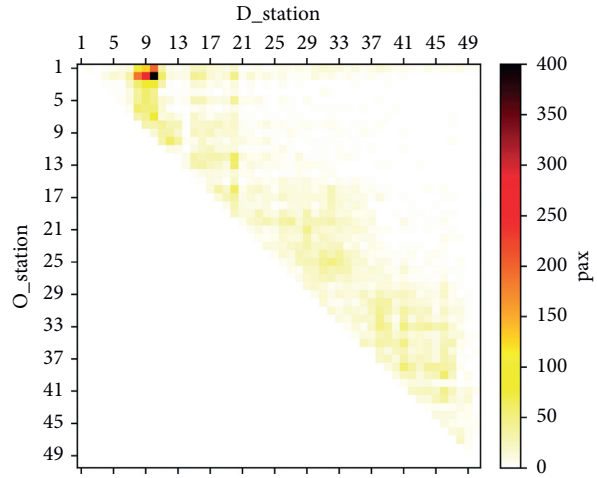


Figure 3: OD data.

Table 5: Passenger arrival rate.

Station	time_1	time_2	time_3	time_4	time_5
1	2.98	3.22	2.04	3.47	2.98
2	3.92	4.45	3.43	4.98	3.92
3	1.60	1.56	1.10	1.52	1.60
4	0.76	0.70	0.79	0.64	0.76

buses are simulated, and the interval of arrival time is 300 s. The initial time of the system is 0 s. These scenarios are described in Table 6.

The scenarios design is summarized as follows.

Scenario 1 indicates that no control strategy is adopted. It can be considered as a basic scenario.

In scenario 2, bus speed is set to a constant. We are able to assess the impact of passengers' arrival rate changes on bus operation by comparison with scenario 1.

In scenario 3, the passengers' arrival rate is unchanged with time. We can analyze the impact of speed changes on bus operation by comparison with scenario 1.

Applying the passenger arrival rate and speed data in the above different scenarios into the model proposed in Section 2.2, we can calculate bus trajectories and bus headway, as shown in Figure 4.

In scenario 1, we can find the phenomenon of a large interval and bus bunching at some stations from Figure 4(a). This phenomenon indicates that the unstable distribution of bus travel speed and passenger arrival rate makes the contribution with unreliability. Meanwhile, the unreliability increases with travel distance. This phenomenon is consistent with our daily bus ride experience, so it also verifies the validity of the model in Section 2.2.

Scenario 2 (Figure 4(d)) reveals that even if the speed of buses between two stations remains unchanged, the unreliability also occurs under the condition that the passenger arrival rate is unstable, and it spreads further backward. Such a conclusion is drawn out of Figure 4(d) that the headway of the bus in the first few stations changes rapidly because there are more bus passengers in the first few stations. Therefore, we believe that the operation of buses is affected by the

Table 6: Scenarios description.

Scenario	Speed	Passenger arrival rate
Scenario 1	Data from smart card data	Data from smart card data
Scenario 2	8 m/s	Data from smart card data
Scenario 3	Data from smart card data	Passenger arrival rate at any time is unchanged

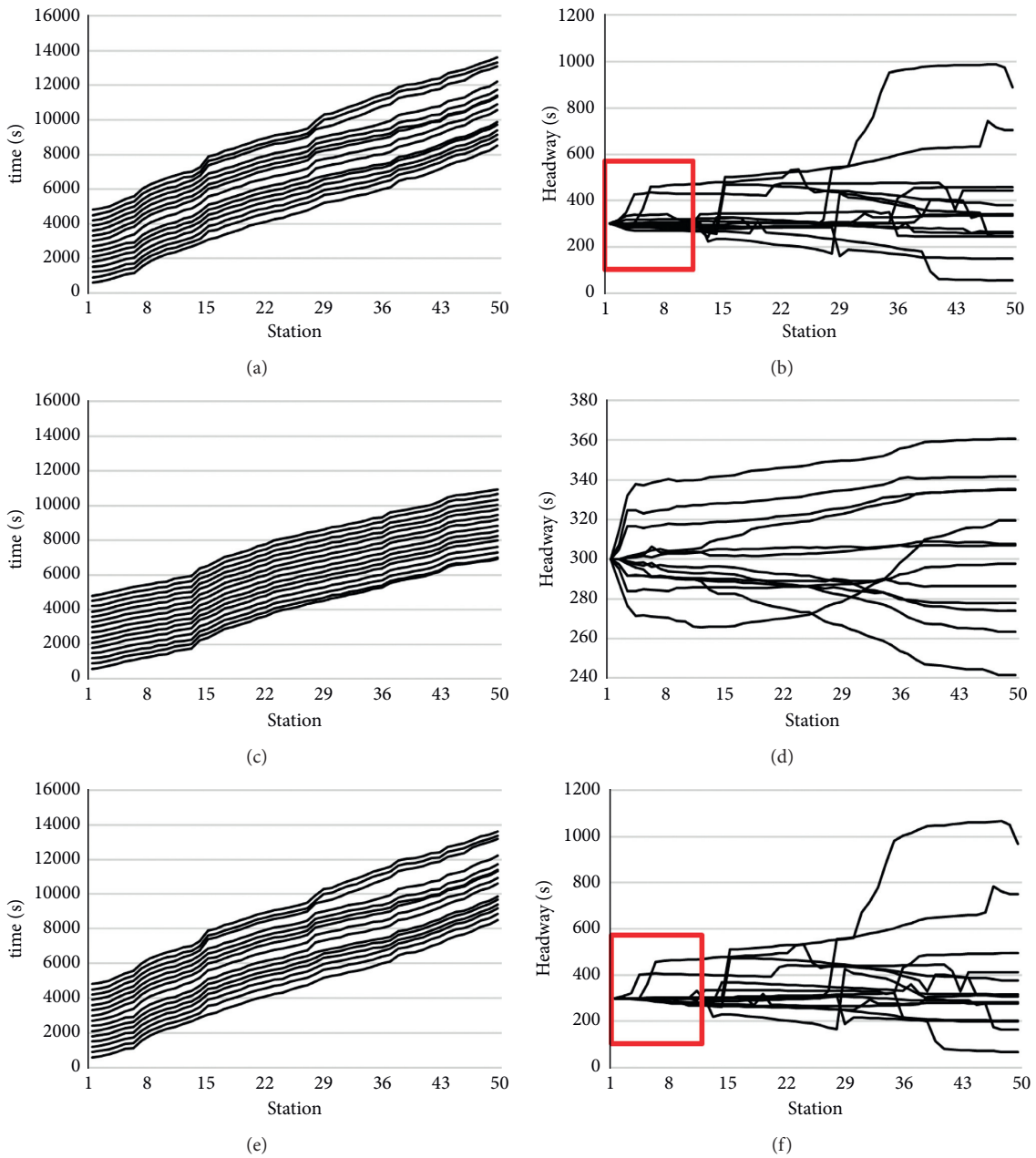


Figure 4: Simulation results in three scenarios: (a) bus trajectories for scenario 1 and (b) headway for scenario 1; (c) bus trajectories for scenario 2 and (d) headway for scenario 2; (e) bus trajectories for scenario 3 and (f) headway for scenario 3.

number passengers boarding and getting off in a certain degree. Unbalanced passenger distribution results in more unreliability.

By comparing scenario 3 (Figure 4(f)) with scenario 1 (Figure 4(b)), it is found that bus travel speed has a higher impact on reliability. When the passenger arrival rate does not change, the operation of the bus is still highly unreliability due to the change of bus travel speed. If the traffic status can be changed or the bus can adjust its traveling speed, the reliability of the bus will be changed. The method of bus speed control has been studied in previous studies [26, 27]. The best case is that the bus speed can be controlled

precisely like the subway and travel with a certain headway. That means, bus lanes may be helpful in improving reliability.

Unfortunately, it is difficult to change the speed of driving freely when buses travel in ordinary roads. In scenario 3, buses travel more reliable in the first few stops comparing with scenario 1. This phenomenon can be found from red box in Figures 4(b) and 4(f). That means, if the passenger arrival rate does not change, it will help to improve reliability, and it has been proved in scenario 2. It shows that the number of onboarding passengers has an impact on the operation of public transport. Moreover, when the number of passengers increases and the arrival rate changes more obviously, it will have a greater impact on the operational reliability just like the result shown in scenario 2 (Figure 4(d)).

In this paper, we focus on releasing the uneven distribution of passengers during peak periods to improve bus operation reliability. A short-turn strategy is proposed. But, different departure time and operation area of short-turn vehicles can make different influences on bus operational reliability. In the next section, we will discuss how to design a short-turn strategy to improve bus operation reliability.

4. Short-Turn Strategy

The most important factor to design a short-turn strategy is to establish the operating area and departure time of the short-turn vehicles. This study suggests that the operating area of short-turn vehicles should be the most concentrated section of the passenger's OD distribution, and the departure time of short-turn vehicles should be at the change point of speed or passenger arrival rate. These methods will be explained in detail in this section.

4.1. Operating Area. When we can get the OD data of the bus line based on smart card data, the interval of stations with larger passenger demand is selected out of OD data as the operating area of short-turn vehicles.

The operating area should satisfy two conditions simultaneously. Firstly, the ratio of OD in the selected area to the total OD of the line is larger, defined as p_i^1. Secondly, the ratio of OD in the selected area to the total number of passengers on board in the area is larger, defined as p_i^2 (in some cases, these two conditions may not be met simultaneously; selection should be done according to the actual situation because the passenger's demand in some lines is more evenly distributed). This method is described as follows:

Step1: given the parameter u, which is the number of stations that short-turn vehicles should operate.

Step2:

$i = 0$

repeat

$i = i + 1$

if $i < s - u$ then

$p_i^1 = (\sum_{a=i}^{i+u}\sum_{b=i}^{i+u+1} \mathrm{OD}_{ab} / \sum_{a=i}^{s}\sum_{b=i}^{s} \mathrm{OD}_{ab})$
and $p_i^2 = (\sum_{a=i}^{i+u}\sum_{b=i}^{i+u+1} \mathrm{OD}_{ab} / \sum_{a=i}^{s}\sum_{b=i}^{s} \mathrm{OD}_{ab})$

else

end if

Step3: we can get set1 and set2 from step2, where set1 = $(p_i^1, p_{i+1}^1, \ldots, p_{s-u}^1)$ and set2 = $(p_i^2, p_{i+1}^2, \ldots, p_{s-u}^2)$.

Get the larger values in set1 and set2. When the values can be found at station i, we can conclude that the operating area is $\mathrm{i} - i + u$.

4.2. Departure Time. The optimal frequency of the combination of the short-turn vehicles and the full-length vehicles has been discussed in previous studies. But, most of them did not give a concrete method to settle the departure time of short-turn buses. The operation of public transport is greatly impacted by the state of traffic flow and the demand of passengers; thus, this research believes that the departure time of the short-turn vehicle should be the time point at which the traffic flow speed is reduced and the passenger demand increases.

The cumulative sum control chart (CUSUM) method in change point theory can be applied to find the time, which is widely used in economics, meteorology, and other fields. The concrete calculation method is described as follows [20].

Assume there is a time series $X_1, X_2, X_3, \ldots, X_T$, and the change point can be found by 5 steps.

Step 1. We can calculate the average value of the series $\overline{X} = \mathrm{sum}(X_1, X_2, X_3 \ldots X_T)/T$.

Step 2. Start the cumulative sum at zero by setting $S_0 = 0$.

Step 3. Calculate the cumulative sums of differences between the values and the average with formula $S_i = S_{i-1} + (X_i - \overline{X})$. A series can be calculated. Then, we can get $S_{\mathrm{diff}} = S_{\max} - S_{\min}$.

Step 4. Generate a bootstrap sample, which means to reorder the origin series randomly. Then, return step2, step 3, and a new difference of the bootstrap CUSUM can be calculated as S_{diff}^1.

Step 5. Repeat step4 and get more S_{diff}^1, and the number of repetitions is N. Let X be the number of bootstraps for which $S_{\mathrm{diff}}^1 < S_{\mathrm{diff}}$. Then, the confidence level that a change occurred as a percentage is calculated as follows:

$$\text{confidence level} = 100\frac{X}{N}\%. \tag{10}$$

It is generally believed that if the confidence level is over 90%, a change point can be detected in the series. A better estimate can be obtained by increasing the number of N. However, 1000 bootstraps are sufficient for most purposes. Once a change point is detected, $S_m = \max_{i=0,1,\ldots,T}|S_i|$ can be calculated. The point m is estimated to be the last point before the change occurs. It can be considered as the departure time of short-turn vehicles.

5. Result

5.1. Short-Turn Strategy. Passenger OD data of bus line No. 973 are selected to analyze, and the method has been proposed in Section 4.1 to determine the starting station and the terminal station of short-turn vehicles. Considering the passenger demand and the total length of the line, we take u value of 12 as an example. The values of set1 and set2 can be calculated, as shown in Table 7.

In Table 7, the value of set2 is also larger when set1 is the largest at the first value. The value of set1 means that the number of passengers boarded on 1–12 station accounts for 22% of the total number of passengers on the entire line. The value of set2 means that passengers who boarded on station 1–12 and still get off on station 1–12 account for 56% of a total number of passengers on station 1–12. Therefore, the operation area of short-turn vehicles is set to station 1–12. Next, we need to determine the departure time of short-turn vehicles.

The analysis in previous scenarios shows that the bus operation is greatly affected by the traffic flow (speed). Therefore, this study takes the change point of the bus travel speed as a condition for judging the departure of short-turn vehicles. Taking the speed of the second interval as an example, the CUSUM method is applied to calculate accumulation difference of this series. The result is shown in Figure 5, and $S_{\text{diff}} = 10.84$.

The original series is reordered to get $S^1_{\text{diff}}, S^2_{\text{diff}}, S^3_{\text{diff}}, \ldots, S^{1000}_{\text{diff}}$, and there are 925 of them less than S_{diff}. It means that the series has a change point at 92.5% confidence level. When $S_m = \max_{i=0,1,\ldots,T}|S_i|$, the change point is m. There are 35 values in the series, which means there are 35! cases of random ordering. In this research, 1000 of them are selected at random. We can get different results for different bootstraps, and the random progress is repeated 10 times to prove that the confidence level can meet the requirements. The results are 91.8%, 92.3%, 89.9%, 92.1%, 91.6, 92.5%, 91.8%, 90.6%, 92%, and 93.1%.

There are 15 buses in the simulation, and the departure frequency is set to 300 s. The departure time of the last bus should be at 4500 s. Because speed changes every 600 s, 4500 s corresponds to the 8th time periods. In the first 8 periods, the cumulative sum $S_5 = 1.97$is the maximum, which means the change point of speed at fifth time periods.

The cumulative sum series of the first 4 station intervals is calculated with the same method. As shown in Figure 6, change points are found in these 4 series, all of them at the same periods. We can confirm that the change points occur in 5th period. That is, the departure time of short-turn vehicles should be in the 5th period (2400–3000 s).

According to the hypothesis in Section 2, when a short-turn vehicle is in operation, passengers can choose either a short-turn vehicle or a full-length vehicle. It is generally believed that passengers choose the bus that arrives first. Passengers continue to accumulate during $H_i^{n-1,n}$. When there is one short-turn vehicle between bus $n-1$ and n, not all the accumulated passengers are suitable for the short-turn vehicle. In order to ensure the short-turn vehicle takes more passengers without causing waste, the departure time of the short-turn vehicle should be as close as possible to bus n, rather than completely in the middle of bus $n-1$ and n [7]. We assume that the two full-length vehicles will depart at 0s

Table 7: Passenger ratio data.

Starting station	set1	set2
1	0.22	0.56
2	0.19	0.55
3	0.12	0.45
4	0.11	0.45
5	0.11	0.44
6	0.1	0.43
7	0.1	0.42
8	0.09	0.4
9	0.1	0.42
10	0.09	0.38
11	0.08	0.35
12	0.09	0.35
13	0.07	0.31
14	0.06	0.26
15	0.07	0.27
16	0.08	0.29
17	0.08	0.29
18	0.08	0.33
19	0.09	0.35
20	0.1	0.37
21	0.09	0.37
22	0.1	0.37
23	0.1	0.37
24	0.09	0.34
25	0.09	0.33
26	0.08	0.31
27	0.09	0.34
28	0.09	0.36
29	0.09	0.4
30	0.11	0.47
31	0.11	0.5
32	0.11	0.55
33	0.11	0.61
34	0.1	0.66
35	0.11	0.79
36	0.1	0.88
37	0.09	0.92
38	0.08	0.98
39	0.06	1

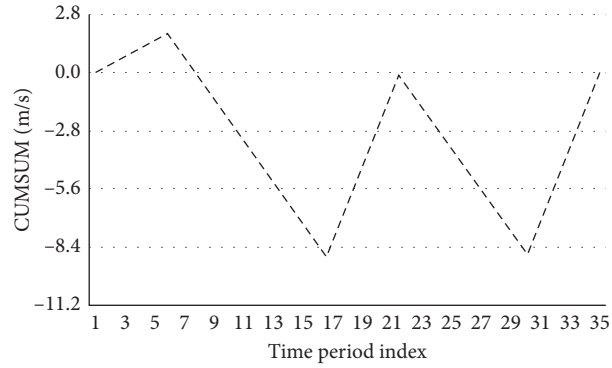


Figure 5: CUSUM chart of the second station interval.

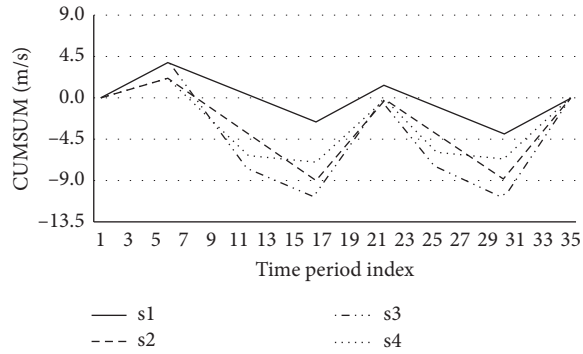


Figure 6: CUSUM charts of the first 4 station interval.

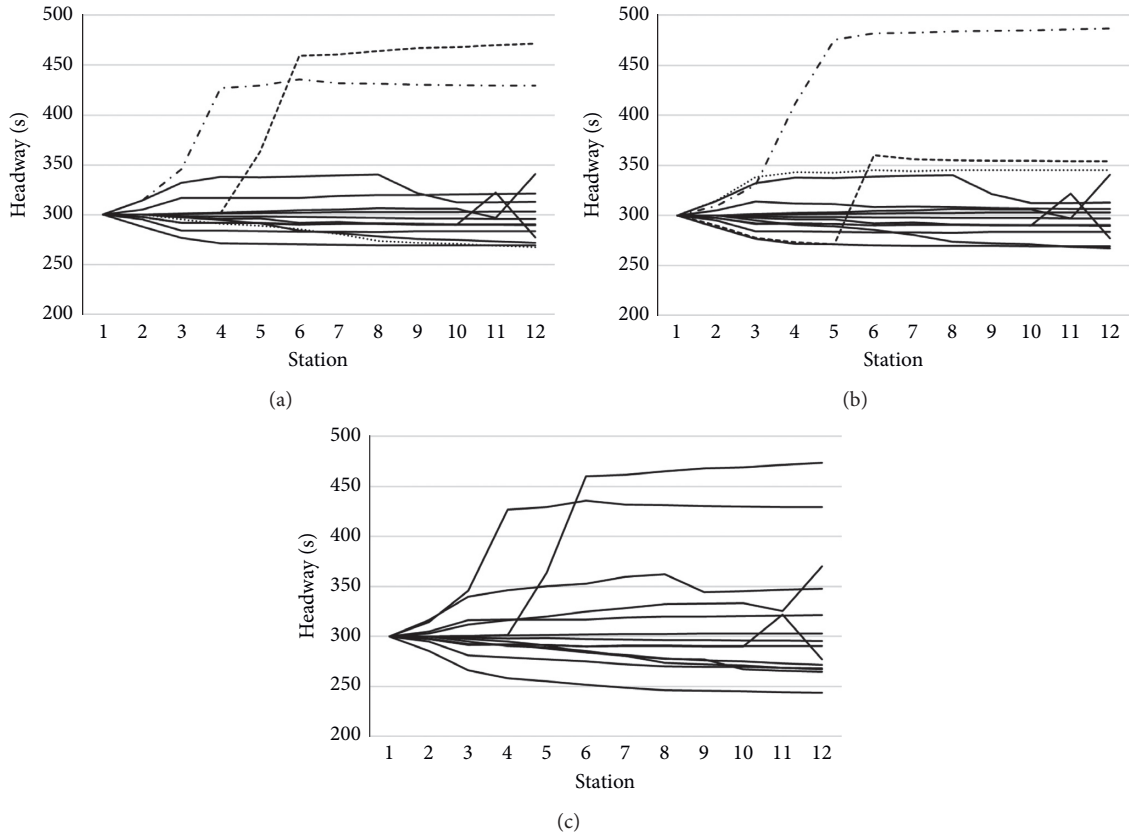


Figure 7: (a) Strategy A, (b) strategy B (dotted lines represent the headway of the full-length vehicles which adjoins the short-turn vehicles), and (c) strategy C.

and 300s, respectively. The departure time of the short-turn vehicle should be after 150 s. Therefore, in this case study, the departure time of short-turn vehicles should be set at the 2600 s and 2900 s because the departure time of the adjacent full-length vehicles is 2400 s, 2700 s, and 3000 s. We define it as strategy B. The basic strategy is defined as strategy A (without short-turn vehicles). Meanwhile, in order to verify the rationality of dispatch the short-turn vehicles at the change point, another departure plan is given randomly, for example, the departure time of short-turn vehicles at the 1100 s and 1700 s. In addition, the departure time of the adjacent full-length vehicles is at 900 s, 1200 s, 1500 s, and 1800 s. The operation area is still in station 1–12. We define it as strategy C.

5.2. Simulation Result. These three strategies are simulated separately. The bus headway of station 1–12 is shown in Figure 7. The mean absolute error between bus headway and the departure interval is used to evaluate different strategies, which can be formulated as follows:

$$\mathrm{MAE} = \frac{\sum_{n=2}^{15}\sum_{i=1}^{12}\left|H_i^{n-1,n} - H_1^{n-1,n}\right|}{15^*12}. \quad (11)$$

Through the numerical simulation results, it can be found that MAE of the first 12 stations is 46.9 s without the short-turn strategy being applied (strategy A), while the value changes to 44.7 s after the strategy B is employed. The overall reliability of the 15 buses in the system increases by 4.7%. It can be seen from the result that strategy B can change full-length vehicles' operating state obviously, especially for these vehicles whose departure time are close to short-turn vehicles'. Dotted lines are used in Figure 7(b) to mark the three buses. The overall reliability of the first 12 stations of these three buses increased by 10.4%. Meanwhile, the reliability of vehicles that depart further behind the short-turn vehicles will also be affected, but the effect will gradually weaken. For strategy C (Figure 7(c)), MAE is 50.2 s, that means the headway distribution is more discrete than the other two strategies. The reliability is even lower than the without short-turn strategy case. Compared with the simulation results of the three strategies, it can be explained that if the short-turn strategy cannot be designed scientifically, it will lead to the reduction of the overall reliability of the bus operation. Therefore, the reasonable design of the departure plan for short-turn vehicles should be considered carefully by the bus operating company. One of the better options is that the short-turn vehicles depart at the change point in the peak period (Figure 7(b)) and operate in the area with larger passenger demand.

6. Discussion and Conclusion

6.1. Discussion. It can be seen from the result that we calculated the operating area should be from station 1 to 12 based on OD data (Section 5.1). However, there are some subjective factors to calculate this area using the method in Section 4.1 such as the value of u. In this paper, u is set 12 because possible carrying capacity of the short-turn vehicle, total length of bus line, and turnover efficiency are considered. Different lines or cities may have different values.

Meanwhile, it should be noted that an implicit assumption in the simulation of this study is that once the change point is found, the short-turn vehicles can arrive at the target station. But, it seems to be unrealistic. Because the short-turn vehicles take a certain amount of time from the departure station to the target station, the travel time needs to be considered to achieve the desired improvement. If we can predict the traffic congestion state and the passenger flow more accurately through the forecasting method, we may be able to find a more suitable timetable of short-turn vehicles. Therefore, when the bus company uses this method, it is necessary to forecast the traffic state for a better short-turn strategy [28, 29].

The essence of the short-turn strategy is to adjust the number of passengers boarding full-length vehicles at some stations. That means, if the passenger number of each station can be well optimized, it will be helpful to improve the reliability of the bus operation. This view also provides a new idea for the operation and management of public transportation. If authorities hope to attract more passengers to choose to take the bus by improving bus operation reliability, not only the method of speed control and the express bus could be useful but the management of passenger demand is also very important. For example, we can restrict passenger flow at some stations to improve the stability of passenger numbers during peak hours [30, 31]. Also, we can use the transportation information system to publish real-time public transportation information such as the degree of congestion in the bus or bus arrival time to guide passenger demand [32].

The result shows in Section 3 that passenger demands may impact bus operation reliability. In addition, the larger the number of passengers, the greater the impact. This also implies that if this method is used in another bus line, it will have different effects. Therefore, we suggest that it would be better to adopt the short-turn strategy in the rush hours and the bus lines with more passengers. The result provides an idea to explain the capacity paradox of the transit system that has been proved [33], which may be caused by the timing of departure.

6.2. Conclusion. In this study, a bus operation state model considering the operation of short-turn vehicles was firstly established. Then, two parameters affecting the operation state of buses are analyzed, which are passenger arrival rate and bus travel speed. Result in Section 3 shows that passenger distribution affects the operating state of full-length vehicles in the system. Then, the short-turn strategy is proposed which contains two parts based on the number of passengers to determine the operation area of short-turn vehicles and the departure time using change point theory. Finally, the smart card data of buses are used to verify the proposed method through numerical simulation. The simulation result shows that if a short-turn strategy is designed by using the method mentioned in Section 4, the reliability of the full-length vehicles in the operation area of the short-turn vehicles increases by 4.7%. But, if the departure time of short-turn vehicles is set randomly, it will reduce the reliability by 7%, even the operating area is reasonable. Affected by passengers' boarding and dropping, the short-turn strategy may affect the operation of full-length vehicles in different degree. When the number of passengers increases and the arrival rate changes more obviously, the departure plan of short-turn vehicles will have a greater impact on the operation reliability of full-length vehicles.

By comparing with different departure strategies, we find that the departure time and operating area of short-turn vehicles determined in Section 4 are reasonable. It can optimize reliability to a certain degree. Meanwhile, the strategy can be deployed in real-time systems, and we can formulate the optimized short-turn strategy simply and quickly.

Since the short-turn vehicles generally run in the upward and downward directions, the short-turn strategy should have an impact on operating of full-length vehicles in two directions. But, this research only covers a certain direction of operation. In future research, the impact of the short-turn

strategy on the entire public transport system in two directions shall be touched. In addition, this study only considers short-turn vehicles depart at the change point. The optimization of the cost is not considered, such as passengers' waiting time and bus company management cost. Also, sensitivity analysis of bus frequency and passengers demand need to be considered. These points will also be supplemented in future research.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (nos. U1811463, 51878020, 51908018, and 51308018) and Natural Science Foundation of Hebei Province (no. E2018407051). Useful suggestions were given by Dr. Yue Liu from the University of Wisconsin-Milwaukee and Dr. Huanmei Qin from the Beijing University of Technology which are also acknowledged.

References

[1] J. A. Sorratini, R. Liu, and S. Sinha, "Assessing bus transport reliability using micro-simulation," *Transportation Planning and Technology*, vol. 31, pp. 303–324, 2008.

[2] J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 2-3, pp. 191–229, 2001.

[3] M. A. Turnquist and L. A. Bowman, "The effects of network structure on reliability of transit service," *Transportation Research Part B: Methodological*, vol. 14, no. 1-2, pp. 79–86, 1980.

[4] X. Chen, L. Yu, Y. Zhang, and J. Guo, "Analyzing urban bus service reliability at the stop, route, and network levels," *Transportation Research Part A: Policy and Practice*, vol. 43, no. 8, pp. 722–734, 2009.

[5] A. Mark, E. Amir, and E. Israel, "Optimal control of headway variation on transit routes," *Journal of Advanced Transportation*, vol. 20, pp. 73–88, 2010.

[6] P. R. Tétreault and A. M. El-Geneidy, "Estimating bus run times for new limited-stop service using archived AVL and APC data," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 6, pp. 390–402, 2010.

[7] P. G. Furth, "Zonal route design for transit corridors," *Transportation Science*, vol. 20, no. 1, pp. 1–12, 1986.

[8] S. Liang, S. Zhao, C. Lu, and M. Ma, "A self-adaptive method to equalize headways: numerical analysis and comparison," *Transportation Research Part B: Methodological*, vol. 87, pp. 33–43, 2016.

[9] S. Liang, M. Ma, and S. He, "Multiobjective optimal formulations for bus fleet size of public transit under headway-based holding control," *Journal of Advanced Transportation*, vol. 2019, Article ID 2452348, 14 pages, 2019.

[10] S. Liang, M. Ma, S. He, H. Zhang, and P. Yuan, "Coordinated control method to self-equalize bus headways: an analytical method," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1175–1202, 2019.

[11] A. Ceder, "Optimal design of transit short-turn trips," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1221, pp. 8–22, 1989.

[12] P. D. Site and F. Filippi, "Service optimization for bus corridors with short-turn strategies and variable vehicle size," *Transportation Research Part A: Policy and Practice*, vol. 32, pp. 19–38, 1998.

[13] C. E. Cortés, S. Jara-Díaz, and A. Tirachini, "Integrating short turning and deadheading in the optimization of transit services," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 5, pp. 419–434, 2011.

[14] A. Tirachini, C. E. Cortés, and S. R. Jara-Díaz, "Optimal design and benefits of a short turning strategy for a bus corridor," *Transportation*, vol. 38, no. 1, pp. 169–189, 2011.

[15] P. G. Furth, "Short turning on transit routes," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1108, pp. 42–52, 1987.

[16] D. Canca, E. Barrena, G. Laporte, and F. A. Ortega, "A short-turning policy for the management of demand disruptions in rapid transit systems," *Annals of Operations Research*, vol. 246, no. 1-2, pp. 145–166, 2016.

[17] A. Ceder, S. Hassold, and B. Dano, "Approaching even-load and even-headway transit timetables using different bus sizes," *Public Transport*, vol. 5, pp. 193–217, 2016.

[18] H. Zhang, S. Zhao, H. Liu, and L. Jin, "A dynamic short-turning bus control for uncertain demand," *Journal of Advanced Transportation*, vol. 2017, Article ID 7392962, 9 pages, 2017.

[19] G. F. Newell, "Dispatching policies for a transportation route," *Transportation Science*, vol. 5, no. 1, pp. 91–105, 1971.

[20] S. J. Berrebi, E. Hans, N. Chiabaut, J. A. Laval, L. Leclercq, and K. E. Watkins, "Comparing bus holding methods with and without real-time predictions," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 197–211, 2018.

[21] W. A. Taylor, "Change-point analysis: a powerful new tool for detecting changes," 2000, http://www.variation.com/files/articles/changepoint.pdf.

[22] L. Horváth and M. Hušková, "Change-point detection in panel data," *Journal of Time Series Analysis*, vol. 33, no. 4, pp. 631–648, 2012.

[23] B. Safari, "Trend analysis of the mean annual temperature in Rwanda during the last fifty two years," *Journal of Environmental Protection*, vol. 3, no. 6, pp. 538–551, 2012.

[24] Y. Ji, X. Yang, and Y. Du, "Optimal design of a short-turning strategy considering seat availability," *Journal of Advanced Transportation*, vol. 50, no. 7, pp. 1554–1571, 2016.

[25] Y. Zhou, L. Yao, Y. Chen et al., "Bus arrival time calculation model based on smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 1634–1638, 2016.

[26] C. F. Daganzo and J. Pilachowski, "Reducing bunching with bus-to-bus cooperation," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 267–277, 2011.

[27] W. Wu, W. Ma, K. Long et al., "Designing sustainable public transportation: integrated optimization of bus speed and holding time in a connected vehicle environment," *Sustainability*, vol. 8, 1170 pages, 2016.

[28] Y. Bai, Z. Sun, B. Zeng et al., "A multi-pattern deep fusion model for short-term bus passenger flow forecasting," *Applied Soft Computing*, vol. 58, pp. 669–680, 2016.

[29] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 453–466, 2018.

[30] C. F. Daganzo, "A headway-based approach to eliminate bus bunching: systematic analysis and comparisons," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 913–921, 2009.

[31] H. Zhang, S. Zhao, H. Liu et al., "Design of limited-stop service based on the degree of unbalance of passenger demand," *PLoS One*, vol. 13, no. 3, 2018.

[32] G. Gentile, S. Nguyen, and S. Pallottino, "Route choice on transit networks with online information at stops," *Transportation Science*, vol. 39, no. 3, pp. 289–297, 2005.

[33] Y. Jiang and W. Y. Szeto, "Reliability-based stochastic transit assignment: formulations and capacity paradox," *Transportation Research Part B: Methodological*, vol. 93, pp. 181–206, 2016.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7981791, 15 pages
https://doi.org/10.1155/2020/7981791

WILEY | Hindawi

Research Article

Assessing the Impacts of Autonomous Bus-on-Demand Based on Agent-Based Simulation: A Case Study of Fuyang, Zhejiang, China

Zhikang Zhai, Ying Yang, Yu Shen, Yuxiong Ji, and Yuchuan Du

Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Yu Shen; yshen@tongji.edu.cn

Received 3 February 2020; Revised 13 October 2020; Accepted 7 November 2020; Published 23 November 2020

Academic Editor: Prakash Ranjitkar

This paper envisions and assesses the performance of an autonomous bus-on-demand (ABoD) system. We take Fuyang, Zhejiang, China, as the study area to investigate the spatiotemporal distribution of bus travel demand during workdays, and we propose replacing inefficient bus routes with the ABoD system. Agent-based models with various bus dispatching and operation control strategies are constructed to evaluate the performance of the ABoD system. The behaviors and interactions of the agents, passengers, autonomous buses, and a control center are designed. After the verification of the simulated bus travel demand with real-world demand, a series of scenarios with various ABoD operation strategies are simulated. The simulation results show that, in comparison with both current fixed-schedule bus services and the optimized bus dispatching strategies, the ABoD system occupies fewer road resources and utilizes bus vehicles more efficiently. Besides, the system is adaptive to the sudden surge in bus travel demand and is economically sustainable.

1. Introduction

The concept of a demand-responsive public transit system has been proposed and experimented with Daganzo [1] and Wilson and Hendrickson [2]. Decades of effort have been invested in designing and operating these on-demand public transit services. Ceder and Wilson [3] constructed a bus network planning model based on real demand to minimize time consumption. Dial [4] designed a dial-a-ride transit system for passengers to book vehicles by phone in advance, which can be regarded as the precursor of modern mobility-on-demand services, such as Uber and Didi. Adebisi and Hurdle [5] showed that demand-responsive bus services are suitable for low ridership, low density, and scattered demand. Chang and Schonfeld [6] designed a flexible route bus subscription system to provide feeder services to a single destination, to minimize total cost in comparison with fixed-route conventional bus systems. Quadrifoglio et al. [7] examined the impact on oil productivity of demand-responsive transit systems and concluded that the same demand can be satisfied employing fewer vehicles. However, due to the constraints of information and communication technology (ICT) and labor costs, demand-responsive public transit services have seldom been successfully implemented.

In recent years, thanks to the rapid development of ICT and significant progress in autonomous driving technologies, autonomous bus-on-demand (ABoD) services are expected to manifest [8]. Although a few studies have envisioned various proposals for on-demand bus systems [9, 10], the majority of works on the potential application of autonomous technologies treat autonomous vehicles (AVs) primarily as personal vehicles [11–13]. And some works look into the fact that how AVs can be embedded into public transit systems. Vakayil et al. [14] explore a hybrid transit system with on-demand AVs as an additional service to improve metro connectivity. Shen et al. [15] propose the integration of shared AVs into public transportation systems for the first-mile service in Singapore. However, only a few studies have focused on the feasibility of an ABoD system replacing traditional bus services completely. Winter et al. [16] design an automated demand-responsive transport system suitable for areas with intensive demand, which offers a direct connection between pick-up and drop-off

locations without any detours or intermediate stops. Navidi et al. [17] demonstrate that replacing traditional public transport with demand-responsive transport can decrease passengers' perceived travel times without any extra costs, under certain circumstances. Nonetheless, these studies mainly focus on demand-responsive transit services between two locations, whereas strategies for on-demand operation between multiple stops are seldom discussed.

In this study, we envision a real-time ABoD system servicing multiple intermediate bus stops. The new system is compared with the current fixed-schedule and fixed-route bus services. We assess the impacts of the envisioned ABoD system under various scenarios via agent-based simulation, which has advantages over mathematical approaches, as it can provide insight into the operation of a system under different scenarios [18].

This paper is organized as follows. In the second section, we conduct a descriptive analysis to investigate current bus travel demand in our study area, Fuyang, Zhejiang China, based on mobile payment records. In the third section, we elaborate on the agent-based modeling framework and the behaviors of the agents and their interactions, including passengers, autonomous buses, and the control center. In the subsequent section, after the verification of the consistency of the random bus travel origin-destination (OD) matrix with real-world data to ensure the reliability of the simulation results, we simulate the scenarios with various ABoD operation strategies and compare this with the current conventional bus services to evaluate whether the ABoD is competent in improving the quality of bus services in Fuyang. In the final section, discussions and conclusions are offered.

2. A Descriptive Analysis of Bus Travel Demand in Fuyang, Zhejiang

We select Fuyang, a prefecture-level city of Zhejiang, China, for our case study. Fuyang city, classified as a small city in China, had a population of approximately 742,000 in 2018. We collect data from DTChuxing, a company that promotes data fusion innovation in the urban public transport sector. The datasets include the boarding time and bus stop of each passenger who pays via IC Card or mobile phone. The period of the data incorporated in this study spans September 2018, consisting of 20 workdays and 10 weekends. In this study, we focus on bus travel demand during workdays. It is worth mentioning that the number of passengers using either IC Card or mobile phone accounts for 50% of total bus travel demand, whereas the remainder of passengers pays with cash [19]. Thus, we empirically scale up the number of passengers uniformly to synthesize aggregate bus travel demand in Fuyang.

The average bus travel demand during workdays in September 2018 is shown in Figure 1. As outlined in Figure 1(a), the number of passengers varies with the nature of land use. For example, bus stops close to railway stations or hospitals attract more passengers. In addition, bus travel demand varies largely by bus route. As outlined in Figure 1(b), the busiest bus lines carry more than 5,000 passengers a day, whereas some other routes transport less than 1,000. These inefficient bus routes are more financially unsustainable for operators (i.e., black bar: Bus Line 622). Thus, in this work, we envision an ABoD system to replace these routes, and we assess the performance of the new system.

The descriptive analysis above demonstrates that one of the greatest challenges faced in Fuyang, as in many small Chinese cities, is the imbalance of transit supply and demand, where the bus supply normally exceeds the demand, which leads to the low utilization of buses and, possibly, great financial deficits. Meanwhile, the development of ICT and driverless technology offers new possibilities to operate ABoD services. A variety of autonomous bus services have been opened for trial operation. For instance, in Dubai, UAE, the Next Future Transportation Inc. deployed a modular autonomous bus—with 6 seats and a capacity of 10 passengers—that can be joined onto or detached from the other bus modules (see https://www.next-future-mobility.com).

In this paper, we take Bus Line 622 as an example to demonstrate the impacts of the ABoD system. Currently, the bus service operates from 6 a.m. to 8 p.m., with a fixed departure interval of 10 minutes. The vehicles with 26 seats are employed in operation with a capacity of 45 passengers. The route travels across Fuyang city center, with 14 stops, including a number of important points-of-interest, such as the First Hospital of Fuyang and Fuyang West Station. Figure 2(a) outlines the detailed routes of this bus line, and the average number of passengers boarding the buses at each stop is also visualized. As shown in the figure, the spatial distribution of bus travel demand is not uniform across the route. Some bus stops attract over 100 passengers, while others have fewer than ten passengers during workdays.

The temporal distribution of travel demand is shown in Figure 2(b). The average travel demand of Bus Line 622 during workdays is measured in 30-minute intervals. Observable morning and evening peak hours exist for this bus line, with around 200 passengers boarding during morning peak hours, whereas passengers during off-peak hours are few.

3. An Autonomous Bus-on-Demand System: Design and Operation

We construct an agent-based model to evaluate the performance of the ABoD system from three perspectives: operator, passenger, and road resources. We compare the ABoD with traditional bus systems under various travel demand scenarios. The model is coded in Java in the AnyLogic platform, with an integrated GIS environment.

The ABoD system primarily consists of three types of agents: passenger, control center, and autonomous bus. In this study, we model the behaviors and interactions among the three types of agents in the system, as illustrated in Figure 3. The control center collects travel demand information from passengers and determines departure intervals and bus operation strategies. Meanwhile, passengers can receive real-time bus information. When the departure

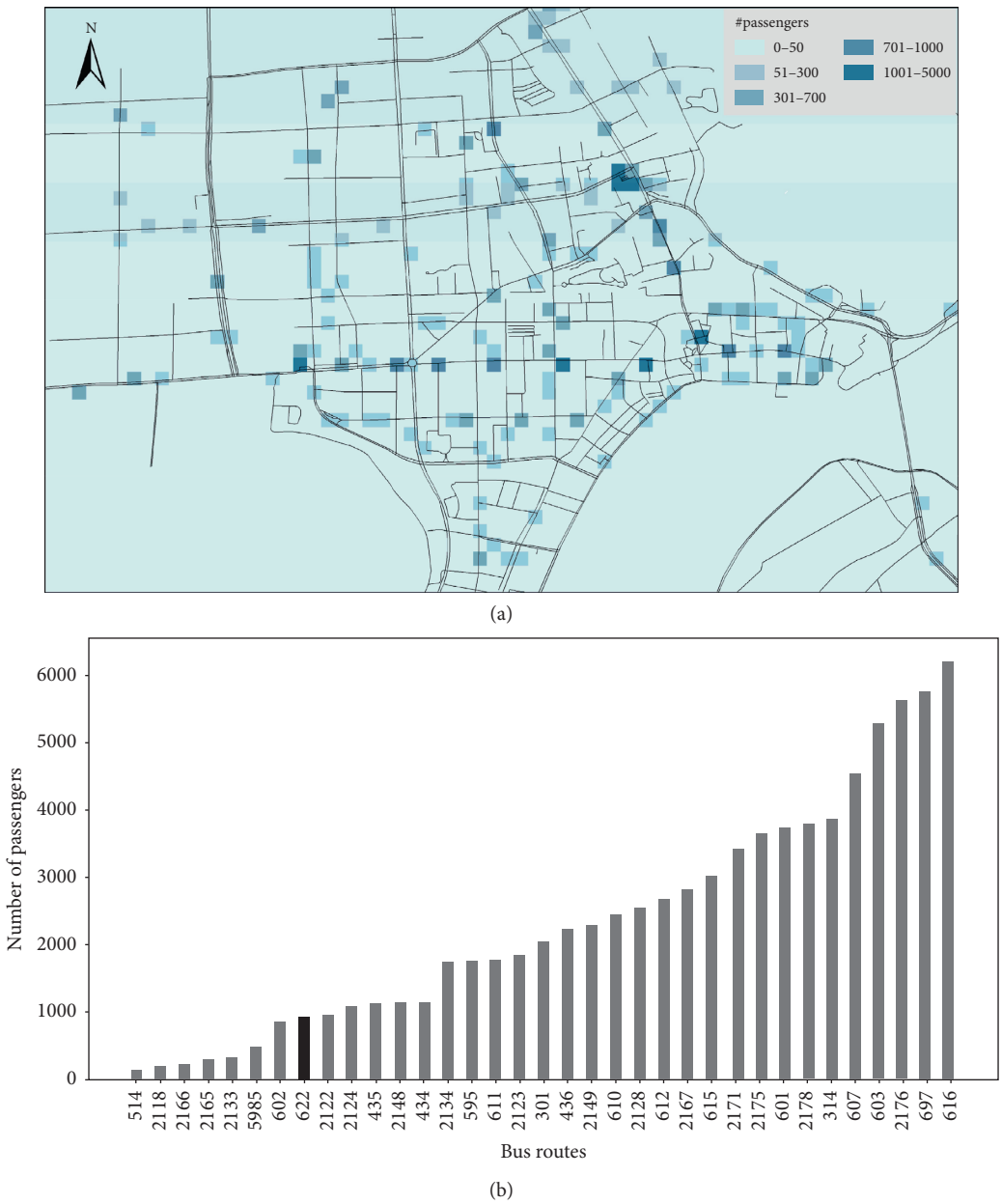


FIGURE 1: Bus travel demand in Fuyang, Zhejiang, China. (a) Spatial distribution of bus travel demand in Fuyang. (b) Bus travel demand by route.

threshold is met, the control center instructs the autonomous buses to depart and continuously receives bus travel information. If necessary, the autonomous buses skip intermediate stops, according to instructions from the control center.

Figure 4 describes the detailed designs of the ABoD system, which are elaborated in the following subsections. In the figure, rounded rectangles refer to the agents; ovals refer to locations; rectangles refer to agents' behaviors.

3.1. Passengers. Passengers send their real-time travel demand information to the control center and walk to the bus stop to wait for the bus. In the simulation, the incidence of passengers follows the real-world statistics of boarding time in each second. Due to the lack of alighting information, we firstly derivate the OD matrix of bus travel in morning and evening peak hours based on the following model, in which the initialized OD matrix can be expressed as in the following equation [20, 21]:

$$\begin{cases} \mathrm{OD}_{ij} = \dfrac{A_i}{\sum_{k=1}^{m} A_k} \times \dfrac{p(j-i)}{\sum_{w=1}^{j-1} p(w)} \times \sum_{z=1}^{m} O_z, & \text{if } i < j, \\ \mathrm{OD}_{ij} = 0, & \text{if } i \geq j, \end{cases} \tag{1}$$

where OD_{ij} is the number of passengers from bus stop i to j during morning peak hours; m is the number of bus stops

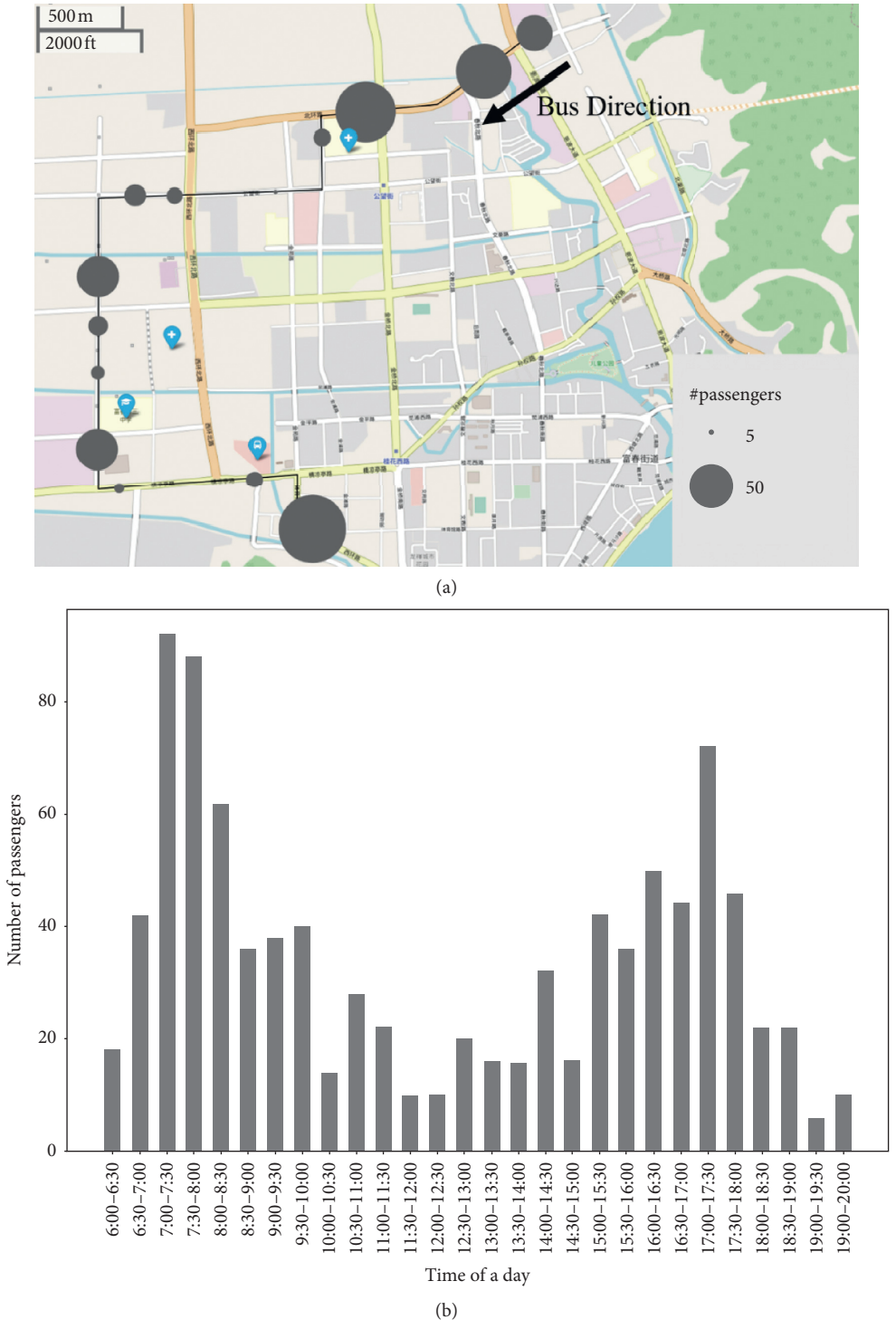


Figure 2: Demand for Bus Line 622. (a) Spatial distribution of bus travel demand. (b) Temporal distribution of bus travel demand.

along the bus route; $\sum_{z=1}^{m} O_z$ represents the total number of passengers boarding the bus at all stops during morning peak hours; A_i is the number of passengers boarding the bus at stop i in the evening peak in the opposite direction; and $p(j-i)/\sum_{w=1}^{j-1} p(w)$ represents the probability that the passengers get off at stop j coming from stop i. We also assume that the number of bus stops a passenger passes by follows the Poisson distribution [22].

The final results can be calculated according to the following equations:

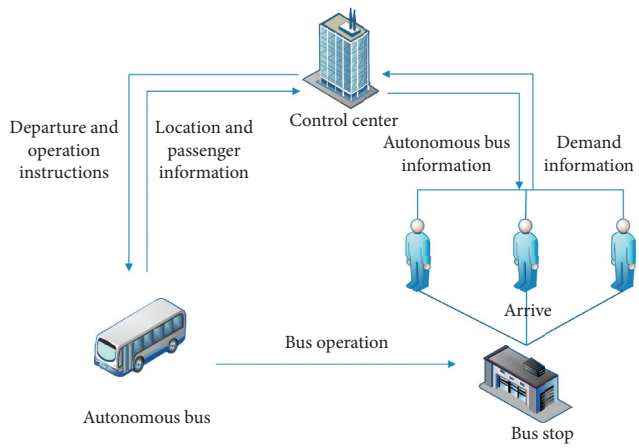


Figure 3: Autonomous bus-on-demand system.

Figure 4: Agent-based simulation of passengers, control center, and autonomous buses.

$$\mathrm{OD}_{ij}^{t+1} = \varepsilon_i \varepsilon_j \mathrm{OD}_{ij}^{t}, \tag{2}$$

$$\varepsilon_i = \frac{O_i}{O_i^*}, \tag{3}$$

$$\varepsilon_j = \frac{D_j}{D_j^*}. \tag{4}$$

In step $t+1$, OD_{ij}^{t+1} can be iterated according to (2). In (3), O_i refers to the number of passengers boarding the bus at

stop i, calculated by the OD matrix, OD_{ij}^{t}, at step t. Furthermore, O_i^* is calculated based on real-world data. The idea of (4) is similar to (3), but for the theoretical number of passengers getting off at stop j. The iteration terminates if the following condition is satisfied:

$$\frac{\max\left(\varepsilon_i\varepsilon_j\right) - \min\left(\varepsilon_i\varepsilon_j\right)}{\left(\left(\max\left(\varepsilon_i\varepsilon_j\right) + \min\left(\varepsilon_i\varepsilon_j\right)\right)/2\right)} \leq \sigma. \tag{5}$$

In (5), σ is a constant that determines the accuracy threshold, empirically set as 0.01 in this paper. Thus, the OD matrix for Bus Line 622 is obtained from 6:30 a.m. to 9:30 a.m., morning peak hours, and from 3:00 p.m. to 6:00 p.m., evening peak hours. The remaining bus travel demand is randomly allocated to off-peak hours. We assume that the passengers do not leave the queues.

3.2. Control Center. The behaviors of control center agents are illustrated in Figure 4. As the core of the ABoD system, the primary functions of the control center are information collection and the analysis of demand data, to control bus departure intervals and operation strategies. The control center determines departure intervals according to the algorithm integrated into Figure 5, based on real-time demand. The control center examines the total number of passengers that should be carried by the autonomous buses, including the number of passengers who are too late to board any bus in service and the number of passengers who are left behind during the constraints of bus capacity. The control center computes bus travel demand every minute, considering the following scenarios.

If the departure interval of the previous bus is less than the maximum departure interval, the number of autonomous buses sent out can be expressed as in the following equation:

$$n = E\left(\frac{\sum_{i=1}^{m} p_i}{c_{\max} S}\right) + E\left(\frac{\left(\sum_{i=1}^{m} p_i\right) \bmod \left(c_{\max} S\right)}{c_{\min} S}\right), \tag{6}$$

where $E(\bullet)$ means the function of taking the integer part of $(\bullet)$; $\sum_{i=1}^{m} p_i$ represents the real-time demand of m stops, which is 14 in this paper; $c_{\min}$ is the ratio of minimum departure condition; $c_{\max}$ is the ratio of maximum departure condition; and S is the number of seats in an autonomous bus, which is six in this paper.

That is, if the real-time demand ranges between $c_{\min}$ and $c_{\max}$, only one autonomous bus is needed. However, if the real-time demand is greater than $c_{\max}$, more than one autonomous bus could be required. In this case, n buses are dispatched together as a platoon. The value of n can be calculated using (6). Assuming that the maximum bus loading rate—the ratio of the maximum capacity of a bus to the number of seats—is R, the values of $c_{\min}$ and $c_{\max}$ should satisfy

$$c_{\min} < c_{\max}, \tag{7}$$

$$c_{\max} < R. \tag{8}$$

If demand reaches a very low level, where the departure interval from the previous bus is equal to the maximum allowed departure interval, a single autonomous bus is sent out regardless.

The maximum departure interval is set at 10 minutes, according to that of the current bus service. We also simulate the scenario of fixed-route, fixed-schedule bus services as the benchmark for comparative study.

3.3. Autonomous Buses. The behaviors of the autonomous buses are also illustrated in Figure 4. The autonomous bus's travel route mirrors that of Bus Line 622, calculated based on the routing service of OpenStreetMap. After receiving instructions from the control center, autonomous buses begin to service passengers from the origin stop, heading towards the terminal. To improve the efficiency of bus operations, autonomous buses may skip intermediate stop m, except for the following situations:

(i) There is at least one passenger on the bus who needs to disembark at stop m.

(ii) There is at least one passenger waiting at stop m for the autonomous bus, and the bus has enough room for the passengers to board, considering the number of passengers alighting at stop m.

The autonomous bus stops if any of the above conditions are met. The specific operation strategy is shown in Figure 5. The autonomous bus receives instructions about whether to skip stop m from the control center before approaching the stop.

The operation details and assumptions about the autonomous buses are summarized as follows. The average speed of the bus is 20 km/h, according to morning peak-hour traffic conditions in Fuyang city center. The penalty of delay due to acceleration and deceleration at each stop is empirically set at 10 seconds, which is consistent with conventional buses. The average boarding or alighting time caused by each passenger is 1 second [23]. According to the technical specifications of the Next Future Transportation Inc., in the simulation, the bus can carry 10 passengers at maximum.

4. Simulation Results and Scenario Analysis

We examine the impacts of the ABoD system from the following perspectives:

(1) Road resources: the concept of passenger car unit kilometers (PCU-km) is employed as the approximation for studying the occupation of road resources by autonomous or conventional buses [24]. The PCU-km is derived as the product of the vehicle kilometers traveled (VKT) and the passenger car equivalent (PCE) factor. According to the design code for urban road engineering in China [25], the PCE factor of a conventional bus is 2. If one autonomous bus or no more than three buses together

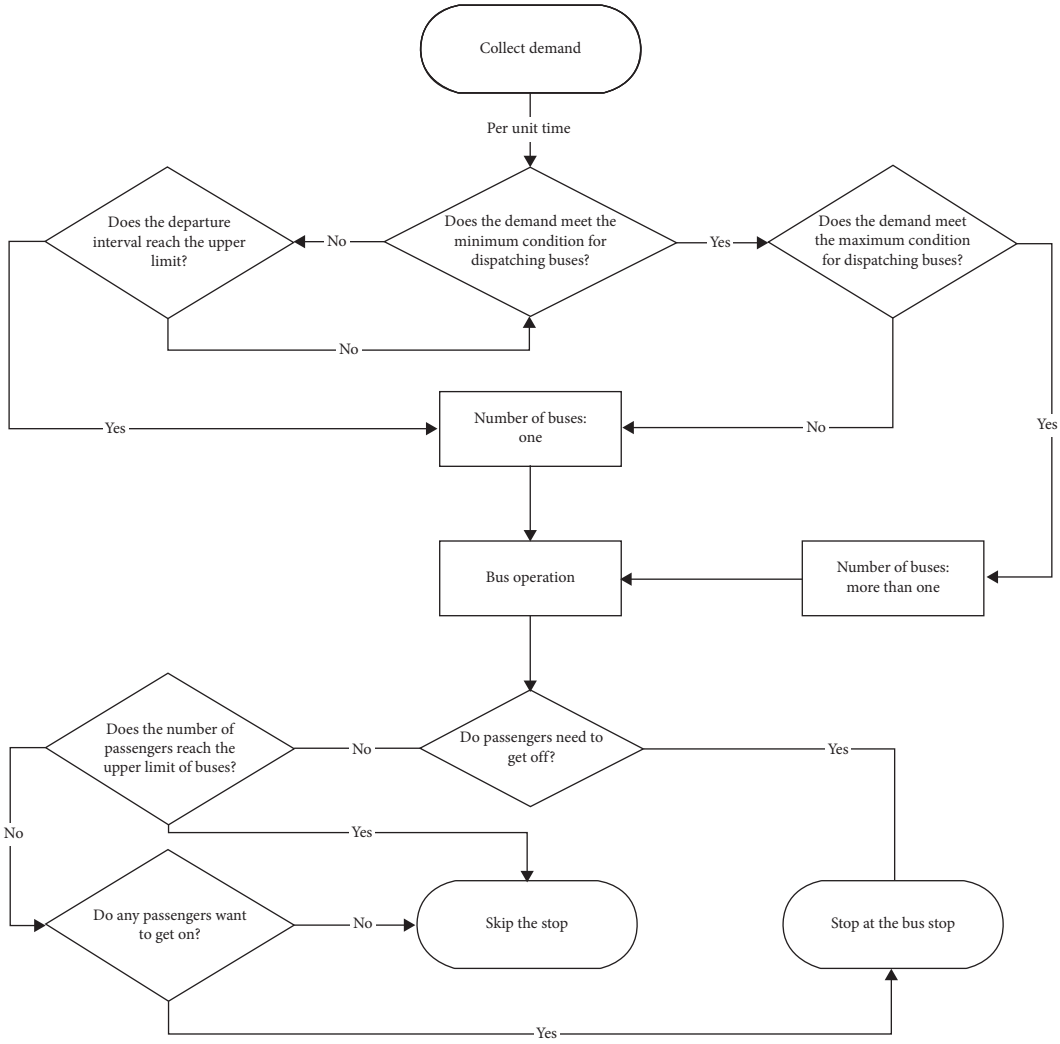


Figure 5: Algorithm of departure interval and bus operation.

as a platoon are dispatched, the PCE factor is 1. Otherwise, the PCE factor is set as 2.

(2) Passenger: the waiting time of passengers is selected as the primary factor for evaluating bus service quality.

(3) Bus operation: passenger load factor (PLF) is used to evaluate the efficiency of vehicle capacity utilization, which is the ratio of passenger-kilometers traveled to capacity-kilometers available, expressed as [26]

$$F = \frac{\sum_{i=1}^{n} p_i d_i}{\sum_{j=1}^{u} S_j^{\max} d_j}. \quad (9)$$

In the equation, $p_i d_i$ refers to the total kilometers traveled by passenger i; and $S_j^{\max} d_j$ refers to the capacity-kilometers available for j-th bus.

4.1. Scenarios. Daily ABoD operation from 6:00 a.m. to 8:00 p.m. is simulated, where 1 second in the simulation equals one in the real world. We study the simulation results by varying the ranges of the following parameters, which are decisive in controlling the ABoD system. The ratio of minimum departure condition, $c_{\min}$, is set as 0.3, 0.6, or 0.9. The ratio of maximum departure condition, $c_{\max}$, is set as 1, 1.3, or 1.6. For comparative purposes, to simulate the performance of the current bus system, the departure interval is fixed at 10 minutes, while each bus vehicle dwells at each bus stop. In addition, we also shift the intervals of the conventional bus system to 5 and 8 minutes, to obtain more comprehensive results. From the perspective of bus travel demand, we also test the impact of various demand levels, ranging from one to ten multiples of current bus demand.

In each scenario, simulation results are the average of 100 runs. Based on actual bus travel demand, we randomize

the arrival time of each passenger according to Poisson distribution. To ensure the reliability of the results, it is necessary to control the error between the randomized travel OD matrix, M^*, and the real-world travel OD matrix, M. We choose the 2-Norm of matrix to assess error, expressed as [27]

$$\text{error} = \frac{\left|\|M\|_2^* - \|M\|_2\right|}{\|M\|_2}. \tag{10}$$

In the equation, $\|M\|_2$ refers to the 2-Norm of matrix M, and $\|M\|_2^*$ denotes the 2-Norm of matrix M^*. We empirically set the upper limit of error as 0.05, and the simulation result is adopted only if the simulation error does not exceed this limit.

Simulation results are shown in Figure 6, where each dot represents the average value of 100 simulation runs of each scenario. The solid lines refer to ABoD, and the dashed-dotted lines represent conventional buses. Each picture includes the simulation results of the ABoD system with various ratios of minimum departure condition $c_{\min}$, with the values of 0.3, 0.6, and 0.9, under conditions from one to ten multiples of current bus travel demand. For comparison, the simulation results of the conventional bus system with different fixed departure intervals (i.e., 5, 8, and 10 minutes) are also plotted in the figure. In addition, the x-coordinates of all figures refer to the multiples of bus travel demand. The figures in each row share the same vertical coordinates, but the ratios of maximum departure conditions differ, at 1, 1.3, and 1.6, respectively.

The figures in the first row show the impact on total PCU-km in each scenario. In the second row, the figures present the impacts on waiting time. Those in the bottom row show the impact on PLF.

In examining the figures in the first row, one can compare the impacts of the ABoD system with differing values for $c_{\min}$ and $c_{\max}$ and PCU-km of autonomous buses increases almost linearly, thanks to their demand-responsive nature. In terms of the ratio departure condition, $c_{\min}$ and $c_{\max}$, larger ratios lead to greater PCU-km, particularly if the multiple of current demand is greater than 1. If bus travel demand is less than five multiples of the current demand, the total PCU-km of autonomous buses is still less than that of conventional bus services. These results show that, under current circumstances, the ABoD system occupies much fewer road resources than the current buses. Even if current bus travel demand were to grow, the performance of the on-demand-bus service would still surpass the current ones.

The figures in the second row indicate the average waiting time of each passenger under different scenarios. With greater bus travel demand, the average waiting time in the ABoD system decreases and tends to remain at the same level. However, in the scenario of the traditional bus system with 10-minute intervals, if demand exceeds five multiples of the current demand, the average waiting time increases sharply, due to the constraints of bus capacity, and passengers are left behind. Compared with current bus services, passenger waiting time in the ABoD system is much lower, regardless of demand.

In the figures in the bottom row, we plot the PLF under different scenarios. Greater bus travel demand leads to an increased PLF, for both ABoD system and conventional buses. If the multiple of demand is less than or equal to ten, the PLF in ABoD is greater than that in the conventional bus system, indicating that it is reasonable to substitute ABoD for the current bus services.

In summary, with less than six multiples of current bus travel demand, the ABoD system presents as overwhelmingly advantageous, with less road resource occupation, shorter passenger waiting times, and more efficient utilization of vehicle capacity. In addition, the ratio of minimum departure condition, $c_{\min}$, and of maximum departure condition, $c_{\max}$, has little impact on the ABoD system with low demand. When the departure intervals of current conventional bus service decline to 5 or 8 minutes, despite shorter passenger waiting times, the total PCU-km of the conventional bus system becomes too large, meaning that, in this case, the fixed-schedule bus system occupies too many road resources compared with the ABoD system proposed in this study.

4.2. Number of Autonomous Buses. The average number of departed autonomous buses is counted every 30 minutes during a day with two multiplies of current demand. In Figure 7, the blue solid lines refer to the number of boarding passengers (i.e., demand), and the dashed-dotted lines represent the number of departed autonomous buses. In each figure, the ratio of minimum departure condition, $c_{\min}$, varies with the values of 0.3, 0.6, and 0.9, and the ratio of maximum departure condition, $c_{\max}$, is set as 1, 1.3, and 1.6. The ABoD system can adapt to changes in demand by adjusting bus dispatching strategies. In terms of the ratio of minimum departure condition, $c_{\min}$, larger ratios lead to smaller numbers of autonomous buses under the same conditions. However, the ratio of maximum departure condition has little impact on the control strategy of buses, due to insufficient demand.

4.3. Cost Analysis. The operating costs for conventional buses and autonomous buses are estimated as follows. For autonomous buses, the operating cost mainly includes energy cost, maintenance, vehicle repair, administrative cost, tax, and dispatching [28]. The operating cost follows a linear function of bus capacity. The operating cost for m bus units can be formed as [29, 30]

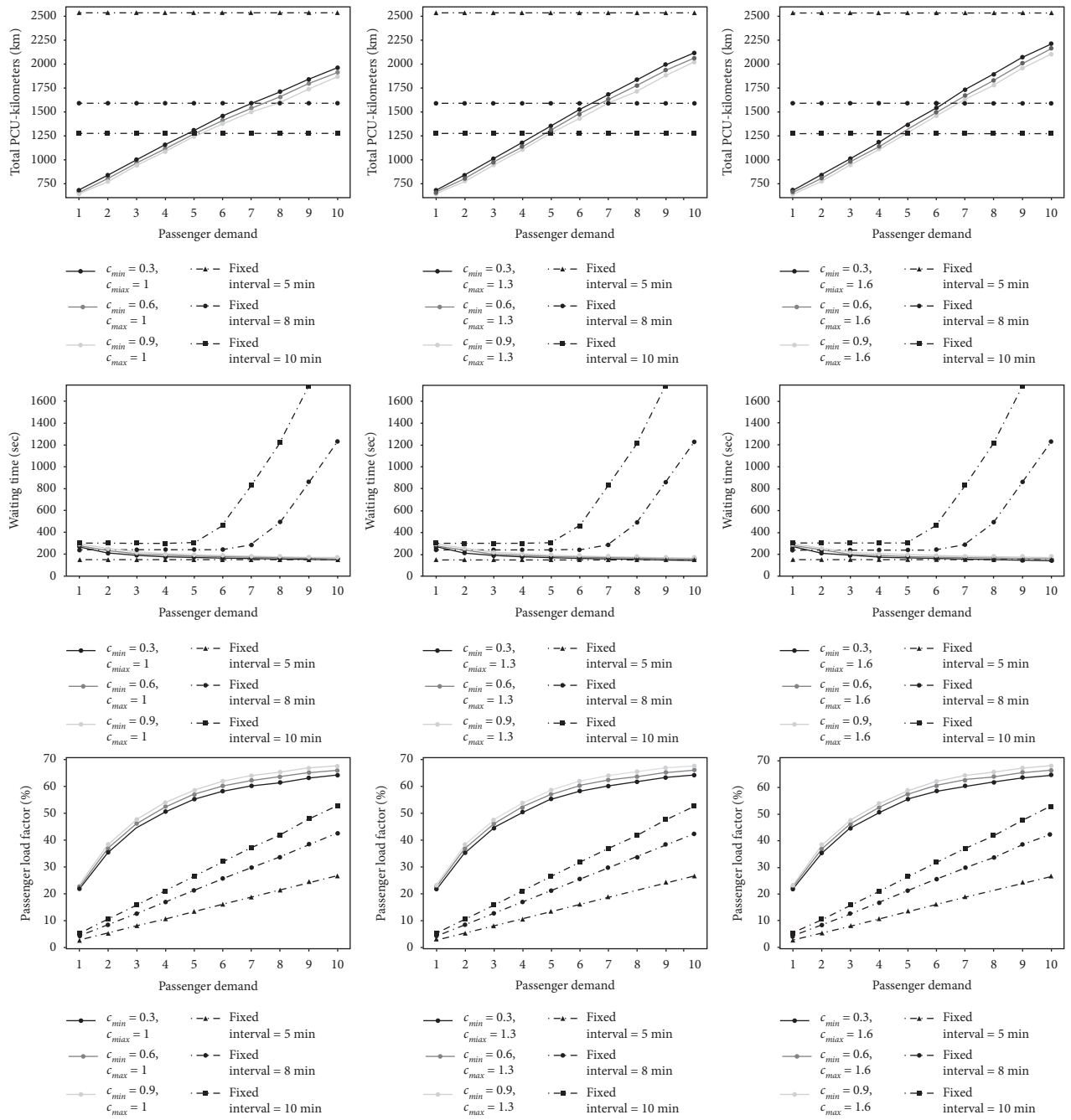


Figure 6: Simulation results.

$$f_a = C_a^F + C_a^A + C_a^V \left(mS_a^{\max}\right)^{\alpha}, \tag{11}$$

where C_a^F means the fixed operating cost and C_a^A means dispatching cost, including the assembling and dissembling cost. C_a^V refers to the marginal operating cost, and $S_a^{\max}$ is the capacity of an autonomous bus, which is 10 in this case. Parameter α ranges between 0 and 1 according to the cost to operate the autonomous buses. We assume that the saving of cost is greater with a larger bus fleet size as the buses are running with greater fuel economy when they are joined together.

For conventional buses, the labor cost is one of the major components of the total cost. Taking into account the labor cost, according to Dai et al. [29], the linear relationship between the operating cost and bus capacity still holds, expressed as

$$f_0 = C_0^F + C_0^V S_0^{\max}, \tag{12}$$

where $S_0^{\max}$ stands for the capacity of a conventional bus, which is 45 in our study.

Based on the work of Dai et al. [29], for an autonomous bus, the fixed operating cost is \$14.49 per dispatch, the

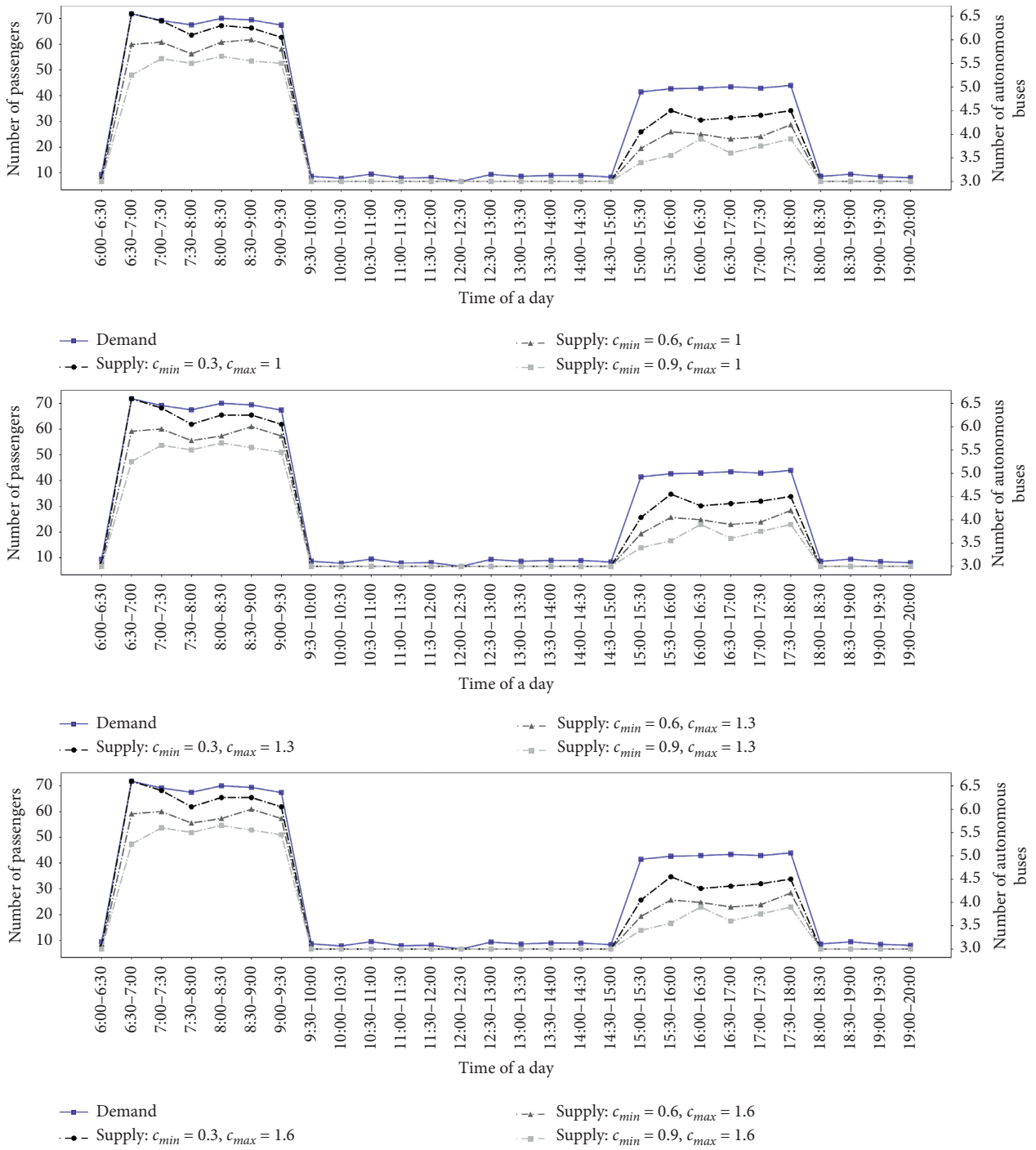


Figure 7: The variety of numbers of autonomous buses in a day.

marginal operating cost is \$0.45 per seat per dispatch, and the assembling and/or dissembling cost is \$1.07. For a conventional human-driven bus, the fixed operating cost is \$39.18 per dispatch; the marginal operating cost is \$0.45 per seat per dispatch. Parameter α is set as 0.9.

In addition to the operating cost, the generalized cost for passengers, including waiting time as bus stops and in-vehicle travel time, is also taken into account [29]. The value of waiting time for each passenger is set as \$0.8 per minute according to the wage level of Fuyang, Zhejiang. Since the in-vehicle travel time is less sensitive, we set the value of in-vehicle travel time as \$0.2 per minute.

The comparison of cost between ABoD and conventional buses is illustrated in Figure 8. The total cost includes

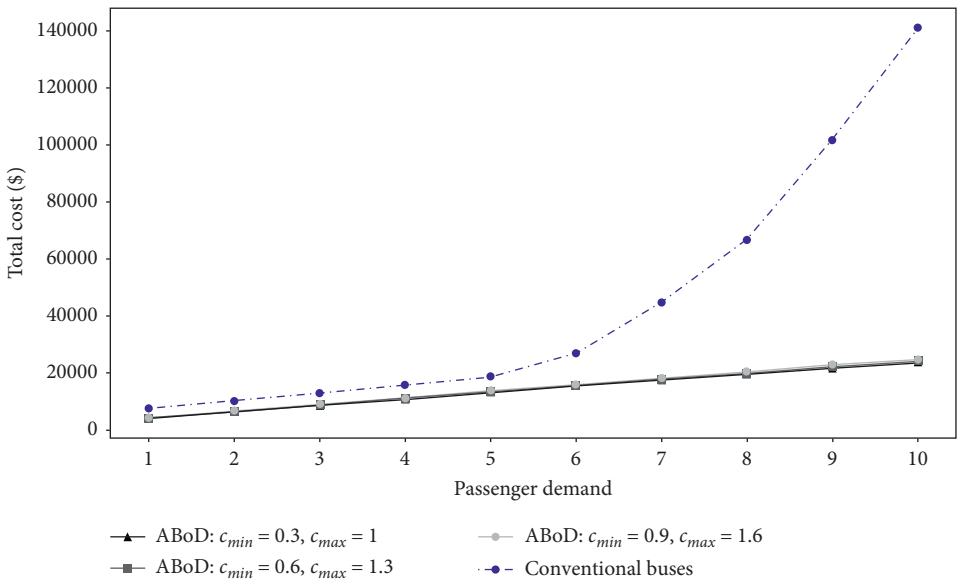


Figure 8: The total cost comparison between ABoD and conventional buses.

operating costs, passenger waiting time cost, and passenger travel time cost. In the figure, the ratios of departure condition of the ABoD system are set to combinations of 0.3 and 1, 0.6 and 1.3, and 0.9 and 1.6, respectively. The simulations show that no matter the ratio of departure condition is, the total costs vary only slightly. The total cost of conventional buses is always greater than that of the ABoD. If the travel demand exceeds five multiples of the current demand, the total cost of conventional buses increases sharply, while the cost of the ABoD increases almost linearly, which indicates that the ABoD system is more adaptive to the change in travel demand than the conventional buses.

4.4. Comparison with an Optimized Bus Dispatching Model. In the section, we focus on the difference between ABoD and an optimized bus dispatching model. The optimal model is calculated based on Newell [31], which is elaborated in Appendix A. The total costs of the ABoD and optimization model under various passenger demands are shown in Figure 9. The excess cost ratio—defined as the ratio of excess cost to total cost of optimization model—is also plotted, represented by the red curve. The results show the total cost of ABoD compared to that of the optimized model. However, if current bus travel demand grows, the excess cost ratio decreases, meaning that the relative cost advantage of the optimization model over ABoD is shrinking.

The simulation results of the total PCU-Kilometers, waiting time, and passenger loads in the ABoD and the optimized model are shown in Figure 10, in which the solid curves refer to the ABoD, and the blue dashed-dotted curves refer to the optimized model. Under the current level of travel demand, the passengers' waiting time of the ABoD is less than that of the optimized model thanks to the mechanism of maximum departure interval in ABoD. If travel demand increases, the passengers may wait for a longer time in the ABoD system. The longer waiting time is due to the fact that the ABoD system balances the waiting time against the PCU-Kilometers and loads of vehicles. As shown in the figures, with larger bus travel demand, the PCU-kilometers in the ABoD are much smaller while the loads are greater than in the optimized model. It shows that the ABoD system is potentially more beneficial to sustainable development for bus operators despite a relatively long waiting time.

In real life, bus travel demand can be affected by emergent incidents like large gathering events or the breakdown of the metro, which causes the outbreak of passenger flows. In this case, the adaptability of the bus system to the fluctuation of demand becomes essential. In this work, we simulate an emergent event (e.g., the metro breakdown) that results in the λ-multiples of current passenger demand from 7:00 a.m. to 9:00 a.m. The total costs of the ABoD and the optimized model with various demand multiples of λ are plotted in Figure 11. The blue dashed-dotted curve for the total cost of the optimization model grows faster than solid curves for ABoD. If λ becomes greater than 3, the ABoD is more dominant than the optimization model in total cost since the ABoD is more responsive to real-time passengers' demand.

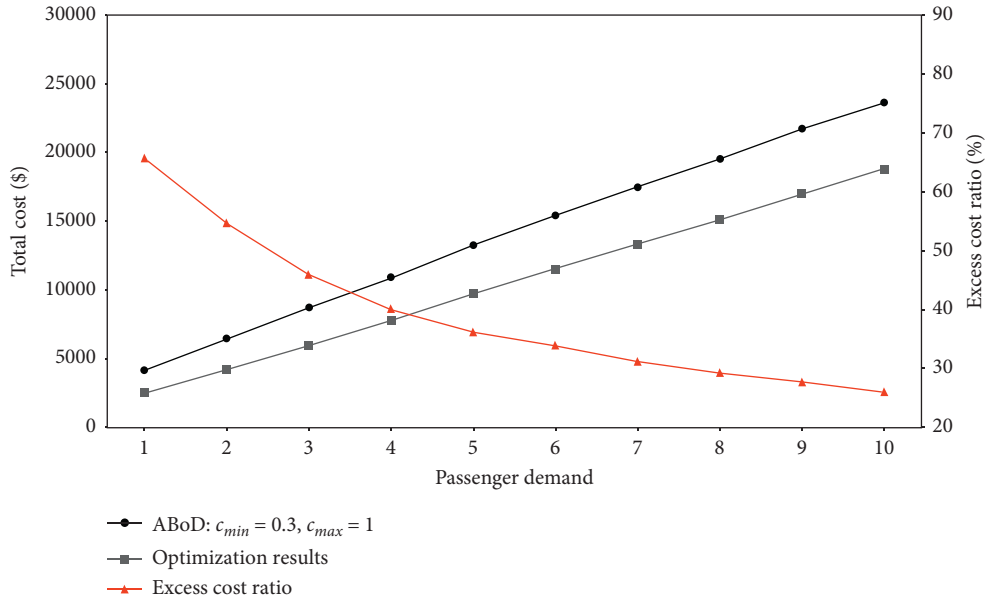


Figure 9: The total cost comparison between ABoD and optimization model.

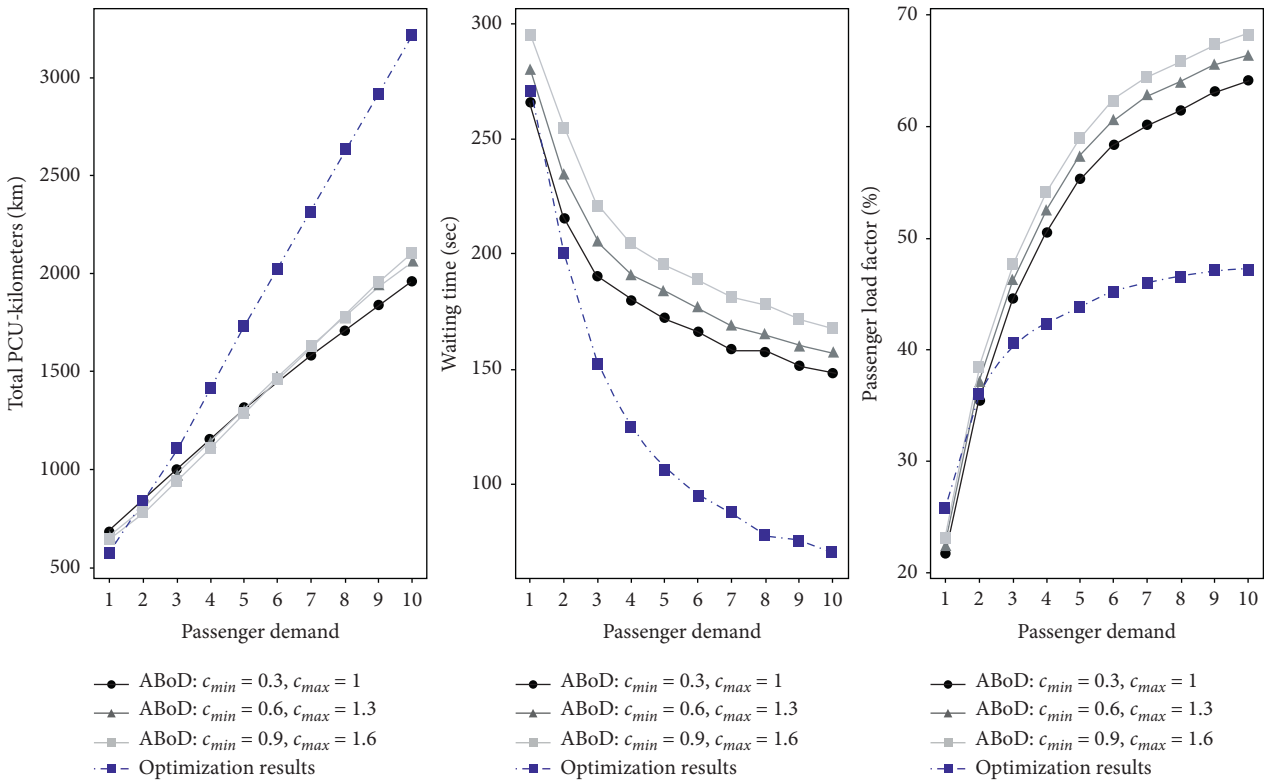


Figure 10: Simulation results of ABoD and optimization model.

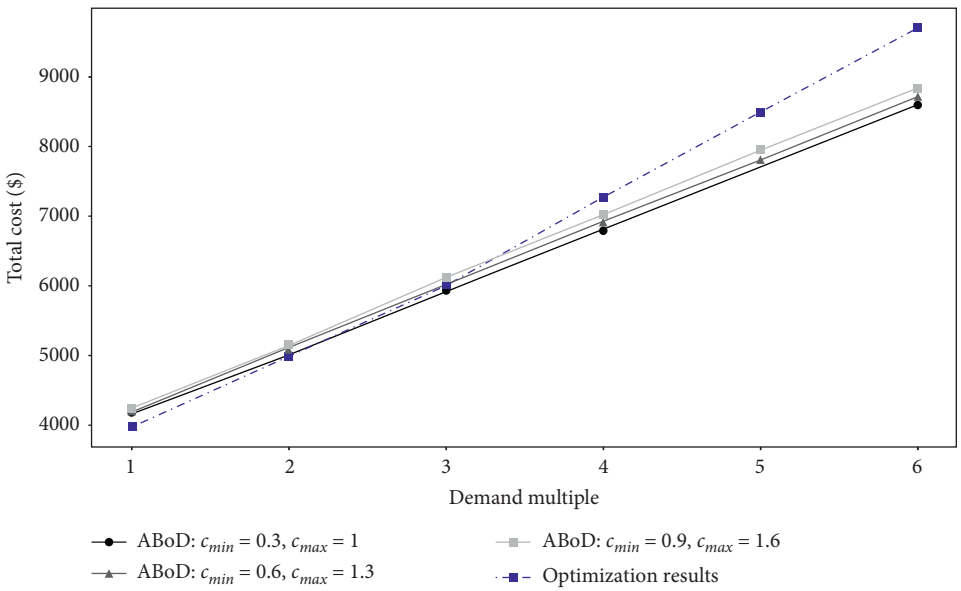


Figure 11: The impact of sudden passenger flow on ABoD and optimization model.

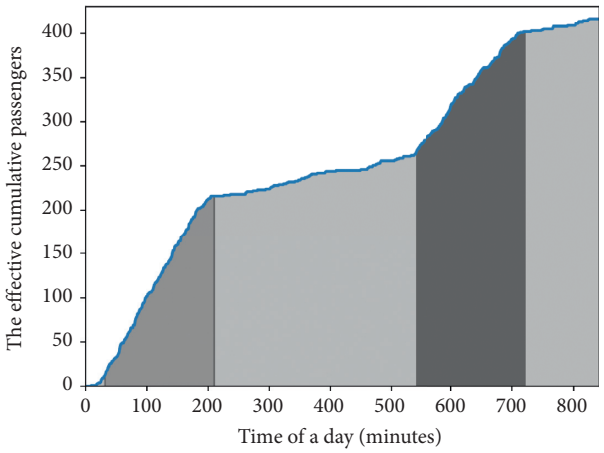


Figure 12: The effective cumulative passenger demand of Bus Line 622.

5. Conclusion and Discussion

This paper has presented a spatiotemporal analysis of current bus travel demand in Fuyang, a prefecture-level city in Zhejiang, China. Using Bus Line 622 as an example, we have envisioned an ABoD system with various bus dispatching and operation strategies. To evaluate the performance of the ABoD system, an agent-based model has been built that includes three primary types of agents: passengers, control center, and autonomous buses. To model bus travel demand, an OD matrix has been deduced, based on real-world payment data. We have investigated the impacts of the ABoD system from the following perspectives: the occupation of road resources, passenger waiting time, and utilization of bus vehicles. Under a series of scenarios, with various bus dispatching rules and travel demands, we have compared the impacts of the ABoD system with those of both conventional fixed-schedule bus services and the optimized bus dispatching strategies. The simulation results show that the ABoD system performs better at levels below triple current demand. Using total PCU-km as an approximation of road resource occupation, the simulation has illustrated that the ABoD system is favorable in terms of saving road resources. The PLF in the ABoD system is greater, showing that vehicle capacity is utilized more efficiently. In addition, the ABoD system is adaptive to the sudden surge in bus travel demand and is economically sustainable.

Room for improvement exists that must be acknowledged here. Control strategies of departure intervals and bus operation in the current models are rule-based, which can be further improved with additional optimization algorithms. Other travel modes that may compete with buses can be incorporated into the model, and dynamic interactions between the supply and demand of autonomous buses should be considered as well.

In this paper, we use one bus line as an example to demonstrate the feasibility of implementing the envisioned ABoD system in small cities. For future work, we could examine the impacts of the entire bus network, considering supply-demand interactions. In addition, the constraints of bus fleet size may also lead to the empty travel of autonomous buses. The assessment of operational costs incurred by

the empty travel of buses and the repositioning of these empty buses could be another important direction for future studies. Finally, with additional types of agents, including other potential stakeholders involved in this system, the agent-based model can be extended to describe more complex realities with additional dimensions, such as modal competition and pricing strategies.

Appendix

A. An Optimal Model for Bus Dispatching

To compare with the proposed ABoD model, we formulate the problem using the optimization methods based on Newell's work [31], which minimizes the operating costs and the passengers' delay. Along a bus line, $F_{ij}(t)$ stands for cumulative passenger demand at time t from point i to point j, i.e., the number of passengers boarding at point i at time t and alighting at point j. Thus, any cumulative passenger demand $F_{ij}(t)$ can be converted into an effective demand at time $t - \tau_i$ from the origin to point j, where τ_i stands for travel time from the origin to point j. $F(t)$ stands for the effective cumulative passenger demand from the origin at time t, which is expressed as

$$F(t) = \sum_{i,j} F_{ij}(t + \tau_i). \tag{A.1}$$

Between time t_o and time T, the sum of passenger waiting time can be approximately expressed as

$$W = \frac{1}{2n}\left[\int_{t_o}^{T} \sqrt{f(t)}\,\mathrm{d}(t)\right]^2, \tag{A.2}$$

where n is the number of dispatches; $f(t)$ is the derivative of $F(t)$. Regardless of the capacity of the vehicle, the optimal headways Δt with minimized total cost are expressed as

$$\Delta t = \sqrt{\frac{2a}{bf(t)}}. \tag{A.3}$$

where a is the operating cost per dispatch and b is the value of passenger waiting time. Taking into account the constraint of vehicle capacity $S^{\max}$, the optimal headways Δt are expressed as

$$\Delta t = \min\left\{\frac{S^{\max}}{f(t)}, \sqrt{\frac{2a}{bf(t)}}\right\}. \tag{A.4}$$

In this case study, we adopt a staged optimization approach since the bus travel demand varies largely during peak and off-peak hours. As illustrated in Figure 12, the effective cumulative travel demand in one day can be divided into five stages. In each stage, the effective cumulative passenger demand $F(t)$ is approximated as a linear function of time t.

Data Availability

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

Conflicts of Interest

The authors declare that no conflicts of interest regarding the publication of this paper exist.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (71901164), the Natural Science Foundation of Shanghai (19ZR1460700), Shanghai Science and Technology Committee (19DZ1208700), and the Fundamental Research Funds for the Central Universities (22120180569). The corresponding author is sponsored by the Shanghai Pujiang Program (2019PJC107).

References

[1] C. F. Daganzo, "An approximate analytic model of many-to-many demand responsive transportation systems," *Transportation Research*, vol. 12, no. 5, pp. 325–333, 1978.

[2] N. H. M. Wilson and C. Hendrickson, "Performance models of flexibly routed transportation services," *Transportation Research Part B: Methodological*, vol. 14, no. 1-2, pp. 67–78, 1980.

[3] A. Ceder and N. H. M. Wilson, "Bus network design," *Transportation Research Part B: Methodological*, vol. 20, no. 4, pp. 331–344, 1986.

[4] R. B. Dial, "Autonomous dial-a-ride transit introductory overview," *Transportation Research Part C: Emerging Technologies*, vol. 3, no. 5, pp. 261–275, 1995.

[5] O. Adebisi and V. F. Hurdle, "Comparing fixed-route and flexible-route strategies for intraurban bus transit," *Transportation Research Record*, vol. 854, pp. 37–43, 1982.

[6] S. K. Chang and P. M. Schonfeld, "Optimization models for comparing conventional and subscription bus feeder services," *Transportation Science*, vol. 25, no. 4, pp. 281–298, 1991.

[7] L. Quadrifoglio, M. M. Dessouky, and F. Ordonez, "A simulation study of demand responsive transit system design," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 4, pp. 718–737, 2008.

[8] R. Bishop, "Intelligent vehicle applications worldwide," *IEEE Intelligent Systems*, vol. 15, no. 1, pp. 78–81, 2000.

[9] C. Csiszár and A. Zarkeshev, "Demand-capacity coordination method in autonomous public transportation," *Transportation Research Procedia*, vol. 27, pp. 784–790, 2017.

[10] B. Wang, S. A. Ordonez Medina, and P. Fourie, "Simulation of autonomous transit on demand for fleet size and deployment strategy optimization," *Procedia Computer Science*, vol. 130, pp. 797–802, 2018.

[11] W. Payre, J. Cestac, and P. Delhomme, "Intention to use a fully automated car: attitudes and a *priori* acceptability," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, no. Part B, pp. 252–263, 2014.

[12] J. Shin, C. R. Bhat, D. You, V. M. Garikapati, and R. M. Pendyala, "Consumer preferences and willingness to pay for advanced vehicle technology options and fuel types," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 511–524, 2015.

[13] M. Kyriakidis, R. Happee, and J. C. F. de Winter, "Public opinion on automated driving: results of an international questionnaire among 5000 respondents," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 32, pp. 127–140, 2015.

[14] A. Vakayil, W. Gruel, and S. Samaranayake, "Integrating shared-vehicle mobility-on-demand systems with public transit," in *Proceedings of the Transportation Research Board 96th Annual Meeting*, Washington, DC, USA, January 2017.

[15] Y. Shen, H. Zhang, and J. Zhao, "Integrating shared autonomous vehicle in public transportation system: a supply-side simulation of the first-mile service in Singapore," *Transportation Research Part A: Policy and Practice*, vol. 113, pp. 125–136, 2018.

[16] K. Winter, O. Cats, G. H. d. A. Correia, and B. van Arem, "Designing an automated demand-responsive transport system: fleet size and performance analysis for a campus-train station service," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2542, no. 1, pp. 75–83, 2016.

[17] Z. Navidi, N. Ronald, and S. Winter, "Comparison between ad-hoc demand responsive and conventional transit: a simulation study," *Public Transport*, vol. 10, no. 1, pp. 147–167, 2018.

[18] N. Ronald, R. Thompson, and S. Winter, "Simulating demand-responsive transportation: a review of agent-based approaches," *Transport Reviews*, vol. 35, no. 4, pp. 404–421, 2015.

[19] F. Hu, "Full coverage of mobile payment for all buses in Hangzhou," 2017, http://zzhz.zjol.com.cn/system/2017/05/23/021520284.s.html.

[20] X. Zhou, X. Yang, and X. Wu, "Origin-destination matrix estimation method of public transportation flow based on data from bus integrated-circuit cards," *Journal of TongJi University (Nature Science)*, vol. 40, pp. 1027–1030, 2012.

[21] J. Doblas and F. G. Benitez, "An approach to estimating and updating origin—destination matrices based upon traffic counts preserving the prior structure of a survey matrix," *Transportation Research Part B: Methodological*, vol. 39, no. 7, pp. 565–591, 2005.

[22] H. L. Dou, H. D. Liu, and X. G. Yang, "OD matrix estimation method of public transportation flow based on passenger boarding and alighting," *Computer and Communications*, vol. 25, no. 135, p. 79, 2007.

[23] X. Chen, B. Hellinga, C. Chang, and L. Fu, "Optimization of headways with stop-skipping control: a case study of bus rapid transit system," *Journal of Advanced Transportation*, vol. 49, no. 3, p. 17, 2015.

[24] Bureau of Infrastructure Transport and Regional Economics (BITRE), *Traffic and Congestion Cost Trends for Australian Capital Cities*, BITRE, Canberra, Australia, 2015.

[25] Ministry of Housing and Urban-Rural Construction of the People's Republic of China, *Code for Design of Urban Road Engineering*, China Construction Industry Publishing House, Beijing, China, 2012.

[26] "Load factor," http://web.mit.edu/airlinedata/www/Res_Glossary.html.

[27] G. V. L. Golub and F. Charles, *Matrix Computations*, Johns Hopkins University Press, Baltimore, MA, USA, 3rd edition, 1996.

[28] W. Zhang, E. Jenelius, and H. Badia, "Efficiency of semi-autonomous and fully autonomous bus services in trunk-and-branches networks," *Journal of Advanced Transportation*, vol. 2019, Article ID 7648735, 17 pages, 2019.

[29] Z. Dai, X. C. Liu, X. Chen, and X. Ma, "Joint optimization of scheduling and capacity for mixed traffic with autonomous and human-driven buses: a dynamic programming approach," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 598–619, 2020.

[30] Z. Chen, X. Li, and X. Zhou, "Operational design for shuttle systems with modular vehicles under oversaturated traffic: discrete modeling method," *Transportation Research Part B: Methodological*, vol. 122, pp. 1–19, 2019.

[31] G. F. Newell, "Dispatching policies for a transportation route," *Transportation Science*, vol. 5, no. 1, pp. 91–105, 1971.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5830593, 19 pages
https://doi.org/10.1155/2020/5830593

Research Article

Optimal Operation Scheme with Short-Turn, Express, and Local Services in an Urban Rail Transit Line

Tao Feng,[1,2] **Siyu Tao**,[1,2] **and Zhengyang Li**[3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China*
[3]*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*

Correspondence should be addressed to Siyu Tao; taosiyu@swjtu.edu.cn

Received 6 February 2020; Revised 1 October 2020; Accepted 19 October 2020; Published 19 November 2020

Academic Editor: Yu Jiang

Flexible railway operation modes combining different operation strategies, such as short-turn, express, and local services, can significantly reduce operator and user costs and increase the efficiency and attractiveness of rail transit services. It is therefore necessary to develop optimization models to find optimal combinations of operation strategies for urban rail transit lines. In this paper, a model is proposed for solving the urban rail transit operation scheme problem. The model considers short-turn, express, and local services with the aim of minimizing the operator's and users' costs. The problem is first decomposed into two subproblems: the service route design problem and the passenger assignment problem. Then, a mixed-integer nonlinear program (MINLP) model is formulated, and linearization techniques are utilized to transform the MINLP model into a mixed-integer linear programming (MILP) model that can be easily solved by commercial optimization solvers. To accelerate the solution process, a heuristic search algorithm is proposed to obtain (nearly) optimal solutions based on the characteristics of the model. The two subproblems are solved iteratively to improve the quality of solutions. A real-life case study in Chengdu, China, is performed to demonstrate the effectiveness and efficiency of the proposed model and algorithm.

1. Introduction

Urban rail transit systems play an important role in meeting passenger demands in our daily lives because of their large passenger transport capacity, low pollutant emissions, and convenience. The increase in the number of long-distance passengers and imbalanced passenger demand due to urban expansion have brought increasing pressure to public transport. The elastic passenger demand cannot be satisfied by adopting only a single operation mode. As a means of improving the quality of public transport, urban rail transit has drawn much attention from operators and governments. Operators aim to improve the service efficiency and quality by adopting flexible service strategies such as short-turn and elastic services and elastic stopping patterns to satisfy the daily travel demand. A local service serves all the stations along the full length of the line, an express service runs from a starting station to an ending station and visits only some intermediate stations, and a short-turn service is provided only on sections with larger passenger flow volumes. Flexible operation schemes integrating different service strategies can markedly improve the service level of urban rail transit. Full-length local services are commonly adopted in urban rail transit lines in China, but may lead to wasted train capacity due to the imbalanced distribution of passenger demand. In many big cities, such as Tokyo, New York, and Paris, routes with local, express, and short-turn services are used to cater to elastic passenger demand [1]. When short-turn, express, and local services are provided, different stations and sections are served at different frequencies so that the transport capacity matches the diversity of passenger demand on the line appropriately [2]. Hence, academic research on the design and financing of (nearly) optimal operation schemes is necessary for urban rail transit operators.

This study investigates the urban rail transit service design problem (URTSDP). The service routes, including which stations the trains turn back at, the stopping patterns, and the train frequencies need to be optimized to reduce the operator's operation cost and the users' time cost. Meanwhile, the capacity of each turn-back station is fixed and cannot be exceeded. After the starting and ending stations have been determined, a train departs from the starting station, serves some or all the intermediate stations, and runs in sections in a predetermined order. Express services can operate on any service route. Hence, the design of urban rail transit services is more complex and challenging compared to the design of bus service networks. The URTSDP can be divided into two subproblems: the service route design problem and the passenger assignment problem. The former determines the service routes and their frequencies, while the latter determines the optimal strategy for assigning users from their origins to their destinations [3].

Under these conditions, we propose an optimized model to solve the urban rail transit service design problem. The solution can be split into two aspects. Firstly, this model considers short-turn, express, and local services and the capacity limitations of the turn-back stations and sections in urban rail transit lines. Secondly, the passenger assignment problem is solved while meeting the capacity limitation constraints. A small-scale problem can be solved efficiently using the model and commercial solvers; however, the problem scale will increase dramatically because of the stopping patterns of the express and short-turn services if the problem is solved directly. To reduce the scale of the problem, a novel strategy called the *line pool* is introduced in the model formulation. Some important stations are designated in advance, and a service route must visit these stations if the route covers them. All the feasible service routes constitute the *line pool*. A mixed-integer nonlinear programming (MINLP) model is formulated. The model can be transformed into a mixed-integer linear programming (MILP) model by using linearization techniques such as the reformulation linearization technique (RLT). In addition, we design a search algorithm to solve the large-scale problem iteratively. Two groups of numerical studies are performed to verify the effectiveness of the model and the search algorithm.

The main contributions of this paper are as follows:

(1) Formulation of a MILP optimization model to solve the URTSDP, which includes the service route design problem and the passenger assignment problem and considers the passenger transfers and the turn-back station capacity restriction

(2) Proposal of a *line pool* strategy and novel search algorithm to solve the large-scale problem

(3) Verification of the performance of the new search algorithm by comparing its performance with that of the MILP model

(4) Demonstration of the effectiveness and efficiency of the model and the new search algorithm using real-life case studies

The remainder of this paper is organized as follows. In Section 2, an overview of the existing literature is presented. Section 3 provides the problem description. Section 4 proposes a model for the URTSDP. Section 5 describes a search algorithm to solve the problem. In Section 6, a practical numerical study is used to verify the effectiveness of the model. Section 8 concludes the paper.

2. Literature Review

Most of the studies on the design of urban rail transit services have focused on the use of express and local services. Service optimization with the synchronous operation of short-turn, express, and local services is much less well studied. It is crucial to determine which stations to stop at or skip in express/local services. Chun et al. [4] proposed a synchronized express and local system implemented to optimize the urban transit system by reducing passengers' waiting times and transfers. Gao et al. [1] analyzed the relationship between the energy consumption and passenger waiting time in the operation of express and local services and formulated a biobjective model. Jamili and Pourseyed Aghaee [5] proposed a model for designing the stopping patterns of service routes under uncertain conditions to help reduce the train fleet size and energy consumption. Based on the assumption that the capacity of the train is unlimited, Abdelhafiez et al. [6] constructed a nonlinear integer programming model, designed a heuristic method to solve the model, and compared the heuristic method to the enumeration algorithm. Luo et al. [7] presented a mathematical model that minimizes the passenger travel time and optimizes the train stopping schedule for express and local services. The model analyzes passenger choice behavior using a logit model and evaluates the transport capacity and energy consumption efficiency. However, this work does not consider the operator's cost. Yang et al. [8] proposed a nonlinear programming model which optimizes the frequency, stopping pattern, and express/local trains on the cross-lines and verified the resulting operation scheme obtained by the genetic algorithm with the actual cases. The model can reduce the users' time cost and the operational investment. It is not the only method to address the imbalanced passenger demand by optimizing the express/local train services in urban rail transit lines. Some researchers have considered other factors wherein the positions of the turn-back stations, that is, the choices of the turn-back stations for the origin and destination, play a significant role in determining the service routes. Considering the capacities of the trains and turn-back stations and other operational constraints, Sun et al. [9] proposed a model to solve the complicated problem of optimizing the operation scheme to minimize the total cost to operators and users based on the actual problems faced in urban rail transit lines. In contrast, Ding et al. [10] constructed a mathematical model to minimize the average waiting time of passengers without considering the turn-back equipment. The resulting solution illustrates that the operation scheme based on this model can effectively alleviate congestion and avoid wasted transport capacity. Zhang et al. [11] presented an MINLP model for train scheduling in which short-turn services and the influence of other impact factors on the scheme formulation

are considered. Li et al. [12] proposed a collaborative optimization model in which all the service routes can use different train compositions, and the frequency and train composition of each service route and the positions of turn-back stations are optimized simultaneously. Cao et al. [13] formulated a model based on marshaling, the skip-stop pattern, and robustness to optimize the timetabling design to minimize the total cost to the operator and users and applied the model to real-life urban rail transit in Beijing.

The aforementioned studies only consider a single operation strategy. Multiple operation strategies can be integrated into a single model to obtain better schemes to relieve the transport pressure on urban rail transit lines. There are a number of related studies on bus network design which are worth learning from. The studies are plentiful, and they mainly focus on factors such as the stop patterns, express and short-turn services, multiple fleets, and bus capacities. Given the service design similarities between urban rail transit and bus corridors, theoretical methods for bus service design can be used as references for optimizing the stopping patterns of express services. Afanasiev and Liberman [14] defined an express service as a service with no intermediate stops between the origin and terminal stations to satisfy passenger demands. The passengers can achieve travel time savings of up to 30%. Vijayaraghavan and Anantharamaiah [15] presented a theoretical model to insert express as well as partial services for better fleet assignment. The model may allow fleet reduction and provide benefits for the operator and users in terms of saved travel time. Conlon et al. [16] described the introduction of express bus services to Chicago, which were preferred by the customers and achieved a customer satisfaction that exceeded the standard target mean score. Wang and Lo [17] proposed a ferry service design model with normal and express service. The problem was formulated as an MILP model, and a solution algorithm was designed to reduce the scale of the problem. Larrain et al. [18] presented an express selection model with capacity restriction and defined the parameters for determining which express service is attractive on a bus corridor. Meanwhile, Ulusoy et al. [19] proposed a model for further improving the optimal strategy by incorporating various factors and including all-stop, short-turn, and express services to meet the heterogeneous demand. Larrain et al. [20] proposed a methodology in which the express services skip all the intermediate stops except for the initial and final stops. The authors reported that the introduction of zonal services can save more than 6% of the social costs. Wang et al. [21] formulated an MINLP model to design bus services from the perspectives of operators and users.

From another perspective, short-turn operation strategies are commonly used in urban rail transit and bus transit systems. The short-turn points for buses can at any stop, but urban rail transit faces more restrictions due to the operational rules. Hence, the consideration of the short-turn points in bus service design is less complicated. Ceder [22] presented a model to determine the feasible short-turn stops. Furth [23] found that it is necessary to coordinate the schedule with the short-turn pattern and proposed a model to obtain the optimal schedule offset. Delle Site and Filippi [24] considered the short-turn location and vehicle size as variables and presented a framework to design the bus service scheme on a given corridor. In comparison, urban rail transit is more complicated because of the limited capacity of the turn-back stations. The selection of the turn-back stations is the main variable in designing a short-turn service. The direction and capacity of the turn-back station must be considered in urban rail transit. In order to analyze the advantages and disadvantages of the various models in more detail, a comparison will be performed between the proposed model and the previous studies (as shown in Table 1). Li et al. [12] proposed a model to optimize the operation scheme that considers short-turn services and the train composition. A flexible operation scheme can be achieved with a variable train composition to cope with varying passenger demand. However, this will increase the purchase cost of trains because the train composition is fixed and the expenditure increases with the size of the train composition. In contrast, express services can relieve the pressure of imbalanced passenger demand without changing the train composition.

The above literature review shows that (i) previous studies have only focused on express and local services or short-turn services, and few studies have integrated express and local services with short-turn services in urban rail transit operation schemes. There are more related studies on the design of bus service networks. These methods cannot be used directly to solve urban rail transit service design problems because of the unique restrictions and considerations that exist in operating urban rail transit services such as the capacities of the turn-back stations and the sections; (ii) there is usually an assumption that express services are operated on full-length services, which may also lead to wasted transport capacity. In this work, express services are operated in conjunction with short-turn services.

3. Problem Description

In this section, we describe the URTSDP, which includes two subproblems: the service route design problem (SRDP) and the passenger assignment problem (PAP). Without loss of generality, we assume that the layout of the urban rail transit line and the turn-back stations are given and the operational rules are known. The URTSDP decides the number of service routes, frequency, and train stopping pattern with the objective of minimizing operator and user costs.

3.1. Service Route Design Problem. An operation scheme that includes the number of service routes and the stopping pattern of the express service is determined given the urban rail line, the positions of the turn-back stations, and the condition that there are no route limitations on the express services, which can operate on both full-length and short-turn routes. In addition, there are two constraints to be satisfied:

(1) Capacity limitation of the turn-back station: the service routes that end at these stations must not exceed the turn-back capacity of the station. Each

Table 1: Comparisons of this paper with previous studies.

Factors	Ulusoy et al. [19]	Larrain et al. [20]	Li et al. [12]	This paper
Assumption				
Demand	Fixed	Fixed	Fixed	Fixed
Transfer	Allowed	Allowed	Allowed	Allowed
Strategy	Short-turn services and one express service	Express service	Short-turn services	Short-turn services and express/local services
Model				
Single	✓	—	✓	✓
Bilevel	—	✓	—	—
Objective				
Operator cost	✓	✓	✓	✓
User cost	✓	✓	✓	✓
Constraints				
Transport capacity	✓	✓	✓	✓
Turn-back capacity	—	—	✓	✓
Minimum frequency	✓	—	✓	✓
Solving method				
Solution method	Exhaustive search	Heuristic + solver	Heuristic + solver	Heuristic + solver

service route incurs extra time when it turns back at the ending station. The time is called the turn-back time in urban rail transit lines.

(2) Maximum frequency limitation of each section: each train needs to operate on its corresponding service route in a predetermined sequence. The total number of all trains in each section cannot exceed the maximum frequency because the rail transit has only one line in the inbound or outbound direction.

We take the urban transit line with 6 physical stations and 3 turn-back stations shown in Figure 1 as an example. The stations are denoted as n_1 to n_8 in the outbound direction of the corridor, while the sections are denoted as e_1 to e_7. Stations n_1, n_4, and n_8 are the turn-back stations, and n_4 is designed for bidirectional turn-back.

Based on the given directions and positions of the turn-back stations and their associated capacities on the line, a turn-back station with an available direction can be used as the starting or ending station of the service route. A set of starting stations and a set of ending stations can thus be obtained in the initial stage. A service route consists of routes and nodes, where each node is associated with a physical station. A *line pool* is then obtained by varying the terminal stations among the starting or ending station sets. Figure 2 shows an operation scheme from the *line pool*. It consists of 3 service routes, where *route* 1 is short turn with local service, *route* 2 is short turn with express service, and *route* 3 is full length with express service.

The SRDP is primarily concerned with the frequency, short-turn, and stop pattern of the express service. The operation scheme including the above factors can be obtained by solving the SRDP. The example in Figure 2 consists of three service routes. The terminal of each route is set in accordance with the available starting or ending station set and the condition that each station is served by at least one service route. However, the operation scheme may be nonoptimal even after solving the SRDP because the frequency of each route was not determined in conjunction with the PAP to consider the total cost. To obtain (nearly) optimal operation schemes, the costs of the operator and users need to be considered. That is, the SRDP and PAP need to be solved simultaneously. This is a matter of great concern in academic studies.

3.2. Passenger Assignment Problem. Before determining the optimal passenger assignment strategy, the service network needs to be constructed in advance. A service network containing the arcs, nodes, and physical stations can be constructed depending on the operation scheme, as shown in Figure 2. A train starts from the origin node, goes through intermediate nodes, and ends at the destination node, wherein arcs are used to connect nodes with the associated physical stations, and passengers board (alight) trains operating on the routes served by these arcs. The arcs contain the specific routes of passenger travel and can be divided into three types of arcs, namely, the boarding, alighting, and riding arcs. A service network, as shown in Figure 3, includes a complete passenger traveling path that begins from his (her) origin station, goes through the boarding, riding, and alighting arcs, and eventually ends at his (her) terminal station.

For instance, a passenger travels from station n_1 to station n_5 along the travel path $n_1 \longrightarrow 17 \longrightarrow 18 \longrightarrow 19 \longrightarrow 20 \longrightarrow n_5$. He (she) will pass through a boarding arc $(n_1 \longrightarrow 17)$, riding arcs $(17 \longrightarrow 18 \longrightarrow 19 \longrightarrow 20)$, and an alighting arc $(20 \longrightarrow n_5)$. Therefore, a connection is made by the boarding arc between physical station n_1 and node 17, while the alighting arc performs the same function along the opposite direction. The riding arc makes the connection between two route stations. If one physical station is not

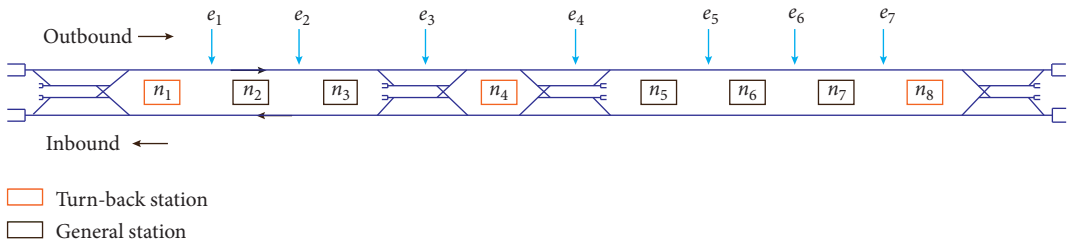


Figure 1: An urban rail transit corridor.

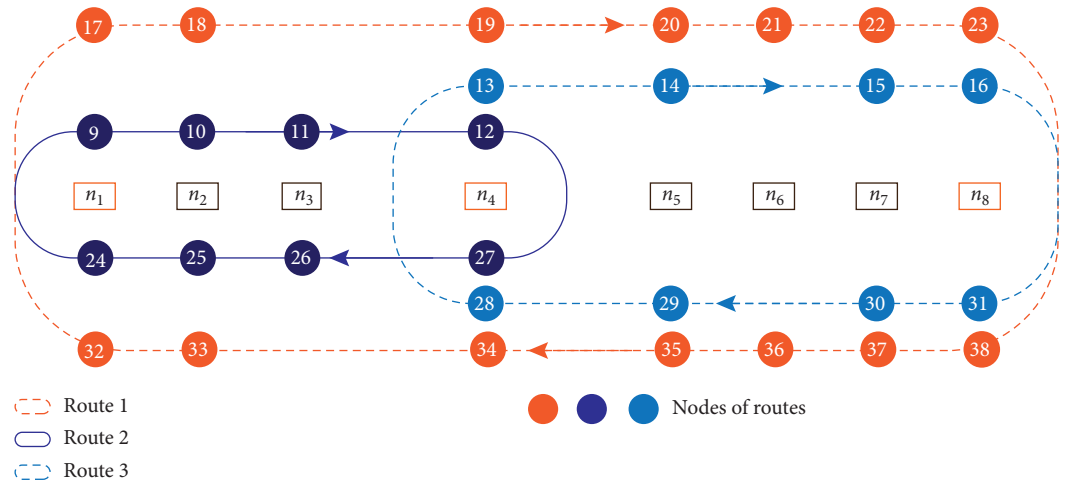


Figure 2: Service description of urban rail transit.

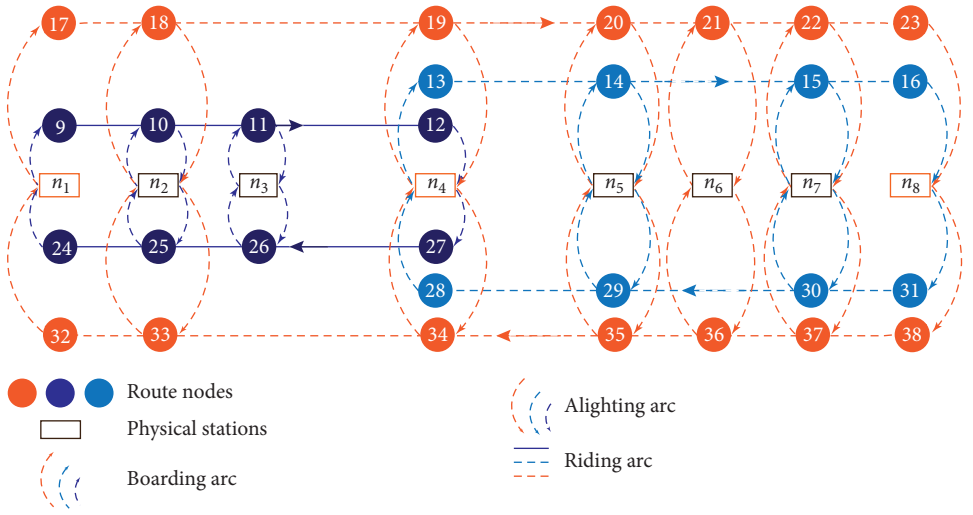


Figure 3: Description of directed service network.

served by a service route, then there is no connection between the physical station and the node. For n_3 shown in Figure 3, there is no boarding arc between n_3 and *route* 1.

As shown in Figure 3, there may be many boarding arcs between the physical stations and services. Therefore, it is important to analyze the service choice behavior of the passengers. To address this *common line* problem, Chriqui and Robillard [25] presented a solution to determine which service passengers would choose by defining the attractiveness of each service based on the passengers' waiting time and in-vehicle time. This approach does not require the prior preparation of fixed service network. Spiess and Florian [3] proposed a strategy for a passenger to reach the terminal station based on the least cost service in a model with a fixed transit line. In this work, we construct a transit assignment model with a known service network for urban rail transit

services based on the optimal strategy and consider the effectiveness of transfers on the passenger distribution further. A detailed discussion is given in the next section.

4. Model Formulation for URTSDP

4.1. Notation Definitions. We formulate the URTSDP problem as an MINLP model and transform the model into an MILP model in this section. The notations to be used in the model in what follows are explained in Table 2.

4.2. Modelling Assumptions. The model should ideally provide station-to-station services for operators and users as quickly as possible. The services proposed for the given stations are determined by the operator's cost and users' cost. These two factors must be considered simultaneously for a more effective operation scheme. The main assumptions of this study are as follows:

(1) The exogenous trip demand is considered to be fixed.

(2) The train composition is an important factor that affects the transport capacity and service level of the urban rail transit line, which are predetermined and fixed. In addition, the fleet size is limited.

(3) The passengers' arrival at each station is considered to be uniformly distributed, and the train arrivals follow an exponential distribution.

(4) We consider the passenger travel behavior under the condition of no congestions.

4.3. Total Costs of URTSDP. The total cost of URTSDP in this study consists of the operator's cost and the users' cost. The former includes two contributions. First, the train ownership cost includes the purchasing cost and related expenses during the operation time of the vehicle. The operation period of a train is typically 30 years [12]. The train ownership cost can be obtained based on the operation time obtained from the model. This promotes sensitivity to the operation time by distributing the cost between the hours on each working day. A second contribution to the operating cost is costs such as the electricity, staff, and track maintenance cost. The user cost can be divided into three parts: the waiting cost, the riding cost, and the transfer cost. The first two are dependent on the time spent and the transfer cost depends on the number of transfers.

Based on the characteristics of the URTSDP, we can divide the model into two parts: (i) a submodel to solve the SRDP and (ii) an optimal strategy to solve the PAP. The generation of the service network requires both models to be solved simultaneously.

4.3.1. SRDP Model. The SRDP determines the service routes and operation frequency on the urban rail transit line. The objective function is formulated as follows:

$$(\text{SRDP})\min z_1 = \sum_{f\in F}\sum_{l\in L} x_{fl}\cdot t_l \cdot c_0 + \sum_{f\in F}\sum_{l\in L} x_{fl}\cdot d_l \cdot c_1. \tag{1}$$

In the objective function (1), the first term is the ownership cost of the trains allocated to the express services and local services during the operation time, and the second term is the service distance-dependent operating cost. The major constraints in the model are as follows:

(i) Service level: the number of service routes should not exceed the limits on the operator's cost and users' travel cost. In addition, the frequency should not be lower than a specified positive number to meet the passenger demand while considering the surplus capacity. The formulation of the service level is as follows:

$$\sum_{l\in L} \mu_l \le r_{\max}, \tag{2}$$

$$\mu_l \cdot \lambda_{\min} \le \sum_{f\in F} x_{fl} \le \mu_l \cdot \lambda_{\max}, \quad \forall l \in L. \tag{3}$$

(ii) Station coverage and passenger demand satisfaction: each station should be served by at least one service route, and the passenger demand at each station should be satisfied because passenger demand is fixed. Constraints (4) describe the variables for the stations:

$$\sum_{f\in F}\sum_{l\in L} x_{fl}\cdot \pi_{li}\cdot u \ge p_e, \quad \forall i \in N. \tag{4}$$

(iii) Section coverage passenger flow volume satisfaction: each section should be covered by at least one service route, and the transport capacity should not be less than the volume of passenger flow in each section. Constraints (5) represents that the total number of trains should not be less than $\phi_{\min}$ and that the maximum frequency $f_{\max}$ in each section is not exceeded because operational rules restrict the frequency in each section to maintain a safety distance between two adjacent trains along the urban rail transit line:

$$\phi_{\min} \le \sum_{f\in F}\sum_{l\in L} x_{fl}\cdot k_{le} \le \phi_{\max}, \quad \forall e \in E, \tag{5}$$

$$(1-\gamma)\sum_{f\in F}\sum_{l\in L} x_{fl}\cdot k_{le}\cdot u \ge q_e, \quad \forall e \in E. \tag{6}$$

(iv) Turning back capacity constraint: the operation scheme can be made more flexible by introducing short-turn services. Each short-turn service route must return at a specified station owing to the properties of the urban transit line. However, each train turning back occupies the track for a specified time. Therefore, the turning back capacity of a station is limited in the inbound and outbound directions, and the number of trains turning back cannot exceed the maximum limitation:

Table 2: Definition of sets, indices, parameters, and variables.

Notations	Description
Sets	
V	Set of service route nodes with index n
N	Set of physical stations in urban rail transit line with index i, $i, \kappa \in N$. $N = N_1 \cup N_2$, where N_1 is the set of low-priority stations with medium or small volume of passenger flow and N_2 is the set of high-priority stations with large volume of passenger flow
N_l	Set of physical stations that service route l visited
M	Set of the turn-back stations
L	Set of service routes in urban rail transit service network with index l. $L = L_1 \cup L_2$, where L_1 is the set of express services and L_2 is the set of local services
E	Set of sections in urban rail transit line with index e
A	Set of arcs in a service network with index a. $A = A_1 \cup A_2 \cup A_3$, where A_1 is the set of boarding arcs, A_2 is the set of riding arcs, and A_3 is the set of alighting arcs
A_i^+	Set of departing arcs in a service network in station i
A_i^-	Set of arriving arcs in a service network in station i
O	Set of inbound and outbound direction with index o
F	Set of the operational trains with index f, $F = \{1, 2, 3, \ldots, f, \ldots, \lvert F\rvert\}$, f denotes that fth train to be operating in service route
W	Set of O-D pairs with index $w(o, d) \in W$, where o and d denote the origin and destination
Parameters	
u	Capacity of each train
$r_{\max}$	The maximum number of service routes
$\phi_{\max}$	Maximum operation frequency of the interval
$\phi_{\min}$	Minimum operation frequency of the interval
$\lambda_{\min}$	Minimum number of trains operation on the service route
c_0	Train operation cost per unit time
c_1	Service route operation cost per unit distance with index l and e
c_2	Passenger waiting cost per unit time
c_3	Passenger riding cost per unit time
c_4	Passenger transfer penalty per person
α_i^o	The turn-back capacity of station i in the direction of o
β	The total number of available vehicles
γ	Train capacity surplus
b_l	Maximum number of serving physical stations for service route l in one direction
t_l	Round-trip time of service route l
d_l	Round-trip distance of service route l
t_a	Time required to pass arc a
p_i	The number of passengers at station i in urban transit line
q_e	The volume of passengers flow in section e in urban transit line
g_κ^w	The number of trips generated by O-D pair w from station κ
G^w	The number of trips generated by O-D pair w
π_{li}	Binary parameter, equals one if service route l visits station i
k_{le}	Binary parameter, equals one if service route l visits section e
σ_l^{io}	Binary parameter, equals one if the service l turns back at station i to direction o
Definitional variables	
μ_l	Binary variable, equals one if the lth service route is activated
θ_f	Binary variable, equals one if the fth train is activated
x_{fl}	Binary variable, equals one if the fth train operation is on service route l
v_a^w	Nonnegative continuous variable, flow of O-D pair w over arc a
y_κ^w	Nonnegative continuous variable, waiting time of O-D pair w at station κ

$$\sum_{f \in F} \sum_{l \in L} x_{fl} \cdot \sigma_l^{io} \le \alpha_i, \quad \forall i \in M, o \in O. \tag{7}$$

(v) Fleet-size constraint: in an urban rail transit line, the total number of available trains is given and limited, and the overall transport capacity of the trains should be able to meet the passenger demand with the minimum number of trains wherein the number of trains operated should not exceed the fleet size:

$$\sum_{f \in F} \sum_{l \in L} x_{fl} \le \beta. \tag{8}$$

(vi) Constraint (10) is a validity inequality and imposes the limitation that the $(f+1)$th train can only be activated if and only if the fth train is activated. Constraint (10) indicates that only one service route can be selected and operated for a train in the activated state:

$$\theta_f \le \theta_{f+1l}, \quad \forall f \in \frac{F}{|F|}, \tag{9}$$

$$\sum_{l \in L} x_{fl} = \theta_f, \quad \forall f \in F. \tag{10}$$

4.3.2. PAP Model. A complete trip of a passenger is composed of three portions: (i) waiting for a train in a station, (ii) riding in the vehicle, and (iii) alighting to egress. There may additionally be transfers. We present a model including the passenger waiting time cost, in-vehicle time cost, and transfer cost. In the optimal strategy, the passengers would board the first train they meet without preferring any other trains [3]. The PAP model is built assuming that there is no congestion and that the train arrivals follow an exponential distribution:

$$\begin{aligned}(\text{PAP}) \min z_2 = {} & \sum_{\kappa \in N} \sum_{w \in W} y_\kappa^w \cdot c_2 + \sum_{w \in W} \sum_{a \in A_2} v_a^w \cdot t_a \cdot c_3 \\ & + \sum_{w \in W} \left(\sum_{a \in A_1} v_a^w - G^w \right) \cdot c_4,\end{aligned} \tag{11}$$

$$s.t.\ v_a^w \le y_\kappa^w \sum_{f \in F} x_{fa}, \quad \forall a \in A_\kappa^+, \kappa \in N, w \in W, \tag{12}$$

$$\sum_{a \in A_\kappa^+} v_a^w = \sum_{a \in A_\kappa^-} v_a^w + g_\kappa^w, \quad \forall \kappa \in N \cup V, w \in W, \tag{13}$$

$$\sum_{w \in W} v_a^w \le \sum_{f \in F} x_{fa} \cdot u, \quad \forall a \in A. \tag{14}$$

In the objective function (11), the first term is the passenger waiting time cost and reflects the service frequency. The second term is the ride time cost in the vehicle. The third term is the transfer penalty put in place to ensure that passengers enjoy direct service as much as possible. Constraint (12) is a relaxation evolved from the original model which contains the waiting time, frequency, and arc flow from the nodes. It distributes the passenger demand from nodes to arcs. The volume of the arc flow is influenced by the service frequency. Constraint (13) imposes flow conservation on the nodes. Constraint (14) ensures that the flow on the arc does not exceed the arc capacity.

5. MINLP Model

The model of URTSDP including two submodels, i.e., model SRDP (20)–(22) and model PAP (11)–(14) which can be simplified as (15)–(19). Objective functions (1) and (15) constitute an integrated optimization model owing to that they are measured by the same dimension.

(MINLP) min $z = z_1 + z_2$

s.t. Express service *line pool*: (20)–(22)

Solving route and frequency problem: (3)–(10)

Solving passenger assignment problem: (16)–(19)

The above model is a mixed-integer nonlinear program (MINLP) since constraints (12) and (16) are nonlinear.

5.1. Model Enhancement

5.1.1. Passenger Assignment Model Simplification. In the optimal strategy model, we obtain the volume of passenger flow by considering each OD pair. As the volume of passenger flow and the number of stations increase, the difficulty in solving the model increases correspondingly. It is a given property that the total passenger waiting time would not be affected if the OD pairs are aggregated. Consequently, we introduce the new variables v_a^i to denote the number of trips to station i over arc a, y_κ^i to denote the waiting time of passengers from station κ to station i, and G_i to denote the number of passengers at destination i. A new model is described as follows:

$$\begin{aligned}\min z_2 = {} & \sum_{\kappa \in N} \sum_{i \in N} y_\kappa^i \cdot c_2 + \sum_{i \in N} \sum_{a \in A_2} v_a^i \cdot t_a \cdot c_3 \\ & + \sum_{i \in N} \left(\sum_{a \in A_1} v_a^i - G^i \right) \cdot c_4,\end{aligned} \tag{15}$$

$$s.t.\ v_a^i \le y_\kappa^i \sum_{f \in F} x_{fa}, \quad \forall a \in A_\kappa^+, i \in N, \kappa \in N, \tag{16}$$

$$\sum_{a \in A_\kappa^+} v_a^i = \sum_{a \in A_\kappa^-} v_a^i + g_\kappa^i, \quad \forall \kappa \in N \cup V, i \in N, \tag{17}$$

$$\sum_{i \in N} v_a^i \le \sum_{f \in F} x_{fa} \cdot c, \quad \forall a \in A, \tag{18}$$

$$v_a^i, y_\kappa^i \ge 0, \quad \forall i \in N, a \in A, \kappa \in N. \tag{19}$$

5.1.2. Problem Reduction Based on Line Pool. As aforementioned, the service network must be completed before using the optimal strategy to tackle the assignment problem. The network is constructed based on the service routes, wherein a choice needs to be made if each station is to be served by an express service or not. Accordingly, for an urban rail transit line with numerous stations, the number of variables will increase exponentially. The binary parameter π_{li} indicates whether service route l stops at station i. There are 2^{n-2} possible service routes for express trains in the service network in a line with n stations. Additionally, many more variables will be generated for the v_a^i, which denotes the number of trips to station i over arc a because the three aforementioned types of arcs in the outbound and inbound directions for each service route must be considered. In fact, express service routes stop at special stations which attract large numbers of passengers in special areas or specific times

instead of at all the stations on the urban rail transit lines. In light of this, we define the set N_2 to account for the high-priority stations on the urban rail transit line. All express service trains must stop at these stations if the service route covers them. Constraint (21) is used to impose the relationship between the express service routes and the stations. Hence, the reduced service routes set can be obtained as the *line pool*. We will analyze the information on the stations as much as possible, especially the volume of the passenger flow. As a result, the *line pool* can be modeled by satisfying the following conditions:

$$\sum_{i \in N} \pi_{li} \leq b_l \cdot \mu_l, \quad \forall l \in L_1, \tag{20}$$

$$\pi_{li} = \mu_l, \quad \forall i \in N_2 \cap N_l, l \in L_1, \tag{21}$$

$$\pi_{li} \leq \mu_l, \quad \forall i \in N_1, l \in L_1. \tag{22}$$

Constraint (20) implies that the total number of stations served by service route l should not exceed the maximum number allowed. Equation (21) ensures that each high-priority station is served by at least one express service. Constraint (22) states that low-priority stations may be served by the express service routes.

5.2. Model Linearization. Constraint (12) is nonlinear because it is the product of a binary variable x_{fa} representing whether the fth train operates on arc a and a continuous variable y_i^w denoting the waiting time to station i for an OD pair w. Here, we transform the constraint into an MILP model and then solve it using CPLEX to obtain the (nearly) optimal solution. The reformulation linearization technique (RLT) replaces the nonlinear polynomials with new variables and adds linear constraints to the newly added variables to achieve the transformation [26]. We take the polynomial xy as an example where x is a binary variable and y is a nonnegative continuous variable. Given the assumption that $\underline{y} \leq y \leq \overline{y}$, the following formula can be obtained via the RLT:

$$\begin{aligned} &0 - \underline{x}\, y - x\, \underline{y} + \underline{x}\, \underline{y} \geq 0, 0 - \overline{x} y - x\overline{y} + \underline{x}\, \underline{y} \geq 0, \\ &\underline{x}\, y + x\overline{y} - \underline{x}\, \overline{y} - 0 \geq 0, \overline{x} y + x\, \underline{x} - \overline{x}\, \underline{y} - 0 \geq 0. \end{aligned} \tag{23}$$

The nonlinearities in (12) and (16) result from the product of the service route frequency variable x_{fa} and the waiting time variable y_κ^i and can be transformed into equivalent linear conditions by using inequality (23):

$$v_{fa}^i \leq M \cdot x_{fa}, \quad \forall f \in F, \forall a \in A_\kappa^+, i \in N, \kappa \in N, \tag{24}$$

$$v_{fa}^i \leq y_\kappa^i, \quad \forall f \in F, \forall a \in A_\kappa^+, i \in N, \kappa \in N, \tag{25}$$

$$v_{fa}^i \geq y_\kappa^i - M\left(1 - x_{fa}\right), \quad \forall f \in F, \forall a \in A_\kappa^+, i \in N, \kappa \in N, \tag{26}$$

$$v_a^i \leq \sum_{f \in F} v_{fa}^i, \quad \forall a \in A_\kappa^+, i \in N, \kappa \in N, \tag{27}$$

$$v_{fa}^i \geq 0, \quad \forall f \in F, \forall a \in A_\kappa^+, i \in N, \kappa \in N. \tag{28}$$

In the above constraints, M is a large positive constant that can be calculated using the equation $M = \sum_{\kappa \in N, \kappa \neq i} g_\kappa^i$. Obviously, we have $v_{\mathrm{fa}}^i = y_\kappa^i$ if and only if $x_{fa} = 1$. Constraint (28) limits the value of the variable v_{fa}^i. Constraints (4) and (5) can be solved by the textbook linearization method.

5.2.1. MILP Model. By the above linearization technique, a mixed-integer linear program (MILP) model can be obtained which is shown as follows. Commercial solvers like CPLEX can be used to solve the model:

(MILP) min $z = z_1 + z_2$

s.t. Express service *line pool*: (20)–(22)

Solving route and frequency problem: (3)–(10)

Solving passenger assignment problem: (17)–(19) and (24)–(28)

6. Solution Method

The MILP model can be solved by the CPLEX solver or other commercial solvers; however, preliminary computations show that the solution computational burden increases significantly as the number of variables and constraints increase. In this case, optimal schemes can be obtained by directly using commercial solvers only for small-scale problems, for example, an urban rail transit line with 10 stations. For a large-scale line, it would take more computational resources and time to obtain solutions than within a feasible limit. Therefore, it is essential to develop an algorithm to efficiently obtain high-quality solutions.

Based on the characteristics of the URTSDP, we design a search algorithm to solve large-scale problems. We first solve the SRDP model to obtain an optimal route operation scheme, wherein the cost of the operator is minimized (denoted as $z_{\mathrm{operator}}^{\min}$) while the users' cost may not be extremal. We then introduce a new constraint for the value of the SRDP model. The (nearly) optimal solution is obtained iteratively by increasing the constraint gradually with a certain step based on $z_{\mathrm{operator}}^{\min}$.

In summary, the search algorithm shown in Figure 4 includes two main parts: (1) determining the search range of the SRDP model and (2) obtaining the solution iteratively.

6.1. Operator Cost Screening Procedure. The passenger assignment problem model is solved based on frequency. Frequency of each route can be obtained by solving SRDP to optimal, wherein we can obtain the x_{fl} and π_{li} for operation scheme. The service network can be generated based on these two parameters. In this case, the PAP model is linear without transformation and can be solved by the commercial solver to obtain the user's cost. To do this, the total cost (denote as z') is easy to calculate, and the lowest value of the operator's cost is determined as $z_{\mathrm{operator}}^{\min}$. The upper bound of

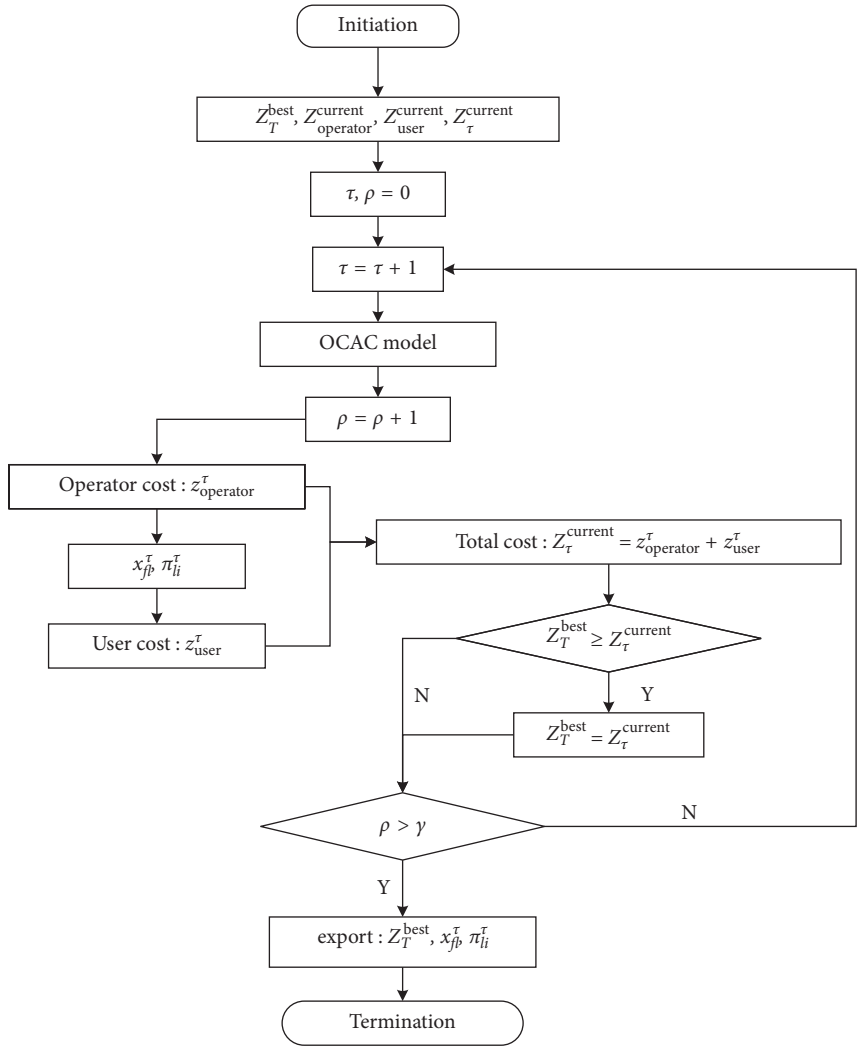


Figure 4: Flowchart of the search algorithm.

the SRDP model needs to be obtained by solving the PAP model to optimal. In general, passengers get better trips without the need for transfers when there is only one all-stop route with the maximal frequency. SRDP and PAP models can be solved based on the service network, which is constructed based on the one all-stop route with maximal frequency. The minimal user's cost hence can be obtained (denoted as z_{user}^{min}). Finally, operator's screening range is generated, and $\Delta = z' - z_{user}^{min} - z_{operator}^{min}$. Here, we assume that there exists an operator cost z_1, corresponding to a user's cost z_2. If $z_1 \geq z' - z_{user}^{min}$, then $z_1 + z_{user}^{min} \geq z'$; i.e., the total cost based on z_1 and z_2 is greater than z', and we therefore do not need to search the operator's cost which is beyond the range.

We can then introduce the operator cost augmented constraint (OCAC) in the following inequality shown in the SRDP model. This constraint is similar to the augmented ε-constraint method [27], and the inequality is a modified version based on it. Additionally, the parameter Ω is changeable and $\Omega = (\Delta \cdot \rho)/\eta$, wherein η is the number to partition the operator cost range Δ and $\rho \in \{0, 1, 2, 3, \ldots, \eta\}$. It is hence explicit to see that the right side of the inequality increases gradually as ρ increases:

$$(\text{OCAC}) z_1 \geq z_{operator}^{min} + \Omega. \tag{29}$$

6.2. Obtaining Solution in Iterative Algorithm

Step 1: initialize the parameters, including the station, and set the number of iterations, $i = 0$. Set the value of γ. Set $Z_T^{best} = \propto +, Z_\tau^{current} = 0$.

Step 2: increase the number of iterations, i.e., $\tau = \tau + 1$.

Step 3: solve the SRDP model, and record the operator's cost which is assigned to $z_{operator}^\tau$ and the value of $x_{fl}^\tau, \pi_{li}^\tau$ in iteration τ.

Step 4: generate passenger assignment service network in accordance with x_{fl}^τ and π_{li}^τ as input to the PAP model. Solve the model to obtain the value of

passenger's waiting cost which is assigned to z^{τ}_{user} and transfer cost in stage τ.

Step 5: set $Z^{\text{current}}_{\tau} = z^{\tau}_{\text{operator}} + z^{\tau}_{\text{user}}$, and record the value $Z^{\text{current}}_{\tau}$ in stage τ.

Step 6: check current value $Z^{\text{current}}_{\tau}$ and Z^{best}_{T}: if $Z^{\text{current}}_{\tau} \leq Z^{\text{best}}_{T}$, $Z^{\text{best}}_{T} = Z^{\text{current}}_{\tau}$; else turn to step 7.

Step 7: increase the value of ρ by 1

Step 8: check the stopping criteria. If $\rho > \gamma$, stop procedure and export Z^{best}_{T}, x^{τ}_{fl}, π^{τ}_{fl}; otherwise, return to Step 2.

7. Numerical Studies

In this section, an urban rail transit line with 19 stations is used to verify the effectiveness of the model for urban rail transit in Chengdu, China. There are three service strategies: short-turn, express, and local services. We generate several instances to evaluate the solution algorithm by comparing the results with the MILP model.

The optimization results of the model coded in C# were obtained by invoking CPLEX 12.8 on the Visual Studio 2017 programming platform on a PC with an Intel Core i5-6200U 2.40 GHz CPU and 12 GB RAM. A limit on the total number of services was imposed throughout the entire optimization process.

7.1. Instance Generation. To verify the computational performance of the solution algorithm, the 5 OD matrices shown in Figure 5 were generated. The lines were generated randomly, and they each consist of six to ten stations with one turn-back station.

For convenience, we describe each line with the triplet (A, B, C). Here, A denotes the line number, B denotes the number of stations, and C denotes the number of turn-back stations in the line. For example, (1, 5, 1) denotes route 1 with five stations and one turn-back station.

7.2. Comparison of MILP Model and Search Algorithm. We analyzed the computational results of the MILP model. When we attempted to solve a line with 12 stations and 1 turn-back station, the optimal solution could not be obtained within the time limit. There was a significant difference in the computation time between 10 stations and 9 stations. The other details are displayed in Table 3.

The results obtained from the preliminary analysis of the two methods are compared in Table 3. These results highlight the differences between the MILP model and the search algorithm. First, the computational time increases exponentially in the MILP model. In contrast, it increases slowly in the search algorithm. This discrepancy can be attributed to the considerable scale of the arcs generated by all the feasible service routes. As Table 3 shows, the computational time of 2535.0 s for the MILP model is much larger than the 172.1 s of the search algorithm when there are 10 stations on the lines. It is apparent that the number of variables is the distinguishing feature between the two methods; there were up to 1754417 variables in the MILP model, while the search algorithm had only 2100 variables in the (5, 10, 1) instance. Notably, small gaps in the results between the two methods were obtained despite the significant difference in the two aspects mentioned above. From the table, the maximum gap is 1.93% and the minimum gap is 0.19%, and there is an average gap of 1.31% in the five instances. In addition, the MILP model took considerable computational time to solve the line with 10 stations, and we could not even obtain a feasible solution when the line contained 12 stations with the time limit 12 h. The small gaps demonstrate that the search algorithm can be used to obtain high-quality operation schemes.

7.3. Practical Case and Results. A practical line with a length of 86.6 km and 5 turn-back stations is shown in Figure 6. n_1, n_{12}, n_{16}, n_{17}, and n_{19} are the turn-back stations at which only trains along one direction can turn back. For the purpose of verifying the search algorithm, we use the demand in the evening peak period shown in Figure7(c).

The parameters are set as follows:

(1) Total available train size = 42 trains

(2) Train capacities: local service = 1520 passenger/train and express service = 1520 passenger/train

(3) Turn-back capacities: total available turn-back stations are n_1, n_{12}, n_{16}, n_{17}, and n_{19}, and turn-back capacities of them are all 20 trains/h. One-way turning back time = 3 min

(4) Operating cost per kilometer: local service = 185.44 RMB/km and express service = 185.44 RMB/km

(5) Ownership cost per hour: local service = 467 RMB/h and express service = 467 RMB/h

(6) Passengers cost: waiting cost = 31 RMB/h, vehicle time cost = 31 RMB/h, and transfer penalty cost = 5 RMB/h

(7) Service level per route: maximum frequency = 20 trains/h, minimum frequency = 6 trains/h, and surplus capacity ratio = 26.3%

Practical case is introduced to determine the (nearly) optimal combinations of service strategies in previous sections, the operation scheme is shown in Figure 8, and details are obtained as follows:

(1) Optimal express service pattern:

Express service route l_1: $1 \longrightarrow 5 \longrightarrow 6 \longrightarrow 10 \longrightarrow 12$

Express service route l_2: $1 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow 6 \longrightarrow 7 \longrightarrow 8 \longrightarrow 9 \longrightarrow 10 \longrightarrow 11 \longrightarrow 12 \longrightarrow 13 \longrightarrow 14 \longrightarrow 15 \longrightarrow 16 \longrightarrow 17$

Local service route l_3: $1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow 6 \longrightarrow 7 \longrightarrow 8 \longrightarrow 9 \longrightarrow 10 \longrightarrow 11 \longrightarrow 12$

Local service route l_4: $1 \longrightarrow 2 \longrightarrow 3 \longrightarrow 4 \longrightarrow 5 \longrightarrow 6 \longrightarrow 7 \longrightarrow 8 \longrightarrow 9 \longrightarrow 10 \longrightarrow 11 \longrightarrow 12 \longrightarrow 13 \longrightarrow 14 \longrightarrow 15 \longrightarrow 16 \longrightarrow 17 \longrightarrow 18 \longrightarrow 19$

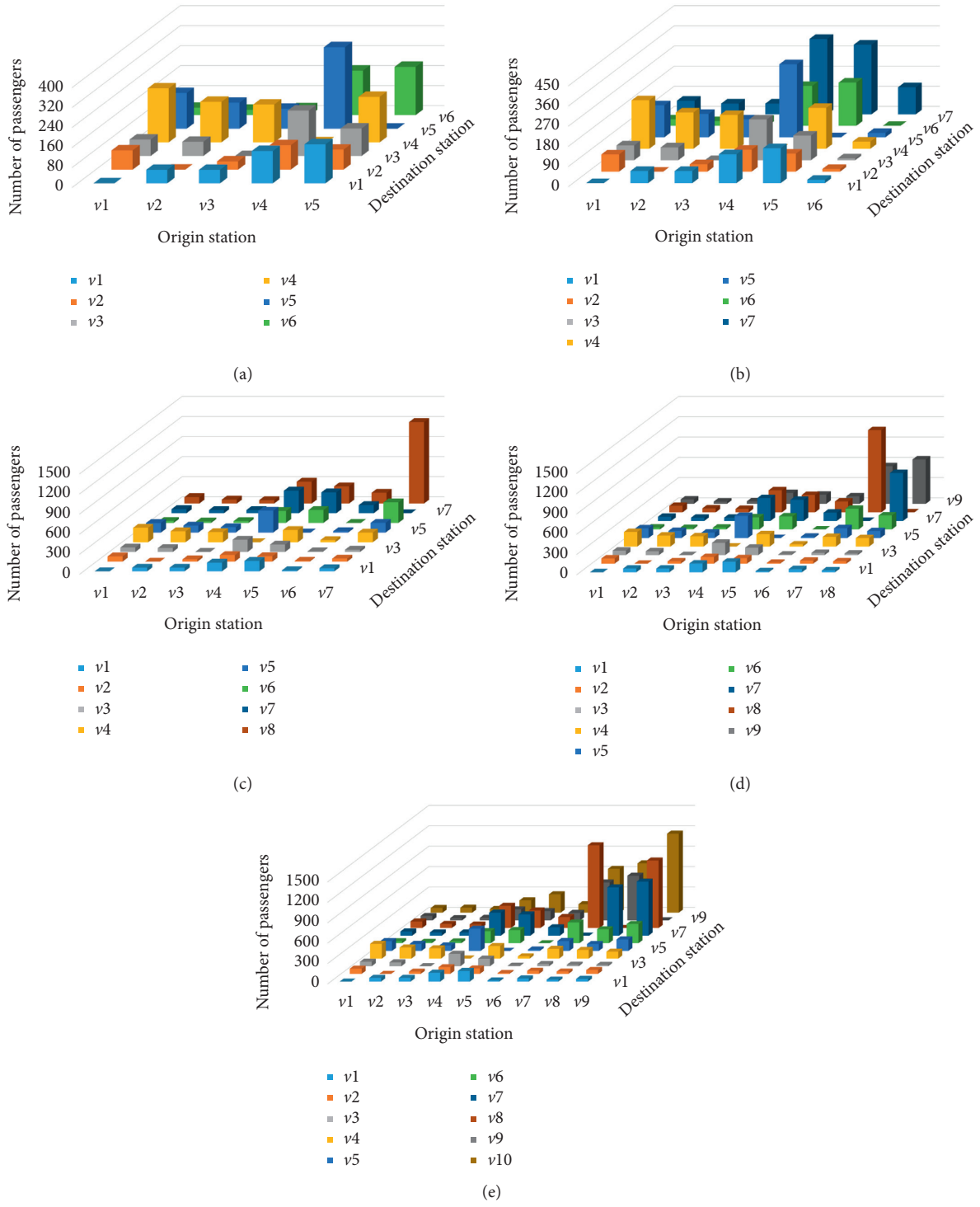


Figure 5: Passenger OD matrices of 5 lines. (a) Six-station OD demand. (b) Seven-station OD demand. (c) Eight-station OD demand. (d) Nine-station OD demand. (e) Ten-station OD demand.

(2) Optimal operation frequency for each service:

l_1: 10 trains/h
l_2: 8 trains/h
l_3: 8 trains/h
l_4: 10 trains/h

(3) Urban rail transit total cost: 825629 RMB

(4) Computation time: 3613.9 s

Figure 8 shows that the operation scheme contains four service routes consisting of two express service routes and two local service routes. One of the most notable aspects of

Table 3: Computational results of the models.

Instances	MILP model					Search Algorithm			
	Variable size	Objective value (RMB)	Best bound (RMB)	Gap (%)	Time (s)	Variable size	Objective value (RMB)	Gap (%)	Time (s)
(1, 6, 1)	51978	46609	46609	0.00	2.8	1044	47253	1.38	74.2
(2, 7, 1)	119147	55378	55374	0.01	5.3	1281	56442	1.93	77.7
(3, 8, 1)	286494	79889	79887	0.00	44.3	1536	80043	0.19	125.8
(4, 9, 1)	697875	124740	124734	0.00	300.3	1809	126401	1.34	127.5
(5, 10, 1)	1754417	198218	198210	0.00	2535.0	2100	201629	1.72	172.1

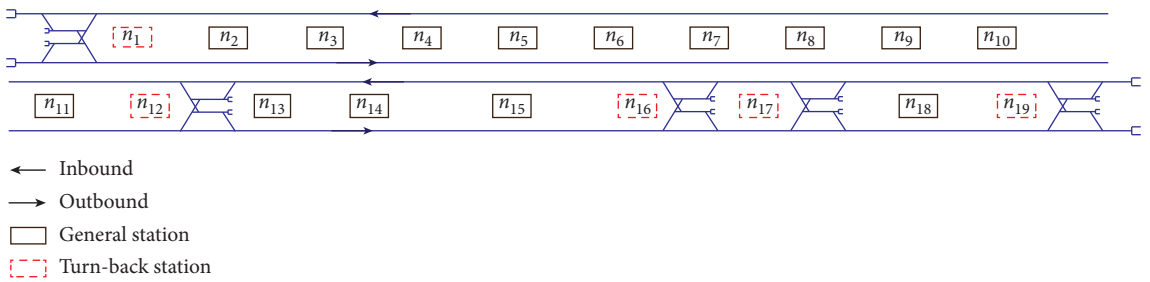


Figure 6: Layout of an urban transit line in Chengdu.

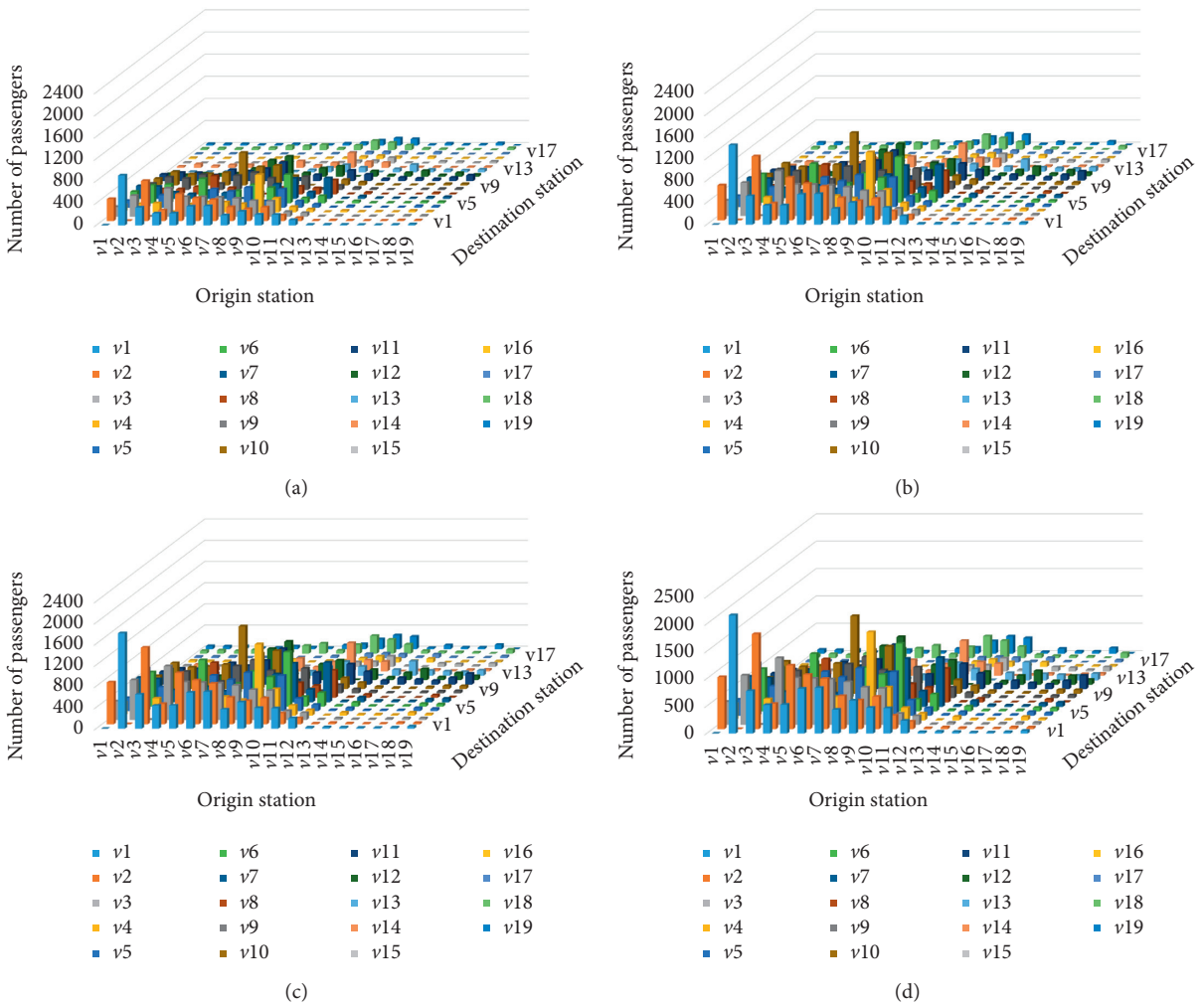


Figure 7: Multiple passenger demand levels. (a) 50% demand with respect to original demand. (b) 80% demand with respect to original demand. (c) Original demand. (d) 120% demand with respect to original demand.

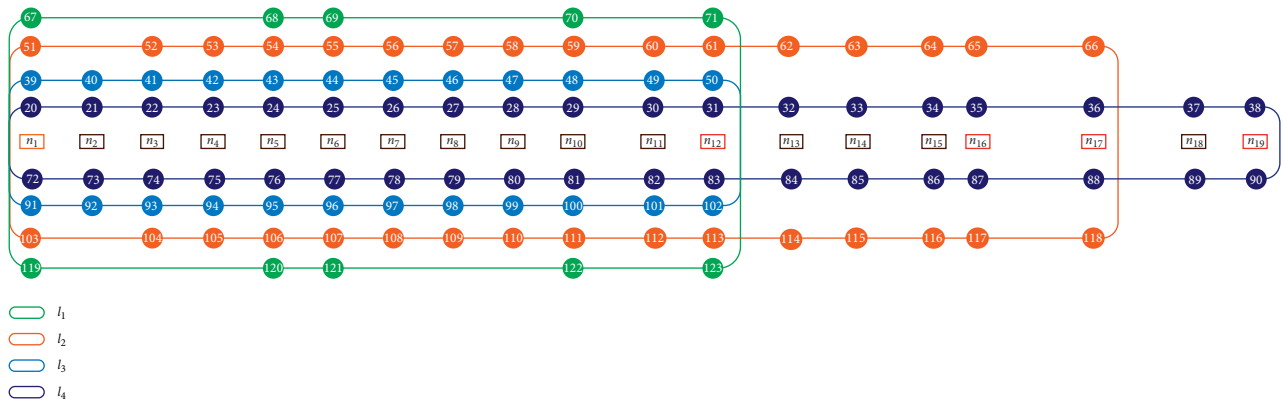


Figure 8: Optimized operation scheme for an urban rail transit line.

the operation scheme is the simultaneous existence of both express and local services in one route in which the express trains along service route l_1 stop at only five stations. Another express service route contains 16 stopping stations, except for station n_2. A possible explanation for the difference in the number of stopping stations in the two express service routes is the passenger demand for medium- to long-distance travel. Skipping stations allows the express service routes to provide different transport capacities to all the stations covered while meeting the passenger demand at each station. Therefore, the operation scheme is different for different passenger demand levels. To verify the effectiveness of the algorithm, the effect of the demand level on the number of service routes and stopping pattern is further studied. Operation schemes were generated with different given levels of demand in the multiple-scenario analysis. Figure 7 shows the multiple demand levels used for evaluation.

7.4. Multiple-Scenario Analysis. Using the solution algorithm, different operation schemes can be obtained under different scenarios, such as varying levels of passenger demand. The impact of the passenger demand can be evaluated by inputting various levels of passenger demand into the solution algorithm. We evaluated the optimization capability of the solution algorithm at different demand levels and obtained distinguishing results with different express service routes, costs, and frequencies. The results for the multiple scenarios are shown in Table 4.

As can be seen from Table 4, high-quality operation schemes were generated for different demand levels by the search algorithm. There are distinctions in the total cost, transit line setting, solution time, and service frequency between the three passenger demand scenarios. The 120% demand scenario has the largest total cost, solution time, and frequency of 912395 RMB, 3620 s, and 38 trains/h, respectively, because of the high demand level in this period. It can be concluded that the higher the level of demand, the higher the cost and the solution time. It is hence necessary for operators and users to design suitable operation schemes.

To further analyze how the parameters influence the operation scheme and total cost, we varied the number of search range partitions. As Figure 9 shows, the number of grids determines the partitioning of the search range. This value is of great importance in exploring the variable space. In this experiment, the number of grids was gradually increased from 0 to 1800. A larger value indicates a more finely divided search range with a higher possibility of finding a higher-quality solution. However, more resources are required. When the number of grids is 0, the augmented model does not place a constraint on the objective function. When the number of grids is between 0–200 and 500–600, the total cost shows a rapid downward trend. When the number of grids reaches 1000, the value of the objective function changes only slightly as the number of grids is further increased. This shows that a rougher grid may miss the optimal solution, whereas an excessively high grid number does not significantly improve the search accuracy.

7.5. Comparison with Empirical Operation Scheme. We compared two schemes, namely, (i) a high-quality operation scheme based on the solution algorithm and (ii) an empirical scheme in operation in the urban rail transit line to verify the search algorithm. The empirical routes and the morning peak passenger volume were given and the differences in the total cost, frequencies, and line settings were determined. Table 5 shows a detailed comparison of the two schemes.

There are three service routes in the empirical scheme, while four service routes exist in the high-quality scheme. This implies that there are many more nodes in the high-quality scheme than in the empirical scheme. However, more trains are employed in the empirical scheme, and the total cost is higher. The cost of 825629 RMB in the high-quality scheme is 7.76% lower than the 895078 RMB in the empirical scheme. Additionally, the costs to the operator and the users in the high-quality scheme are both lower than those of the empirical scheme with relative cost reductions of 2.18% and 12.96%, respectively. There are two likely causes for the total cost differences between the two schemes. From the operator's perspective, the higher cost to the operator in the empirical scheme is attributed to the larger number of operating trains due to the positive correlation between the

Table 4: Scenarios analysis with respect to original demand levels.

Demand with respect to original demand (%)	Total cost (RMB)	Solution time (s)	Transit lines setting	Service frequency (trains/h)
120	912395	3620	1 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17	8
			1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12	20
			1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17 ⟶ 18 ⟶ 19	10
80	630096	1681.1	1 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 12	8
			1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12	8
			1 ⟶ 2 ⟶ 3 ⟶ 5 ⟶ 6 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17 ⟶ 18 ⟶ 19	10
50	421167	1305.5	1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12	8
			1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17 ⟶ 18 ⟶ 19	8

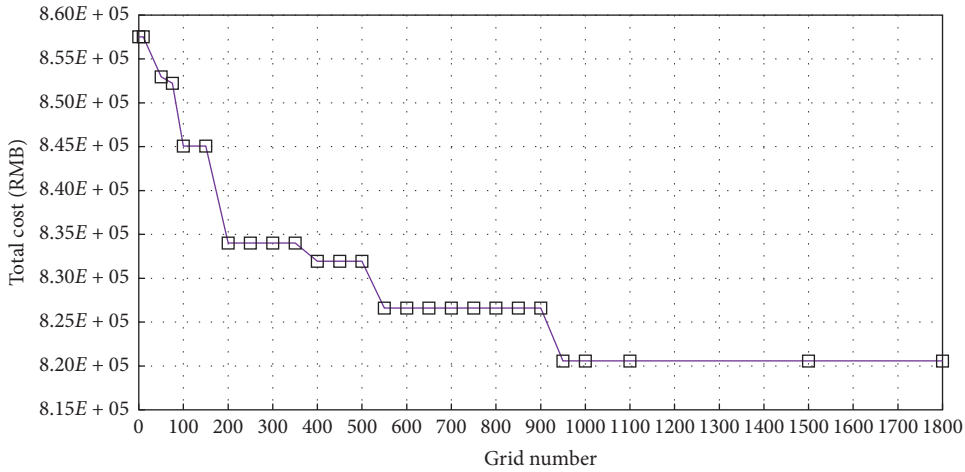


Figure 9: Sensitivity studies on total costs based on different partitions.

Table 5: Result comparison of the two schemes.

Scheme type	Total cost (RMB)	Operator cost (RMB)	User cost (RMB)	Transit lines setting	Service frequency (trains/h)
High-quality scheme	825629	422333	403296	1 ⟶ 5 ⟶ 6 ⟶ 10 ⟶ 12	10
				1 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17	8
				1 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12	8
				1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17 ⟶ 18 ⟶ 19	10
Empirical scheme	895078	431728	463350	1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 7 ⟶ 10 ⟶ 15 ⟶ 16	8
				1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12	24
				1 ⟶ 2 ⟶ 3 ⟶ 4 ⟶ 5 ⟶ 6 ⟶ 7 ⟶ 8 ⟶ 9 ⟶ 10 ⟶ 11 ⟶ 12 ⟶ 13 ⟶ 14 ⟶ 15 ⟶ 16 ⟶ 17 ⟶ 18 ⟶ 19	8

cost and the number of trains. 40 trains are employed in the empirical scheme, while the employment of only 36 trains in the high-quality scheme can save considerable vehicle purchase costs. From the passengers' perspective, the provision of more service routes to passengers gives them the choice for flexible travel options without waiting for a specific service. For example, passengers can use the express service route from station 1 to station 12 or other stations, but they can also use the local service route when more waiting time is needed for the express train.

Accordingly, we carried out a detailed analysis of the section flows and capacity utilization ratios (CURs) in the two schemes. By comparing the passenger flow volumes in the sections, the influence of the service setting on the passenger flow accumulation can be understood. The CUR indicates whether the service design is reasonable. The empirical scheme and high-quality scheme are, respectively, denoted in the figure as ES and OS for short. ES and OS, respectively, signify the capacities of the empirical and optimized schemes in each section.

As shown in Figure 10, the CUR of the OS is higher than that of ES in most sections because the transport capacity of most sections in the OS is higher than that of the ES, indicating that the trains are distributed more efficiently. The maximum OS CUR is 63.56%, which is considerably higher than the ES CUR of 57.20% at e_6 in the inbound direction. Moreover, the average OS CUR is 0.52% higher than the average ES CUR. In the outbound direction, the maximum OS CUR is 57.65% higher than the ES CUR of 51.89% at e_8. The reasons for the differences are clear: the OS uses express services in conjunction with short-turn services, and service routes from station 1 to station 12 employ both express service and local services. The transport capacity can thus be fully utilized. In addition, the passenger demand is mainly concentrated at the first twelve stations, which account for 81.37% of the total passenger demand. These indicators explain the rationality and efficiency of the scheme based on the proposed model.

Regarding the waiting time, Figure 11 shows that the waiting time of the ES at most stations is lower than that of the OS, that is, passengers have to wait longer for the same service under the high-quality scheme. This is because the train frequency in the ES is higher than that in the OS. Although there is

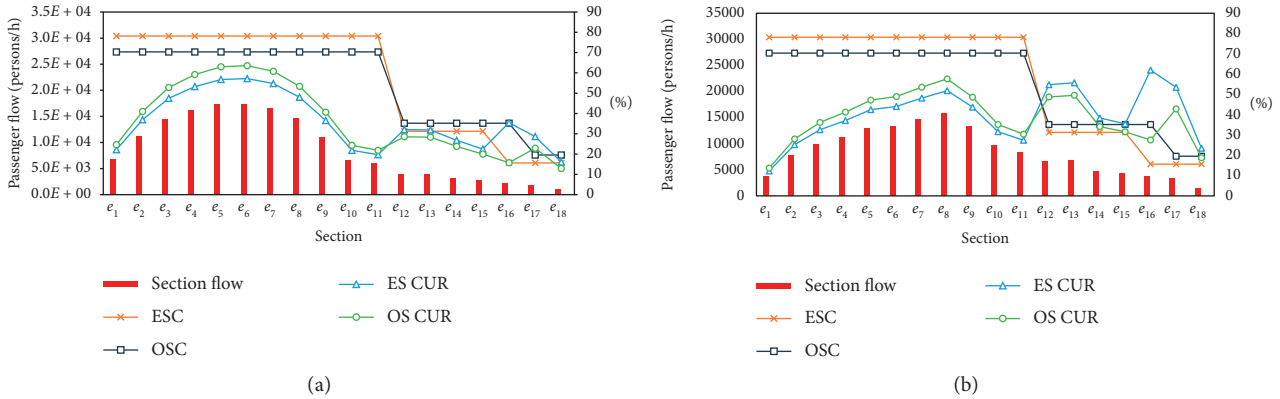


Figure 10: Section flow and capacity utilization ratio. (a) Inbound direction. (b) Outbound direction.

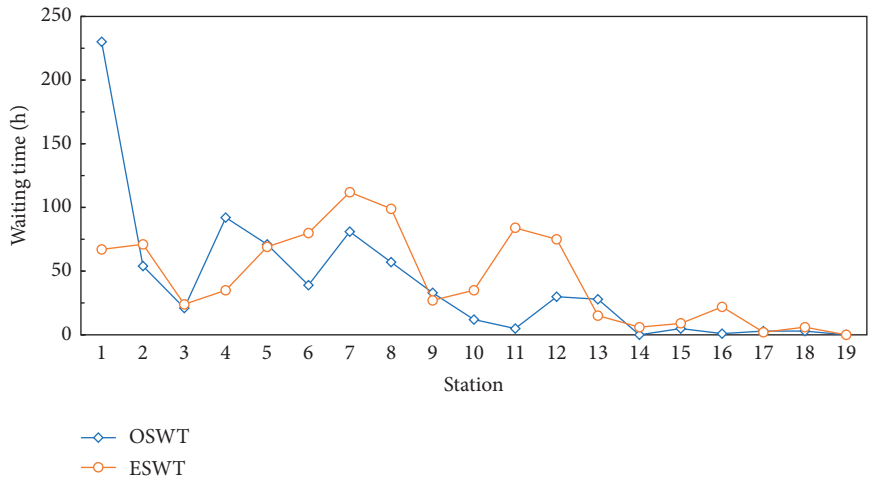


Figure 11: Passenger waiting time distribution.

a large difference in the waiting time in station 1 of up to 163 h, the total waiting time in the OS is lower than that of ES, and the reduction in waiting time is 73 h.

In summary, the present results are significant in two major respects. First, the operator's cost can be reduced by reducing the number of vehicle purchases. In practice, the number of rolling stocks can be reduced through a reasonable and high-efficiency operation scheme as verified by the aforementioned example. Furthermore, a high-quality operation scheme also benefits the passengers and saves their waiting time and riding time by integrating multiple strategies.

8. Conclusions

In this paper, we proposed an MILP model for the URTSDP, wherein short-turn, express, and local services are provided. This problem can be analyzed from two aspects. Firstly, the SRDP determines which station to turn back at, the numbers and frequencies of the service routes, and the routes to run and stations to serve for each express service route. Secondly, an optimal strategy is applied to solve the PAP while considering the capacity restrictions. Because the MILP model can be solved directly by commercial solvers for small-scale problems, we introduced the *line pool* strategy to reduce the scale of the problem and designed a search algorithm to solve the large-scale problem by minimizing the total cost. In addition, we applied the model and search method in a practical urban rail line under multiple scenarios. This differs from previous works that tested their models and methods only on small-scale scenarios.

Numerical studies were conducted to analyze the performance of the proposed model and the search algorithm. An instance comparison verified that the search algorithm could find high-quality solutions with a gap of 1.31% compared to the MILP model. In particular, compared to the empirical operation scheme, the high-quality scheme demonstrated practicability and effectiveness through a total cost saving of 7.76%. The performance of the search algorithm was tested under multiple scenarios and different parameter values. The multiple-scenario analysis illustrated the applicability of different operation schemes to meet different demands. A sensitivity eperiment was performed to investigate the effects of different partitionings of the search range on the total cost and solution time.

This study contributes to academic research by introducing the practical strategy of the *line pool* and the solution algorithm to solve a large-scale problem. However, there are several future refinements that are worth studying. (i) The train size of each service route is not included into the model. It may be worthwhile to use trains of different sizes to cope with various demand levels to further conserve transport capacity. (ii) The local service stopping pattern is not addressed in this model. Addressing this issue can alleviate the demand imbalance and save the operator's cost. (iii) In the future, the model can be developed in conjunction with additional constraints, such as real-time adjustments [28], robustness, and other factors to create a better match between supply and demand.

Data Availability

The data are available by contacting Siyu Tao through taosiyu@swjtu.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (Project no. 2017YF B1200701), the Research on Complete Technology of Integrated Transportation Hub in Complex Urban Environment (Grant no. cstc2014yykfb30003), and the Research on Functional Layout and Planning of Shapingba Integrated Transportation Hub (Grant no. 2015H01372).

References

[1] Y. Gao, L. Yang, and Z. Gao, "Energy consumption and travel time analysis for metro lines with express/local mode," *Transportation Research Part D: Transport and Environment*, vol. 60, pp. 7–27, 2018.

[2] S. Li, R. Xu, and K. Han, "Demand-oriented train services optimization for a congested urban rail line: integrating short turning and heterogeneous headways," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1459–1486, 2019.

[3] H. Spiess and M. Florian, "Optimal strategies: a new assignment model for transit networks," *Transportation Research Part B: Methodological*, vol. 23, no. 2, pp. 83–102, 1989.

[4] J. H. Chun, R. M. Anderson, and D. Paik, "The S-train system: synchronized express & local trains for urban commuter rail systems," in *Proceedings of the 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1586–1591, IEEE, Washington, DC, USA, October 2011.

[5] A. Jamili and M. Pourseyed Aghaee, "Robust stop-skipping patterns in urban railway operations under traffic alteration situation," *Transportation Research Part C: Emerging Technologies*, vol. 61, pp. 63–74, 2015.

[6] E. A. Abdelhafiez, M. R. Salama, and M. A. Shalaby, "Minimizing passenger travel time in URT system adopting skip-stop strategy," *Journal of Rail Transport Planning & Management*, vol. 7, no. 4, pp. 277–290, 2017.

[7] Q. Luo, Y. Hou, W. Li, and X. Zhang, "Stop plan of express and local train for regional rail transit line," *Journal of Advanced Transportation*, vol. 2018, Article ID 3179321, 11 pages, 2018.

[8] A. Yang, B. Wang, J. Huang, and C. Li, "Service replanning in urban rail transit networks: cross-line express trains for reducing the number of passenger transfers and travel time," *Transportation Research Part C: Emerging Technologies*, vol. 115, p. 102629, 2020.

[9] Y. Sun, P. M. Schonfeld, Y. Lu, and M. Zhou, "Redesigning rail transit short-turn operations: case study of line 2 of the Shanghai metro in China," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2540, no. 1, pp. 46–55, 2016.

[10] X. Ding, S. Guan, D. J. Sun, and L. Jia, "Short turning pattern for relieving metro congestion during peak hours: the substance coherence of Shanghai, China," *European Transport Research Review*, vol. 10, no. 2, 2018.

[11] M. Zhang, Y. Wang, S. Su, T. Tang, and B. Ning, "A short turning strategy for train scheduling optimization in an urban rail transit line: the case of Beijing subway line 4," *Journal of Advanced Transportation*, vol. 2018, 2018.

[12] Z. Li, J. Zhao, and Q. Peng, "Optimal train service design in urban rail transit line with considerations of short-turn service and train size," in *Proceedings of the 8th International Conference on Railway Operations Modelling and Analysis (ICROMA) RailNorrköping 2019*, pp. 665–687, Linköping University Electronic Press, Norrköping, Sweden, June 2019.

[13] Z. Cao, A. Ceder, D. Li, and S. Zhang, "Robust and optimized urban rail timetabling using a marshaling plan and skip-stop operation," *Transportmetrica A: Transport Science*, vol. 16, no. 3, pp. 1217–1249, 2020.

[14] L. L. Afanasiev and S. Y. Liberman, "Principles for organizing express bus services," *Transportation Research Part A: General*, vol. 17, no. 5, pp. 343–346, 1983.

[15] T. A. S. Vijayaraghavan and K. M. Anantharamaiah, "Fleet assignment strategies in urban transportation using express and partial services," *Transportation Research Part A: Policy and Practice*, vol. 29, no. 2, pp. 157–171, 1995.

[16] M. T. Conlon, P. J. Foote, K. B. O'Malley, and D. G. Stuart, "Successful arterial street limited-stop express bus service in Chicago," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1760, no. 1, pp. 74–80, 2001.

[17] D. Z. W. Wang and H. K. Lo, "Multi-fleet ferry service network design with passenger preferences for differential services," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 798–822, 2008.

[18] H. Larrain, R. Giesen, and J. C. Muñoz, "Choosing the right express services for bus corridor with capacity restrictions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2197, no. 1, pp. 63–70, 2010.

[19] Y. Y. Ulusoy, S. I.-J. Chien, and C.-H. Wei, "Optimal all-stop, short-turn, and express transit services under heterogeneous demand," *Transportation Research Record*, vol. 2197, no. 1, pp. 8–18, 2010.

[20] H. Larrain, J. C. Muñoz, and R. Giesen, "Generation and design heuristics for zonal express services," *Transportation Research Part E: Logistics and Transportation Review*, vol. 79, pp. 201–212, 2015.

[21] D. Z. W. Wang, A. Nayan, and W. Y. Szeto, "Optimal bus service design with limited stop services in a travel corridor," *Transportation Research Part E: Logistics and Transportation Review*, vol. 111, pp. 70–86, 2018.

[22] A. Ceder, "Optimal design of transit short-turn trips," *Transportation Research Record*, vol. 1221, no. 557, pp. 8–22, 1989.
[23] P. G. Furth, "Short turning on transit routes," *Transportation Research Record*, vol. 1108, pp. 42–52, 1987.
[24] P. Delle Site and F. Filippi, "Service optimization for bus corridors with short-turn strategies and variable vehicle size," *Transportation Research Part A: Policy and Practice*, vol. 32, no. 1, pp. 19–38, 1998.
[25] C. Chriqui and P. Robillard, "Common bus lines," *Transportation Science*, vol. 9, no. 2, pp. 115–121, 1975.
[26] H. D. Sherali and W. P. Adams, *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Springer Science & Business Media, Berlin, Germany, 2013.
[27] G. Mavrotas and K. Florios, "An improved version of the augmented constraint method (AUGMECON2) for finding the exact pareto set in multi-objective integer programming problems," *Applied Mathematics and Computation*, vol. 219, no. 18, pp. 9652–9669, 2013.
[28] Z. Cao, A. Ceder, and S. Zhang, "Real-time schedule adjustments for autonomous public transport vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 109, pp. 60–78, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7436710, 13 pages
https://doi.org/10.1155/2020/7436710

Research Article

An Empirical Analysis for Mode Choice in a Short-Distance Trip with Personal Rapid Transit

Hyunmyung Kim,[1] Haneum Seok,[2] Soyoung Iris You,[3] and Changju Lee[4]

[1]*Department of Transportation Engineering, Myongi University, Yongin 17058, Republic of Korea*
[2]*Hanam Police Station, Gyeonggi Nambu Province Policy Agency, Hanam 13024, Republic of Korea*
[3]*Innovative Transport Policy Division, Korea Railroad Research Institute, Uiwang 16105, Republic of Korea*
[4]*Virginia Transportation Research Council, Charlottesville, VA 22903, USA*

Correspondence should be addressed to Changju Lee; cl8ax@virginia.edu

Received 21 January 2020; Revised 2 September 2020; Accepted 5 October 2020; Published 17 November 2020

Academic Editor: Prakash Ranjitkar

Recently, there have been emerging demands for new transportation modes, such as personal rapid transit (PRT), to improve the connectivity of first and last mile travel. Advancement of ICT and growing concerns over environmental issues reinforce such demands through which specific transportation modes can satisfy the need of each individual for short-distance trips. Although PRT has received particular attention for short-distance trips, it is true that recent approaches have been developed to analyze the behavior of travelers for mid- to long-distance trips that are not relevant for short-distance trips. This study proposed a suitable approach using logistic regression models that could assist the understanding of features which determine mode choice in a short-distance trip. The mode choice for PRT in short-distance trips in this study was based on the data from the survey. After considering various factors, it was apparent that the purpose of the trip together with weather conditions impacted significantly on travelers' mode choices to PRT in short-distance trips. Additionally, it is expected that this study will play an important initial role in analyzing emerging transportation modes that can more easily respond to new demands for short-distance trips.

1. Introduction

In the four-step travel demand modeling process, mode choice models compute the proportion of trips that use a specific transportation mode between each origin and destination estimated from the trip distribution step. The capacity to explain the mode choice models following the introduction of a new mode or improvement of transportation service directly affects the accuracy of estimated trips by each mode, e.g., [1–3]. Mode choice models, consequently, play a critical role in decisions about investments in transportation [4].

Most current mode choice models target inter- or intra-regional travel of mid/long-distance trips. The assessment of investments or plans for transportation infrastructure, considering the movement of transportation modes separately and commonly, does not include the entire trip of travelers. Consequently, travel distance and/or time of each transportation mode (e.g., automobile, rail, bus, subway, and taxi) have been employed as significant factors that are suggested to be implemented from previous studies, e.g., [2, 3, 5, 6].

The advent of a new user-oriented paradigm requires that traditional supply-oriented transportation services are changed to user-centered ones. Concurrently, road infrastructure for mid/long-distance trips in many metropolitan cities has already reached the maturity stage. A novel approach of smart mobility using small-sized modes for the short-distance trip has been gaining in popularity. Environmental issues arising from traffic problems linked to the Paris Agreement [7] have created a new social norm to deliver a convenient, healthy, and sustainable society, with interest in special transportation modes for short-distance trips. A short-distance trip requires analyzing new factors

including the level of service by public transportation, which have not been conducted due in part to the lack of suitable models or usable data [8].

The short-distance refers to a walkable distance on which walking, motorcycling, and bicycling currently comprise the major transportation modes. Travel of less than one km has often been ignored even in the short-distance trip. Recently, in tandem with the development of information and communication technologies, new transportation modes such as electric scooters, and autonomous and demand-responsive personal rapid transit (PRT) have also arisen as a means of a short-distance trip. These modes are key to connecting the terminal points of mid/long-distance trips by subways or buses, and to link the first and last miles to origins and destinations [9, 10]. Thus, transportation modes for short-distance trips enhance seamless services, user satisfaction, and popularity of mass transit systems.

Transportation modes for short-distance trips are increasing in importance. However, traditional travel demand models have difficulty in explaining the impact of such trips [8]. Because current transportation systems allow free transfers and easy alighting and boarding, this limitation hinders reliable analysis of contemporary transportation techniques and policies. As most mode choice models have been developed for mid/long-distance trips, e.g., [11, 12], factors other than travel time and cost have been rationally neglected that have minor impacts on mode choice. However, the impact of other factors is significant as the traveler's time and cost is trivial in a short-distance trip. Examples include accessibility to terminals, discomfort of using stairs, adverse weather conditions, requirements of children/the elderly, and carrying of baggage. As previously mentioned, traditional mode choice approaches based the disaggregated behavioral models used in the four-step travel demand process are limited to fundamentally reflect these factors [8].

2. Scope and Objectives

The main objectives of this study are to identify and quantify factors affecting mode choice and, to provide fundamentals for the analysis of emerging transportation modes for the new demands of a short-distance trip. As a primary approach, a survey was conducted for inputs to identify the likely PRT choice considering the trip purpose, weather condition, and if passengers were accompanied by children. Here, the concept of PRT, one of desirable transportation modes, was used in this study as a competitive means of walking in a short-distance trip. To ascertain the objectives of this study, a literature review was conducted to identify a suitable methodology for the assessment. An analysis was then performed resulting in an updated interpretation.

3. Literature Review

This study is among initial attempts to understand factors that determine transportation modes in a short-distance trip and three points are of major interest in the literature review which include

(1) defining of the range of a short-distance trip;

(2) alternative modes considered for a short-distance trip; and

(3) explaining variables that determine the transportation modes for a particular short-distance trip.

3.1. Range Definition for the Short-Distance Trip. Travel distance is the most influential factor in determining a transportation mode. For example, the airplane is preferable for distances over 1,000 km, whereas walking, motorcycling, or cycling is a reasonable choice for distances less than 2 km. To develop mode choice models for a short-distance trip, the range of distance is therefore required to be defined for this study.

As shown in Table 1, although 5 km was frequently used as the maximum boundary for the short-distance trip [15, 16], various definitions of the short-distance trip have been suggested recently. That is, it is noted that a distance of less than 3 km was adopted in the most recent study [17] in order to respond to the need for the analysis of nonmotorized trips such as walking and bicycles.

3.2. Mode Alternatives for a Short-Distance Trip. Even for a short-distance trip, travelers frequently determine their travel from a specific set of available modes clustered by travel distance. For example, in Shanghai, China (see Figure 1), differences in average travel distance determined which mode was chosen for the short-distance trip. For example, walking was selected for an average travel distance of 1 km, while motorized modes (i.e., scooter and bus) were taken above an average travel distance of 2 km.

In addition to the preferred travel mode by distance, the average travel distance of all short-distance trip modes of 1.687 meters was used as a reference for the range of short-distance trips in this study.

To develop suitable mode choice models, possible alternative modes for a short-distance trip need to be explored which are shown in Table 2.

From the literature review, walking is the most prevalent mode for a short-distance trip (11 out of 11), followed by cycling (10 out of 11), individual vehicles (8 out of 11), and buses (7 out of 11). Interestingly, the competitor to nonmotorized modes (i.e., walking and cycling) is not public transportation such as bus, but rather, individual vehicles. This means that for a short-distance trip, the use of individual vehicles is common because of their superior access and comfort. In other words, travelers favor their personal vehicle as competitive for the first and last mile as it also satisfies their need for access and comfort.

Comparing the number of considered modes from the previous literature shows the range that is a maximum of 7, a minimum of 2, and an average of 4. Thus, in general, four separate mode choices were considered for the short-distance trip.

3.3. Explanatory Variables for Mode Choice in a Short-Distance Trip. Table 3 outlines the variables reviewed for mode choice models for a short-distance trip from the previous literature.

Table 1: A comparison of definitions used for a short-distance trip.

Authors	Used distance	Reasons
Kim and Ulfarsson [13]	Below 2.25 km	A short-distance was defined by the 95th percentile of walking distances on the basis of 12,900 trip data sets
Tran et al. [14]	Below 1 km	A distance under 1 km was regarded as a short-distance because of its competitiveness for walking rather than other modes, such as motorcycles and bicycles
Li et al. [15]	Below 5 km	A distance under 5 km was considered as a short-distance because of the majority, being over 40% of all trips, caused traffic congestion
Pu et al. [16]	Below 5 km	A distance under 5 km accounted for 90% of all short-distance trips for the elderly
Ferrer and Ruiz [17]	Below 3 km	A short-distance was defined as the distance taking 45 minutes on foot or less

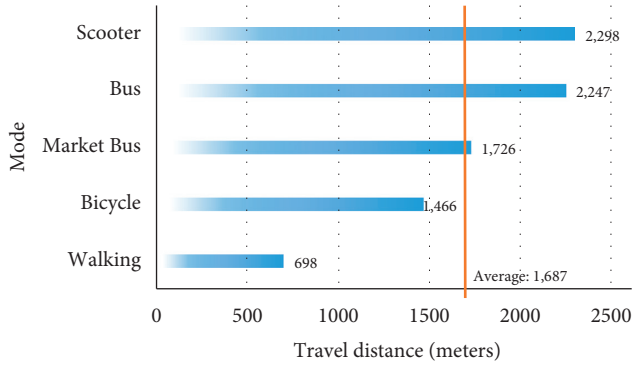


Figure 1: Average travel distances of each mode taken for the short-distance trip [16].

In total, 16 studies were reviewed, with 12 factors, which included socioeconomic factors such as gender, age, household income, and vehicle ownership that were mostly used as explanatory variables. Additionally, trip-related factors such as trip distance, walking comfort, travel time, trip purpose, traffic safety, and traffic condition/congestion were key factors applied in many studies. Some previous studies adopted weather (5 out of 16 studies) and crime (4 out of 16 studies) as explanatory variables. On average, 4.5 explanatory variables were used for 16 reviewed studies.

From this review, two features of the explanatory variables were identified:

(1) Multiple factors were considered for mode choice models

(2) Traditional factors (e.g., travel distance and time) were not always involved in mode choice studies

The mode choice for a short-distance trip may not follow the traditional approaches of mid/long-distance trips. Another reason is that in the short-distance trip, travel distance and time would likely have a minor impact on the choice of mode. Interestingly, these two reasons are correlated so that for short-distance trips other factors largely determine the mode. Travel distance and time are essentially non-determinants of the transportation mode.

3.4. Takeaways from the Literature Review. There are several key points from the literature review that can be extracted as inputs to analyze the mode choice of travelers for a short-distance trip:

(1) Many studies have considered up to 5 km as a short-distance trip. With this, considering the average speed of walking (4 km/h) [29, 30] and the average speed of PRT (40 km/h) [31], this study adopted 4 km as a short-distance trip.

(2) Among diverse transportation modes used in a short-distance trip, walking is the most common mode for a short-distance trip. Therefore, two modes being walking and PRT were selected as the alternatives for this study to develop the survey and mode choice models.

(3) Various factors were reviewed as explanatory variables from previous studies. These trip features (e.g., travel time, travel distance, and trip purpose), environmental features (e.g., walking comfort and weather), and individual characteristics (e.g., gender and age) most affected the mode choice for a short-distance trip. In addition to these factors, the trip purpose, weather condition, PRT fare, travel time, and accompaniment were selected as determinants for the mode choice model in the study.

4. Methodology

It is useful to elucidate the overall procedure described in Figure 2. The proposed methodology in this study comprises four parts which consider these objectives:

(1) Undertaking a pilot survey to test the predefined approach

(2) Completing a directional setup with preliminary results for the main survey

(3) Conducting the main survey

(4) Modeling and calibrating the main survey results

The first step was to examine the factors (i.e., explanatory variables) through the pilot survey which were identified from the literature review. Two factors were selected in this step being a weather factor representing traffic environments, and an accompanying children factor which could affect the mode choice in a leisure trip. Note that the major objectives of the pilot survey were checking the applicability of selected factors in this study, and providing preliminary results to be used to set up the main survey.

In considering the weather factor, two different regions comprising large-sized residential areas with parks being

Table 2: Comparison of transportation modes considered regarding a short-distance trip.

Authors	Transportation modes							Number of considered modes
	Walking	Bicycle	Scooter	Taxi	Bus	Subway	Individual vehicle	
Mackett [18]	○	○	○	○	○	○	○	7
Olszewski and Wibowo [19]	○	○			○	○	○	5
Müller et al. [20]	○	○	○		○		○	5
Kim and Ulfarsson [13]	○	○			○		○	4
Walton and Sunseri [21]	○						○	2
Sidharthan et al. [22]	○	○						2
Halldórsdóttir et al. [23]	○	○			○		○	4
Tran et al. [14]	○	○	○					3
Li et al. [15]	○	○					○	3
Pu et al. [16]	○	○	○		○			4
Prato et al. [24]	○	○			○	○	○	5
Frequency	11	10	4	1	7	3	8	—

Table 3: Factors considered in previous studies of a short-distance trip.

	Gender	Age	Household income	Vehicle ownership	Trip distance	Walking comfort (slope/ stairway)	Travel time	Trip purpose	Traffic safety	Traffic condition/ congestion	Weather	Crime
Mackett [18]	○	○			○		○	○				
Olszewski and Wibowo [19]					○	○			○	○		
Müller et al. [20]				○	○						○	
Kim and Ulfarsson [13]	○	○			○		○					
Walton and Sunseri [21]	○				○	○	○				○	○
Sidharthan et al. [22]			○	○		○			○	○	○	
Halldórsdóttir et al. [23]	○			○		○	○	○			○	
Lee et al. [25]									○	○		
Marquet and Miralles-Guasch [26]	○	○	○					○				
Rybarczyk and Gallagher [27]					○				○	○		○
Tran et al. [14]	○	○	○	○				○	○	○		○
Li et al. [15]	○	○	○	○								
Pu et al. [16]		○	○	○			○	○				
Ferrer and Ruiz [17]						○						○
Guang and Reid [28]			○			○						
Prato et al. [24]	○	○	○			○	○	○			○	
Frequency	8	7	7	6	6	7	6	6	5	5	5	4

Gwanggyo and Ilsan in Gyeonggi Province, Republic of Korea, were selected. During the pilot survey, autonomous and demand-responsive PRT had been introduced as a new and more practical public transit mode to the respondents for the pilot survey. This PRT is an unmanned mode with a maximum of 6 passengers operating on a designated or virtually marked road where a user can access a ride anytime.

The second step to design the main survey was based on preliminary results from the pilot survey. Basically, two main purposes being mandatory and nonmandatory were considered with two factors, which were tested in the pilot survey, using two different fare schemes. The type of scenarios was accordingly set up with factors that needed to be considered for the model calibration. The next step was to conduct the main survey comprising 60 questions, followed by model calibration and logistic regression. Despite the availability of emerging techniques, such as machine learning methods, the logistic regression model was selected as the tool to analyze the behavior of travelers for short-distance trips.

This decision has two significant reasons:

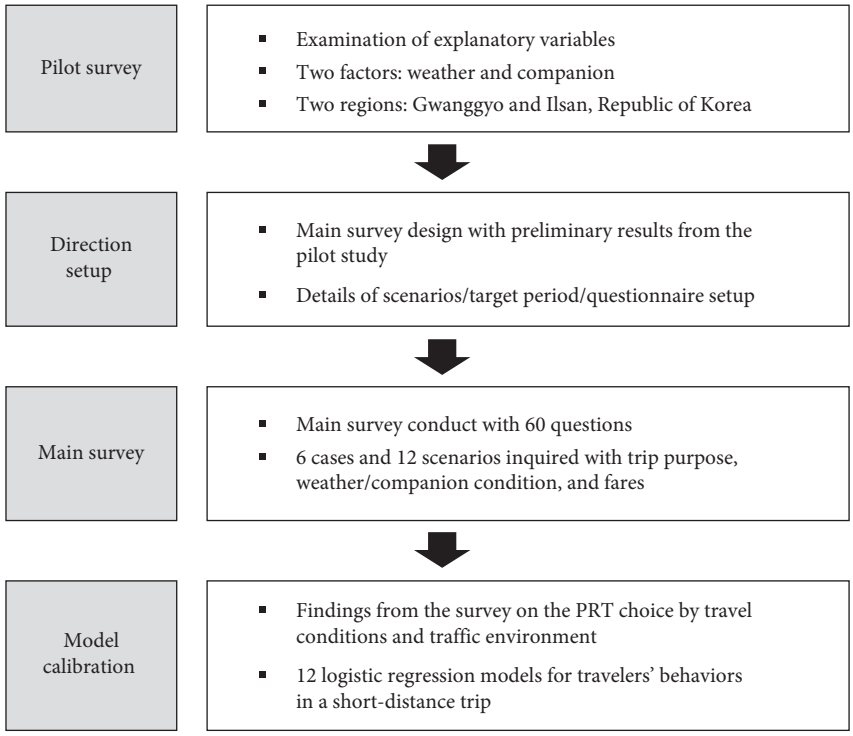


Figure 2: Overall procedure of the methodology.

(1) The logistic regression models have been regarded as a standard method for describing the relationship between a response variable and explanatory variables over the last decade [32]

(2) Machine learning techniques are gaining in popularity, but can become inefficient with a relatively small size datasets as it requires massive datasets to train the models [33]

5. Analysis and Results

5.1. Pilot Survey: Preliminary Findings for Travelers' Behaviors in Short-Distance Trips. From the pilot survey 1,030 responses were received by surveyors from the two regions (515 responses each). However, only 1,006 responses were valid (effectively 97.7%). The basic question had two choices being good versus bad weather (snow, rain, and/or cold temperatures). In cases where the travel time when walking was within 20 minutes, the intervals for the questionnaire were 5, 10, 15, and 20 minutes, with the mode choice being either walking or PRT. The flat fare for PRT was assumed as 500 Korean won (KRW) per person, regardless of the distance traveled. Figure 3(a) describes the result of the pilot survey for the weather condition.

Another location (Uiwang in Gyeonggi Province, Republic of Korea) being a lake park region where family leisure activities frequently occur was chosen to examine the effect of the travelers' companions. A total of 501 valid responses were obtained from nearby the Uiwang subway station and rail bike station. The PRT option was presented in terms of an increase in travel cost from 0 to 1,500 (KRW per person). In this pilot survey, only travel cost was given as a determinant in choosing PRT because the distance between the Uiwang subway station and the rail bike station was fixed at 1 km, and the actual cost for accompanying children had not been confirmed. Figure 3(b) shows the results of this pilot survey.

The pilot survey revealed that travelers preferred PRT when the weather condition was not good, compared to when the weather was obviously good. The relationship between travel time and the weather condition was also observed. As travel time increased, the extent of impacts from the weather condition decreased. If travel time was relatively long, it became a key determinant of choosing PRT regardless of the weather condition. In contrast, when the travel time was relatively short, the weather condition became the crucial factor in selecting PRT. In terms of accompanying children, when the PRT fare was zero, most of the interviewees would choose PRT, but the probability of PRT choice decreased in two distinct cases. This was when the PRT fare increased, or when traveling without children. With a travel cost up to 500 KRW per person, the probabilities for PRT choice showed little difference, regardless of accompanying children.

5.2. Direction Setup for Main Survey: Case Development and Calibration Process. From the pilot survey, 6 cases (see Figure 4) were generated to develop main questionnaires in the survey and eventually to the development of mode choice models.

Six cases were originally generated, with each case having two different fare schemes: 500 and 1,000 KRW per person. Thus, in total, 12 scenarios were considered in this study. Because a mandatory trip rarely occurs with children, the

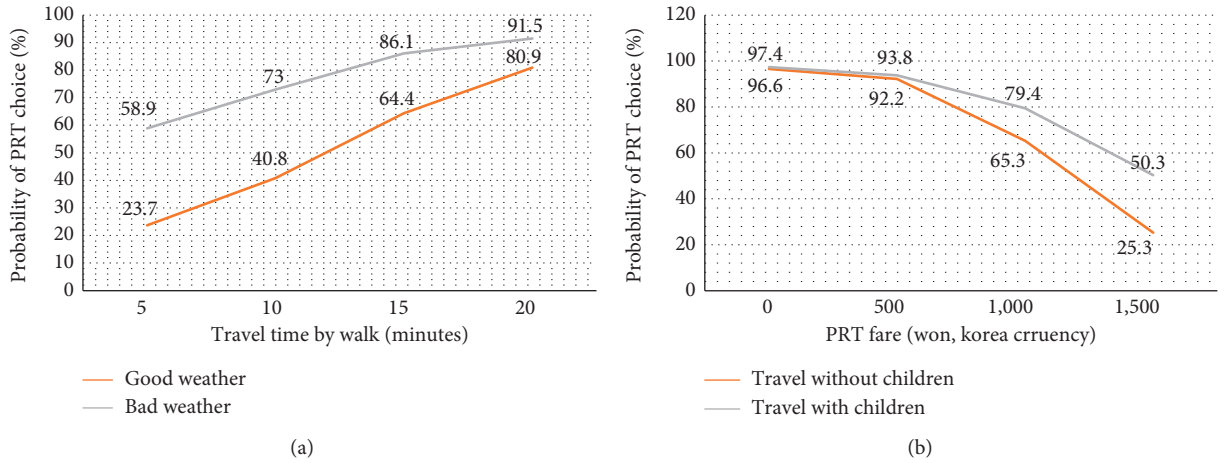


Figure 3: Pilot survey results of weather and companion conditions: (a) weather condition; (b) companion condition (travel with children).

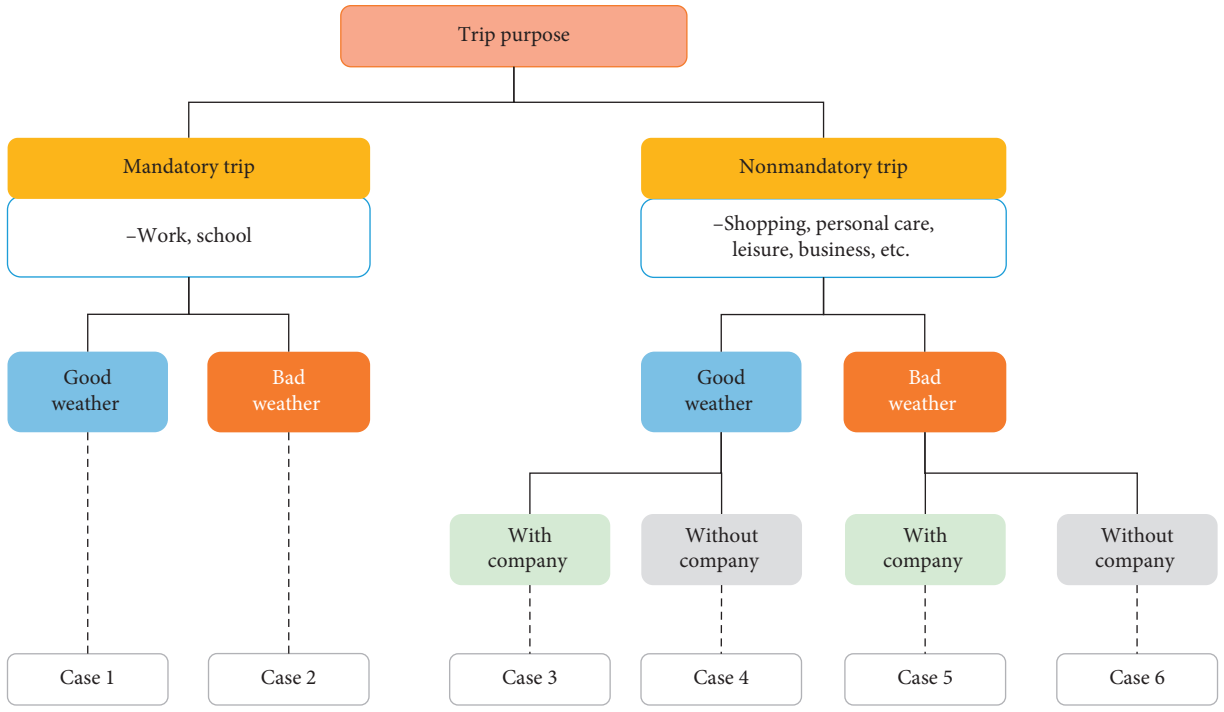


Figure 4: Case tree for mode choice models.

conditions for accompanying children were included only in a nonmandatory trip (see Table 4).

A four-step approach was taken (see Figure 5), of which questionnaires were developed from steps 1 to 3 and step 4 was about how to calibrate the mode choice models in this study:

(1) A total distance of 4 km was divided into five intervals of 800 meters per each (defined as a short-distance trip in this study)

(2) Travel distance was selected randomly from each interval

(3) Travel times with walking and PRT, respectively, were estimated from each randomly selected travel distance

(4) At every 100 meters, the probabilities of PRT choice were fitted into the diversion curve

As previously mentioned, 60 questions were posed to each interviewee in the main survey, consisting of six scenarios with two fare options (500 and 1,000 KRW), and five intervals separated by 800 meters (out of 4 km).

Table 4: List of selected cases for mode choice models.

Case number	Trip purpose	Weather condition	Companion condition
Case 1	Mandatory	Good	—
Case 2		Bad	—
Case 3	Nonmandatory	Good	No
Case 4		Bad	No
Case 5		Good	Yes
Case 6		Bad	Yes

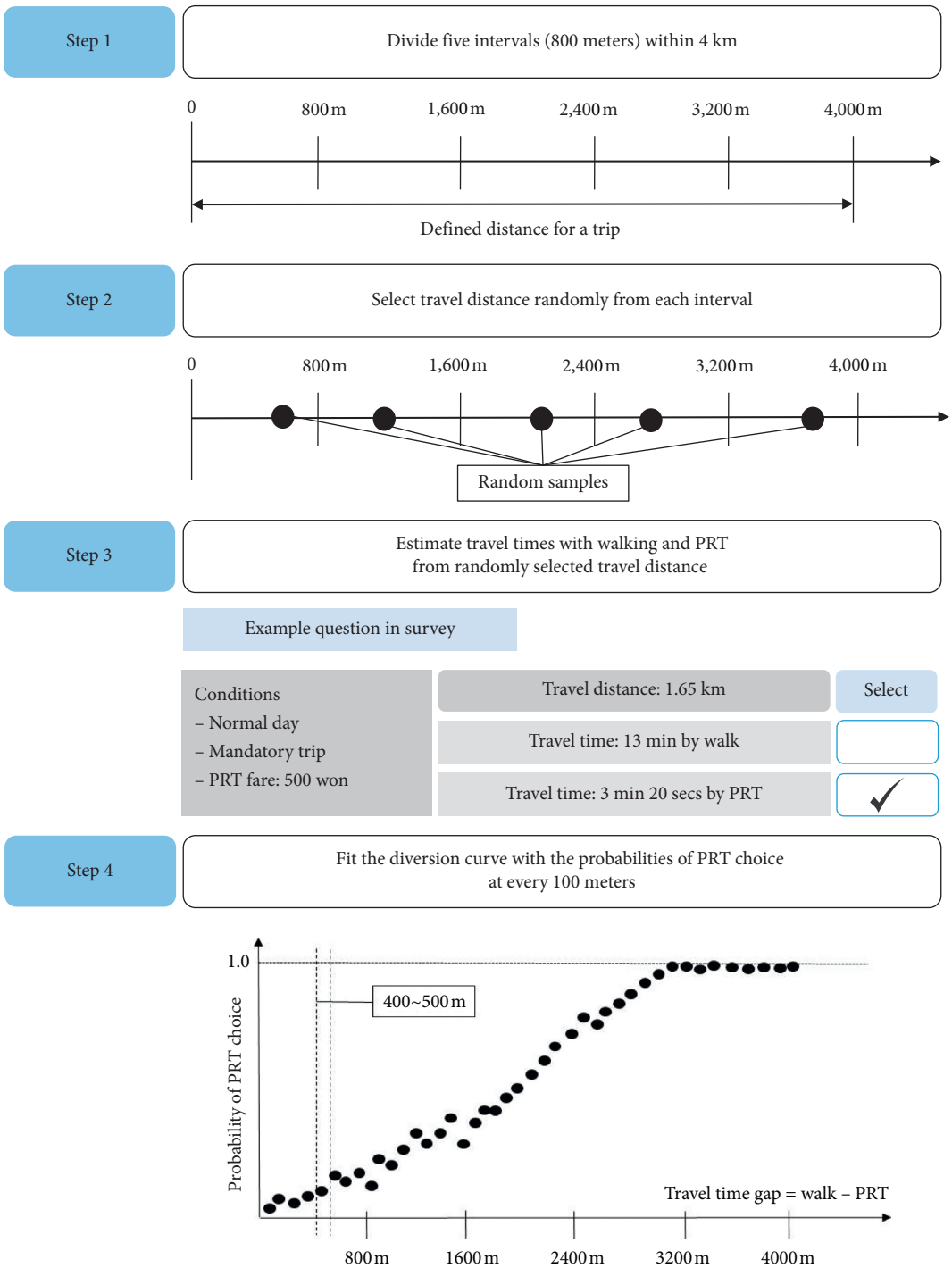


Figure 5: Process of developing mode choice models.

Table 5: Interviewees' characteristics.

Categories	Frequency (respondent, %)
Gender	
Male	609 (50.0)
Female	608 (50.0)
Age	
0–19	128 (10.5)
20–29	240 (19.7)
30–39	239 (19.6)
40–49	241 (19.8)
50–59	240 (19.7)
60+	129 (10.6)
Occupation	
Employee (full-time)	633 (52.1)
Self-employed individual	107 (8.8)
Sstudent	234 (19.2)
Housewife	157 (12.9)
Unemployed person	54 (4.4)
Others	32 (2.6)
Number of family members	
1	99 (8.1)
2	164 (13.5)
3	334 (27.4)
4	491 (40.3)
5	105 (8.6)
6+	24 (2.1)
Family income (unit: thousand KRW/month)	
0–999	61 (5.0)
1,000–1,999	117 (9.6)
2,000–2,999	205 (16.8)
3,000–3,999	236 (19.4)
4,000–4,999	200 (16.4)
5,000–5,999	186 (15.3)
6,000–6,999	85 (7.1)
7,000+	127 (10.4)

5.3. Main Survey: Results for Travelers' Behaviors in Short-Distance Trips

5.3.1. Description and Demographic Statistics. The main survey was conducted through the Internet to (1) obtain a significant number of responses for all types of traveler's groups and (2) have more balanced responses regardless of the locational features. 1,217 respondents were interviewed with gender parity for four days (from 30th of September to 3rd of October 2016). Table 5 presents the details of the interviewees' responses.

5.3.2. Description of Logistic Regression Models. On the basis of the main survey, there are preliminary observations regarding the impacts of travel conditions or the traffic environment on choosing PRT, in addition to the calibration of mode choice models, as follows:

(1) In the mandatory trip, (i) PRT was preferred during inclement weather conditions regardless of PRT fares, and (ii) when the PRT fare was lower, travelers obviously preferred to use PRT even during good weather conditions (see Figure 6). To be specific, both scenarios with travel costs of 500 KRW and 1,000 KRW during bad weather (in case 2) showed higher probabilities than case 1 (good weather conditions). However, in case 1 (good weather), a scenario with 500 KRW showed a higher probability than a scenario with 1,000 KRW (see a horizontal axis of 10-minute difference of travel time). If the weather was good, a cheaper PRT fare increased the probability of PRT choice by around 30% when the travel time was different by 10 minutes.

(2) In the nonmandatory trip without company scenario, the same result was observed. That is, (i) PRT fares did not significantly impact the choice of PRT in bad weather and (ii) a lower fare affected PRT choice with good weather conditions (see Figure 7). In case 4, both scenarios with 500 KRW and 1,000 KRW with bad weather had higher probabilities than case 3 (good weather conditions), whereas with 500 KRW, a higher probability was observed in case 3 compared to the one with 1,000 KRW (being an approximate 15% gap in 18-minute difference of travel time between 500 KRW and 1,000 KRW).

(3) When comparing a mandatory trip against a nonmandatory trip without company with a 500 KRW fare, the nonmandatory trip without company was more sensitive to weather conditions as shown in the left graphs in Figures 6 and 7. This sensitivity became similar with a 1,000 KRW fare (see the right graphs in Figures 6 and 7). That is, the decision for PRT was obviously affected by weather conditions in both a mandatory trip and nonmandatory trip without company.

(4) In the nonmandatory trip with company, (i) the weather condition made a difference to the probability of a PRT choice regardless of fares and (ii) the impact of differing fares showed a minor difference under the same weather conditions both in cases 5 and 6 (see Figure 8). Specifically, with company, having bad weather conditions was a valid factor to determine the choice of PRT. As indicated in Figure 8, a gap of around 30% of 9-minute difference of travel time between bad and good weather conditions (in cases 5 and 6) was found in both fares. As shown in a horizontal axis of 15-minute difference of travel time with bad weather, similar probabilities were observed. This was regardless of the fare difference which led to the conclusion that trips with company were not affected by fares, but rather, by weather conditions.

(5) Under good weather conditions, (i) differing fares influenced the choice of PRT regardless of trip purposes and (ii) the trip purpose affected PRT choice, particularly with a lower fare (see Figure 9). That is, with the same trip purpose, the probability of choosing PRT with 500 KRW was generally higher than the one with 1,000 KRW. Looking at the horizontal axis of 15-minute difference of travel

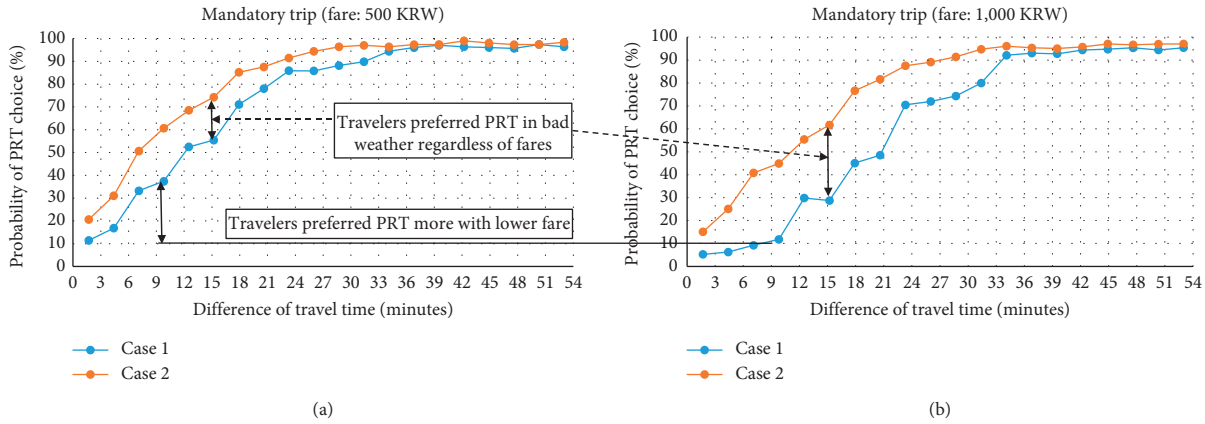


Figure 6: Impacts by weather and fare on PRT choice in the mandatory trip.

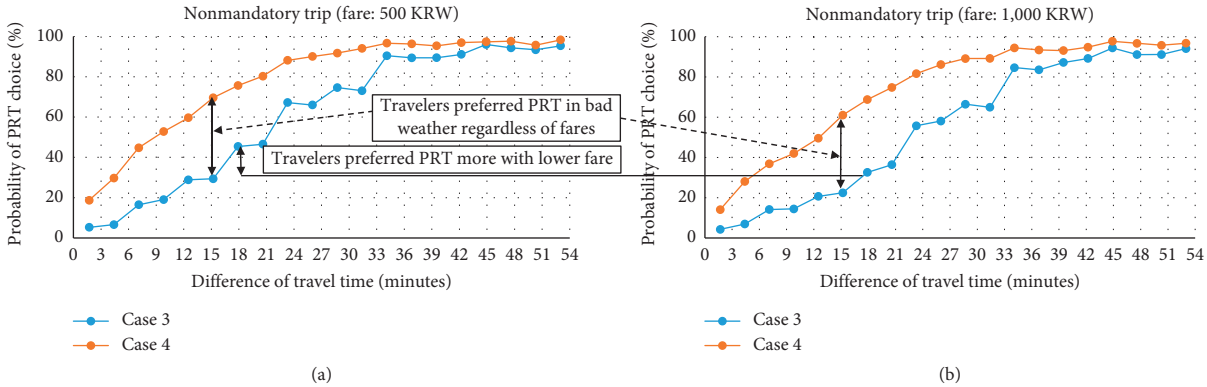


Figure 7: Impacts by weather and fare on PRT choice in the nonmandatory trip without company.

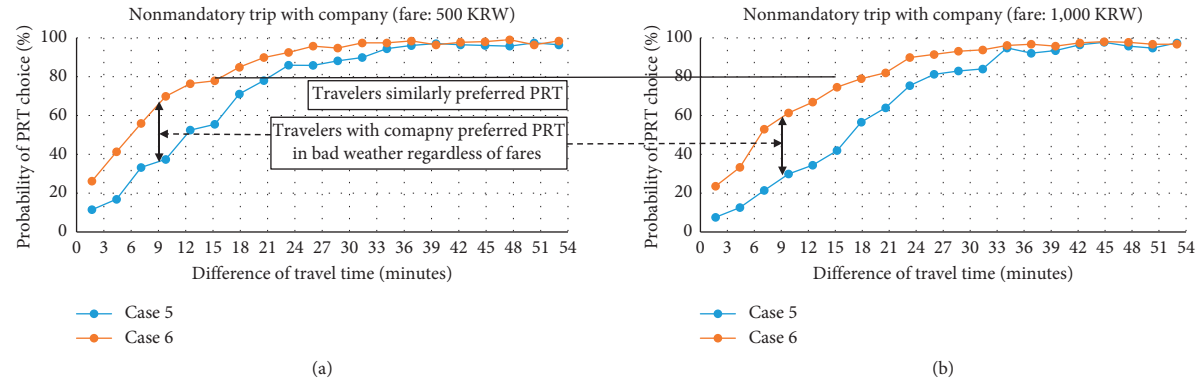


Figure 8: Impacts by weather and fare on PRT choice in the nonmandatory trip with company.

time, the probability of PRT choice in a mandatory trip with 500 KRW was higher by around 25% than the one with 1,000 KRW. Also, with different trip purpose, the probability of PRT choice for the mandatory trip was always higher than for nonmandatory trips owing to the higher value placed on time by those who need to travel obligatorily.

(6) Under bad weather conditions, (i) the fare difference was not a strong factor to determine the choice of PRT regardless of trip purposes and (ii) the trip purpose did not affect the choice of PRT strongly (see Figure 10). With the same trip purpose, a similarity was found in terms of the probabilities of choosing PRT from the two fares. Likewise, with the same fare condition, the probabilities of PRT choice became

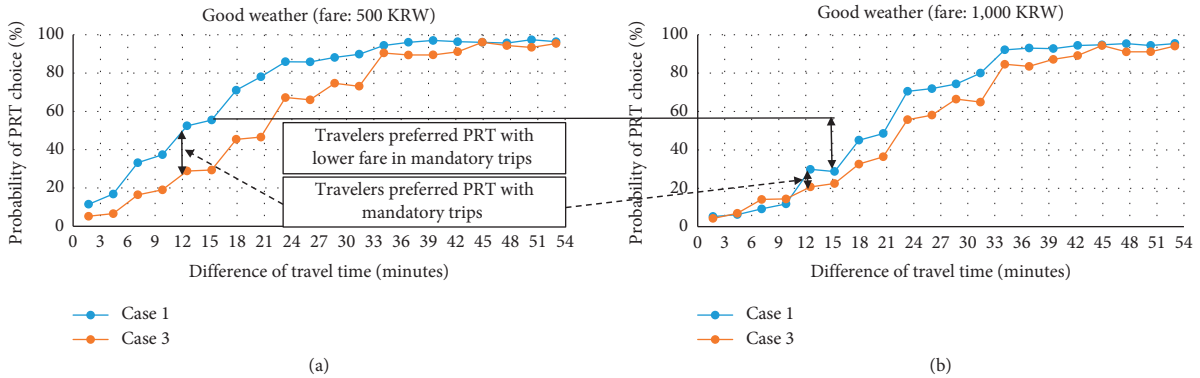


Figure 9: Impacts by fare and trip purpose on PRT choice under good weather conditions.

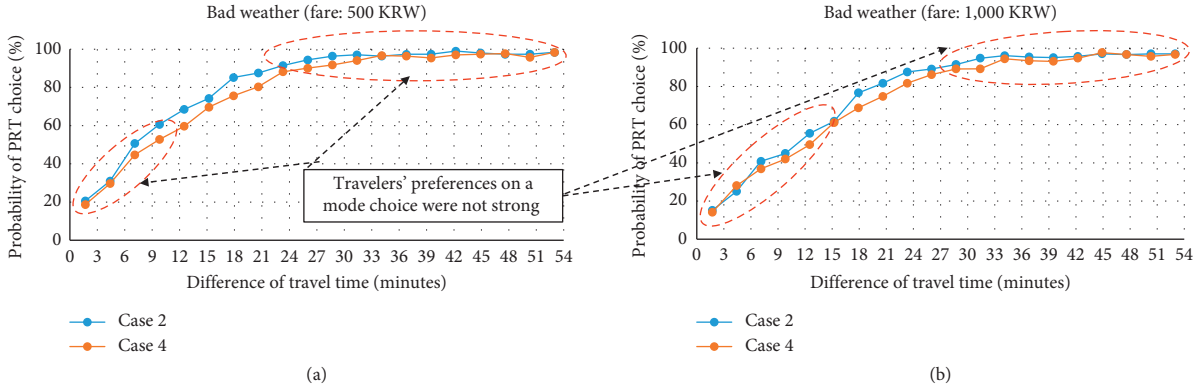


Figure 10: Impacts by fare and trip purpose on PRT choice under bad weather conditions.

Table 6: Logistic regression models for 12 scenarios.

Cases	Scenarios	Trip purpose	Weather condition	Companion condition	PRT fare (KRW)	Functions
Case 1	Scenario 1	Mandatory	Good	—	500	$p = \frac{\exp(-1.651+0.126\times\text{difference of travel time})}{1+\exp(-1.651+0.126\times\text{difference of travel time})}$
	Scenario 2				1,000	$p = \frac{\exp(-2.831+0.137\times\text{difference of travel time})}{1+\exp(-2.831+0.137\times\text{difference of travel time})}$
Case 2	Scenario 3		Bad	—	500	$p = \frac{\exp(-1.034+0.133\times\text{difference of travel time})}{1+\exp(-1.034+0.133\times\text{difference of travel time})}$
	Scenario 4				1,000	$p = \frac{\exp(-1.344+0.122\times\text{difference of travel time})}{1+\exp(-1.344+0.122\times\text{difference of travel time})}$
Case 3	Scenario 5	Nonmandatory	Good	No	500	$p = \frac{\exp(-2.548+0.122\times\text{difference of travel time})}{1+\exp(-2.548+0.122\times\text{difference of travel time})}$
	Scenario 6				1,000	$p = \frac{\exp(-2.825+0.117\times\text{difference of travel time})}{1+\exp(-2.825+0.117\times\text{difference of travel time})}$
Case 4	Scenario 7		Bad	No	500	$p = \frac{\exp(-1.139+0.119\times\text{difference of travel time})}{1+\exp(-1.139+0.119\times\text{difference of travel time})}$
	Scenario 8				1,000	$p = \frac{\exp(-1.365+0.111\times\text{difference of travel time})}{1+\exp(-1.365+0.111\times\text{difference of travel time})}$
Case 5	Scenario 9		Good	Yes	500	$p = \frac{\exp(-2.001+0.134\times\text{difference of travel time})}{1+\exp(-2.001+0.134\times\text{difference of travel time})}$
	Scenario 10				1,000	$p = \frac{\exp(-2.182+0.128\times\text{difference of travel time})}{1+\exp(-2.182+0.128\times\text{difference of travel time})}$
Case 6	Scenario 11		Bad	Yes	500	$p = \frac{\exp(-0.657+0.122\times\text{difference of travel time})}{1+\exp(-0.657+0.122\times\text{difference of travel time})}$
	Scenario 12				1,000	$p = \frac{\exp(-0.826+0.112\times\text{difference of travel times})}{1+\exp(-0.826+0.112\times\text{difference of travel time})}$

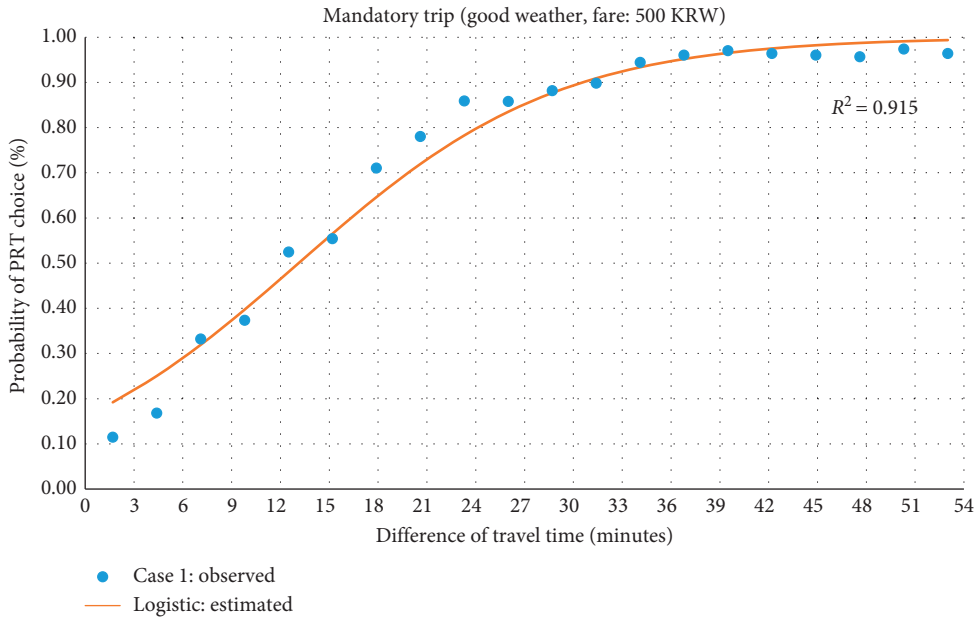


Figure 11: Logistic regression curve of the mandatory trip with 500 KRW under good weather (scenario 1).

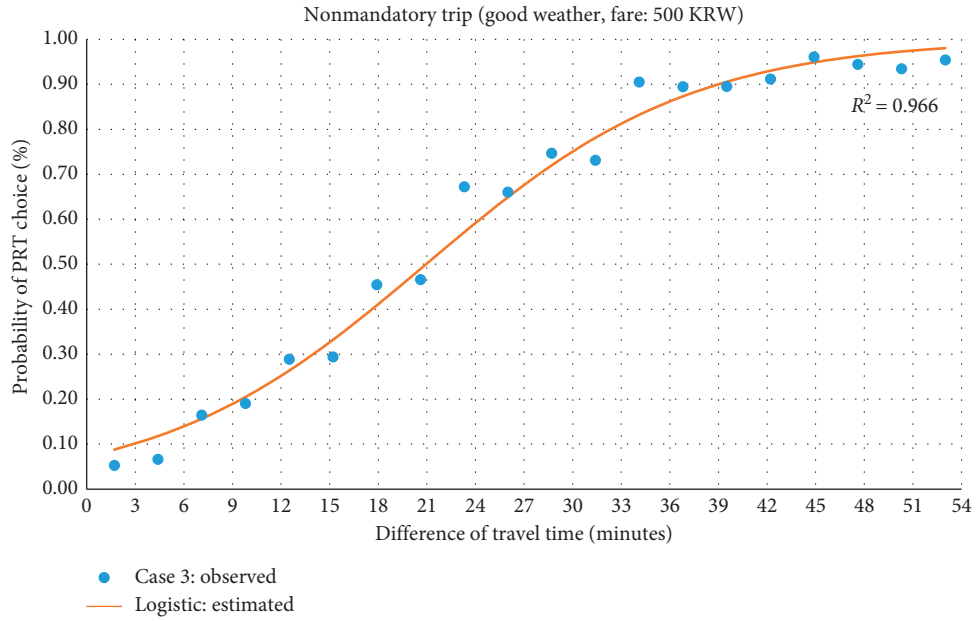


Figure 12: Logistic regression curve of the nonmandatory trip with 500 KRW under good weather (scenario 5).

similar on both mandatory and nonmandatory trips without company.

5.4. Model Calibration: Logistic Regression Models for Travelers' Behaviors in Short-Distance Trips. Based on the results from the main survey, the probability function for PRT choice was calibrated. After testing a variety of functions, a logistic regression curve was determined because of its high goodness-of-fit. In total, 12 logistic regression curves are presented in Table 6.

Among the 12 scenarios, two examples have been presented based on the observed data from the main survey and calculated values from logistic regression functions 1 and 5. The R-squared measures per each model showed high goodness-of-fits which were 0.915 (Figure 11) and 0.966 (Figure 12), respectively.

6. Conclusions

There are emerging demands for connectivity of the first and last mile of travel. According to the advancement of ICT and

growing interest in environmental issues, special transportation modes in short-distance trips that satisfy an individual's various needs for convenience and comfort have been identified. Previous studies have mostly dealt with mid/long-distance trips due in part to the lack of suitable models or usable data for the short-distance trip. Some studies that have focused on short-distance trips conducted a simple analysis and were limited in scope.

The main features of the current study are

(1) to propose a suitable approach that can understand the features determining mode choice of transportation for a short-distance trip;

(2) to quantify the impact of various factors, including the trip purpose, weather condition, fare, travel time, and accompanying children, which affect the mode choice for a short-distance trip;

(3) to reaffirm the applicability of the logistic regression model for the analysis of traveler's behavior for a short-distance trip; and

(4) to provide the groundwork for the analysis of active transportation modes, such as electric scooters, which are expected to grow in response to new demands (including Mobility-as-a-Service) for the short-distance trip.

In conclusion, the trip purpose and weather condition impacted as factors on the mode choice being consistent in part with Ashkrof et al. [34]. For a mandatory workplace or school trip under similar travel times, travelers preferred PRT. A preference for PRT was also observed in bad weather conditions, regardless of the type of trip purposes and PRT fares. When traveling with company, the weather condition became a key factor again in choosing PRT, while the variation of fares did not comparatively have an impact on mode choice. In summary, when the weather was not good, PRT fare difference and trip purposes were not a strong factor when deciding to choose PRT. However, during good weather, other factors contributed to the choice of PRT. These included differing fares regardless of trip purposes, and the trip purpose (i.e., mandatory trips) affected PRT choice, particularly with a lower fare. Given that fares become an effective factor under good weather conditions, this finding is also similar to the one from Liu et al. [35] who concluded that cost significantly influenced mode choice for short-distance trips.

The weather has a significant effect on our lives. For example, the average number of days with precipitation (including snowfall and rainfall) is 103.5 days per year in the Republic of Korea [36]. As such, the extent of the impact of weather conditions on PRT choice is noteworthy. This might require further careful and thorough investigation because of the recent spotlight on diverse environmental factors, such as global warming, climate change, yellow dust, and fine dust.

In comparison to previous disaggregated behavioral models (e.g., logit models), this study employed a logistic regression model to calibrate the outputs from the survey. Note that the short-distance trip is less affected by travel deterrents such as travel time, distance, and fares, when compared with the long-distance trip. However, various other factors, such as weather, security, and slope, determine mode choice for the short-distance trip. Because a logit model requires many explanatory variables, the process of calibration becomes complicated, which in turn defines the model's explanatory significance as low.

To overcome this problem, an alternative model considering diverse variables for mode choice was developed by using a logistic regression model, which replicated actual mode choice from the survey. Nevertheless, as the logistic regression model has a limitation when expressing more than three multiple modes, alternative approaches need to be considered. These include machine learning techniques that can incorporate a number of factors and modes together. In this aspect, decision analysis techniques, such as the conjoint method, Naïve Bayes algorithm decision tree algorithm, or random forest algorithm, can be options used for mode choice modeling with various explanatory factors within future studies.

Data Availability

All data used in this study are available upon request to the author (Hyunmyung Kim, khchlsy@gmail.com).

Disclosure

Changju Lee is currently at Transport Division, United Nations Economic and Social Commission for Asia and the Pacific, Bangkok 10200, Thailand. The views expressed herein are those of the authors and do not necessarily reflect the views of the United Nations.

Conflicts of Interest

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

Acknowledgments

This study was supported by a grant from R&D program of the Korea Railroad Research Institute, Republic of Korea. In addition, the improvement of this study was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT, and Future Planning (no. 2016R1A2B2012722).

References

[1] C. A. Klöckner and T. Friedrichsmeier, "A multi-level approach to travel mode choice-how person characteristics and situation specific aspects determine car use in a student sample," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 4, pp. 261–277, 2011.

[2] A. Reichert and H. R. Christian, "Mode use in long-distance travel," *The Journal of Transportation and Land Use*, vol. 8, no. 2, pp. 87–105, 2015.

[3] A. L. L. Olsson, *Factors that Influence Choice of Travel Mode in Major Urban Areas: The Attractiveness of Park & Ride*,

Infrastruktur, Stockholm, Sweden, 2003, http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-1627.

[4] Minal and C. R. Sekhar, "Mode choice analysis: the data, the models and future ahead," *International Journal for Traffic and Transport Engineering*, vol. 4, no. 3, pp. 269–285, 2014.

[5] M. Spies, "Distance between home and workplace as a factor for job satisfaction in the North-West Russian oil industry," *Fennia*, vol. 184, no. 2, pp. 133–149, 2006.

[6] R. Ewing, W. Schroeer, and W. Greene, "School location and student travel: analysis of factor affecting mode choice," *Journal of the Transportation Research Board*, vol. 1895, no. 1, pp. 55–63, 2004.

[7] United Nations Framework Convention on Climate Change, *The Paris Agreement*, United Nations Framework Convention on Climate Change, New York, NY, USA, 2019, https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

[8] L. Ming, H. Zhu, L. Xia, and L. Lei, "Intercity travel demand analysis model," *Advances in Mechanical Engineering*, vol. 6, Article ID 108180, 2015.

[9] Institute for Transportation and Development Policy, "E-scooters could be a last-mile solution for everyone," *Transport Matters*, Institute for Transportation and Development Policy, New York, NY, USA, 2019, https://www.itdp.org/2018/12/14/e-scooters-last-mile-solution/.

[10] A. Singh, *Uber Electric Scooters for Affordable Last-Mile Connectivity: What They Are and Will They Come to India*, Financial Express, Noida, India, 2019, https://www.financialexpress.com/auto/bike-news/uber-electric-scooters-for-affordable-last-mile-connectivity-what-they-are-and-will-they-come-to-india/1530478/.

[11] R. Moeckel, R. Fussell, and R. Donnelly, "Mode choice modeling for long-distance travel," *Transportation Letters*, vol. 7, no. 1, pp. 35–46, 2015.

[12] C. Llorca, J. Molloy, J. Ji, and R. Moeckel, "Estimation of a long-distance travel demand model using trip surveys, location-based big data, and trip planning services," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 47, pp. 103–113, 2018.

[13] S. Kim and G. F. Ulfarsson, "Curbing automobile use for sustainable transportation: analysis of mode choice on short home-based trips," *Transportation*, vol. 35, no. 6, pp. 723–737, 2008.

[14] M. T. Tran, M. Chikaraishi, Q. H. Pham, J. Zhang, and A. Fujiwara, "Effects of perceived neighborhood walkability on mode choice of short-distance trips in Hanoi city," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 11, pp. 1328–1345, 2015.

[15] M. Li, G. Song, Y. Cheng, and L. Yu, "Identification of prior factors influencing the mode choice of short distance travel," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 795176, 9 pages, 2015.

[16] X. Pu, W. Wang, and Y. Wu, "Short-distance Trip Mode Choice Behavior of the Elderly," in *Proceedings of the 15th COTA International Conference Of Transportation*, pp. 3854–3865, Beijing, China, July 2015.

[17] S. Ferrer and T. Ruiz, "The impact of the built environment on the decision to walk for short trips: evidence from two Spanish cities," *Transport Policy*, vol. 67, pp. 111–120, 2018.

[18] R. L. Mackett, "Why do people use their cars for short trips?" *Transportation*, vol. 30, no. 3, pp. 329–349, 2003.

[19] P. Olszewski and S. Wibowo, "Using equivalent walking distance to assess pedestrian accessibility to transit stations in Singapore," *Journal of the Transportation Research Board*, vol. 1927, pp. 38–45, 2005.

[20] S. Müller, S. Tscharaktschiew, and K. Haase, "Travel-to-school mode choice modelling and patterns of school choice in urban areas," *Journal of Transport Geography*, vol. 16, no. 5, pp. 342–357, 2008.

[21] D. Walton and S. Sunseri, "Factors influencing the decision to drive or walk short distances to public transport facilities," *International Journal of Sustainable Transportation*, vol. 4, no. 4, pp. 212–226, 2010.

[22] R. Sidharthan, C. R. Bhat, R. M. Pendyala, and K. G. Goulias, *A Model of Children's School Travel Mode Choice Behavior Accounting for Spatial and Social Interaction Effects*, The University of Texas at Austin, Austin, TX, USA, 2010.

[23] K. Halldórsdóttir, L. Christensen, T. C. Jensen, and C. G. Prato, "Modelling mode choice in short trips-shifting from car to bicycle," in *Proceedings of the European Transport Conference*, Glasgow, UK, 2011.

[24] C. G. Prato, K. Halldórsdóttir, and O. A. Nielsen, "Latent lifestyle and mode choice decisions when travelling short distances," *Transportation*, vol. 44, no. 6, pp. 1343–1363, 2017.

[25] C. Lee, X. Zhu, J. Yoon, and J. W. Varni, "Beyond distance: children's school travel mode choice," *Annals of Behavioral Medicine*, vol. 45, no. 1, pp. 55–67, 2013.

[26] O. Marquet and C. Miralles-Guasch, "Walking short distances. The socioeconomic drivers for the use of proximity in everyday mobility in Barcelona," *Transportation Research Part A: Policy and Practice*, vol. 70, pp. 210–222, 2014.

[27] G. Rybarczyk and L. Gallagher, "Measuring the potential for bicycling and walking at a metropolitan commuter university," *Journal of Transport Geography*, vol. 39, pp. 1–10, 2014.

[28] T. Guang and E. A. Reid, "Walk trip generation model for Portland, OR," *Transportation Research Part D*, vol. 52, pp. 340–353, 2017.

[29] G. Chiaranda, E. Bernardi, L. Codecà et al., "Treadmill walking speed and survival prediction in men with cardiovascular disease: a 10-year follow-up study," *BMJ Open*, vol. 3, no. 10, 2013.

[30] M.-C. Chiu, H.-C. Wu, and L.-Y. Chang, "Gait speed and gender effects on center of pressure progression during normal walking," *Gait & Posture*, vol. 37, no. 1, pp. 43–48, 2013.

[31] D. Jeffery, *Guidelines for Implementers of Personal Rapid Transit (PRT)*, European Commission, Brussels, Belgium, 2010.

[32] J. Fang, "Why logistic regression analyses are more reliable than multiple regression analyses," *Journal of Business and Economics*, vol. 4, no. 7, pp. 620–633, 2013.

[33] P. Gulhane and B. V. Chaudhari, "Study of machine learning and deep learning of various factors," *International Journal of Research in Computer & Information*, vol. 4, no. 3, pp. 16–21, 2019.

[34] P. Ashkrof, G. Homem de Almeida Correia, O. Cats, and B. van Arem, "Impact of automated vehicles on travel mode preference for different trip purposes and distances," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 5, pp. 607–616, 2019.

[35] Y. Liu, J. Chen, W. Wu, and J. Ye, "Typical combined travel mode choice utility model in multimodal transportation network," *Sustainability*, vol. 11, no. 2, p. 549, 2019.

[36] Korea Meteorological Administration, *Average Number of Days with Precipitation*, Korea Meteorological Administration, Seoul, South Korea, 2017, https://data.kma.go.kr/stcs/grnd/grndRnDay.do?pgmNo=156.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 9020574, 15 pages
https://doi.org/10.1155/2020/9020574

Research Article

Effects of Link Capacity Reductions on the Reliability of an Urban Rail Transit Network

Jie Liu[1,2,3] **Paul M. Schonfeld**[4] **Yong Yin**[1,2,3] **and Qiyuan Peng**[1,2,3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 611756, China*
[3]*National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China*
[4]*Department of Civil and Environmental Engineering, University of Maryland, College Park 20742, MD, USA*

Correspondence should be addressed to Yong Yin; yinyong@swjtu.edu.cn

Received 7 February 2020; Revised 11 September 2020; Accepted 5 October 2020; Published 29 October 2020

Academic Editor: Prakash Ranjitkar

Link capacity reductions, which occur often, degrade the service quality and performance of urban rail transit (URT) networks. To measure the reliability of a URT network when link capacity reductions occur in a given time period, the passengers' generalized travel cost (GTC) is computed and passengers are divided into three categories. The GTC considers here the crowding in trains, seat availability, and perceived travel time. Passengers whose relative increase in GTC on a URT is below or above a preset threshold belong to category I or II, respectively, while passengers who cannot travel on the URT due to insufficient capacities on their paths belong to category III. Passenger trips in categories I are acceptable since their GTC increases only slightly with link capacity reductions. The fraction of acceptable trip (FAT) and total GTC increase ratio (TGCR) in a given time period are defined here as the network's reliability and unreliability metrics, respectively. The ratio of affected passenger trip (RAPT) is proposed to identify each line's most critical links. The reliability and unreliability metrics of Wuhan's URT network during evening peak hours are computed when the capacities of the most critical link or multiple most critical links are reduced. The results show that the proposed RAPT indicator is effective in identifying the most critical links that greatly affect the reliability and performance of a URT network. For capacity reductions on a line's most critical link, the proposed method can determine the capacity reduction ratio corresponding to network's high FAT and low TGCR as well as the priorities of lines needing emergency measures to maintain high network reliability and performance. For capacity reductions on critical links of multiple lines, the proposed method can identify the number of reduction links and the capacity reduction ratio that the network can withstand while maintaining its reliability and performance above a specified level.

1. Introduction

To deal with the urban challenges of congestion, noise, and emissions due to rising demand, urban rail transit (URT) development is promoted by China's government for its desirable economic, social, and environmental benefits [1]. Hence, URT networks in large urban areas have expanded rapidly in recent years and have a major role in public transport. As economic and social development advances, passengers expect from high quality and reliable transportation services from URT [2, 3]. However, URT networks are often subject to disturbances (e.g., public events, vehicle breakdowns, and equipment failures) in daily operations. Those disturbances cause link capacity reductions on the URT network, which results in degraded network performance and low reliability. As passenger volume on a URT network increases, the sensitivity of the network to the disturbances also increases. This may inhibit the passengers' willingness to shift from private transportation to public transportation. Accordingly, measuring and increasing the reliability of a URT network when link capacity reductions

occur are essential in URT policymaking, planning, and operations.

The rest of the paper is organized as follows: first, the related work is presented. Second, the metrics for measuring the reliability and unreliability of a URT network are proposed. A traffic assignment model is introduced to compute the metrics when link capacity reductions occur. Third, the reliability and unreliability of Wuhan's URT network during evening peak hours when link capacity reductions occur is analyzed. Finally, the main conclusions are summarized and suggestions are offered for maintaining the high reliability and performance of a URT network.

1.1. Literature Review. To measure the probability of disturbances and the impacts of disturbances on transportation networks as well as networks' ability to absorb disturbances, many studies use concepts such as reliability, vulnerability, robustness, and resilience [4]. Vulnerability is defined as the susceptibility to disturbances, which emphasizes measurement of the consequences of disturbances [5]. Robustness represents the ability to maintain system performance under disturbances, which is the inverse concept of vulnerability [6]. The resilience of a transportation network is defined as the network's ability to resist, absorb, adapt to, and recover from negative impacts of disturbances [6]. The reliability of a transportation network has been defined in dissimilar ways in various studies. Bell defines that a network is reliable if the expected trip costs are acceptable even when passengers are very pessimistic about the state of the network [7]. Kim et al. define the reliability of a transportation network as the possibility that the network performed its intended tasks satisfactorily for a certain period of time [8, 9]. Gu et al. specify that transportation network reliability is the probability that the network could remain satisfactory in terms of service level provision when perturbation occurs [4].

In practice, the degradation of a transport network could be due to many interruptions, such as operational accidents, natural disasters, and terrorist activities [10]. Those serious events cause complete failures of stations or links. The vulnerability of the transport network was measured based on network performance with and without complete failures at stations or links. Many network performance metrics were proposed that reflected the topology of the transport network [11], such as global connectivity (the minimum number of links whose removal disconnects the remaining nodes from each other), the largest connected component of the network, average shortest-path length, and other derivative metrics [12–15]. The vulnerability and resilience of a URT network were measured according to the relative differences in the above metrics due to network damage events compared to the normal situation. Simulations that removed stations or links were performed to analyze the vulnerability, robustness, and reliability of URT networks [12–14, 16]. The influence of removing stations and links on topologies of URT networks was analyzed deeply in the above studies. The research focused on connectivity issues when the network becomes disconnected due to complete failures at stations or links [17–19]. Although the network performance and service quality of URT's were heavily affected by total loss of station or link capacities, the likelihood of such serious events is low. On the contrary, small disturbances that reduce the capacities of links and stations often occur. In addition, the passengers' travel behavior and transport capacity were neglected in some studies when measuring the vulnerability, reliability, resilience, and robustness of URT networks. For example, Sun et al. analyzed the impact of removing stations on passengers' volume but without considering passengers' travel behaviors and the transportation service quality [20]. Kim and Song measured the reliability of a mass transit system based on the topological structure of the network and the passenger flow among pairs of stations [21]. They also neglected the effects of passengers' travel path selection behaviors on the reliability of a transportation network. Lu measured the resilience of the Shanghai metro network by integrating the network topology and passenger volume [22]. He assumed that passengers travel on the shortest travel paths between pairs of stations. De-Los-Santos et al. measured the robustness of a rail transit network when links were removed from it based on the passengers' overall travel time [23]. However, they assumed that passengers only chose the shortest paths and waited for the failure to be repaired when no alternative path existed, which was inconsistent with passengers' travel characteristics.

Researchers had noted the necessity of considering traffic demand and transport supply in studying the reliability, vulnerability, and robustness of transport networks [6, 24]. Snelder et al. also proposed a framework for analyzing the robustness of road networks for short-term variations in supply [25]. Considering traffic demand and transportation supply could measure the influence of full and partial link capacity reductions on network performance, which helped to measure the reliability, vulnerability, and robustness of transportation networks. Jiang et al. assessed the vulnerability of the transport network that considers both equilibrium flows and the probability of the disturbance [26]. Cats and Jenelius researched the impact of partial capacities degradation of lines on public transport network vulnerability [27]. Jenelius measured the reliability of a high-frequency bus line in Stockholm in terms of the passengers' perceived travel time [28]. While he analyzed the service reliability when bus lines operate normally, the effects of transport interference on the reliability of bus lines had not been analyzed. Liu et al. measured the travel reliability of a URT network with connectivity reliability, travel time reliability, and capacity reliability during normal operations of the URT network [24]; although passengers' route choices behavior was considered, the travel reliability of a URT network under disturbances was not measured.

The transportation network reliability is closely related to the acceptable level of service. Regarding the reliability of public transportation network reliability, many researchers had defined the acceptable level of service in terms of punctuality (i.e., on-time performance) and the coefficient of variation of vehicle headways. However, most standards for the acceptable level of service are proposed from the

perspective of suppliers rather than passengers [29, 30]. In addition, the standards focused on travel time or vehicle running time, which reflects the extent to which vehicles ran on time. However, the transportation service quality perceived by passengers on the journey is neglected. Even if passengers arrive at their destination on time, they may consider their travel unacceptable because of low perceived service quality to passengers.

1.2. Objective and Contribution. This paper evaluates the reliability and unreliability of a URT network when link capacity reductions occur. It analyzes cases not only where the link capacity is reduced to zero, but also cases where the link capacity is partially reduced. The passengers' generalized travel cost (GTC) which is the sum of the monetary value of perceived travel time and monetary cost of the trip is computed to reflect the service quality perceived by passengers. If link capacity reductions occur, the reliability of the URT network decreases because some passengers' perceived service quality is not acceptable. To determine whether the service perceived by a passenger is acceptable, the relative increase in GTC for passengers is computed. Passengers are classified into three categories according to their relative increase in GTC and whether paths with spare capacity exist between station pairs when link capacity reductions occur. Passengers whose relative increase in GTC on a URT is below or above a preset threshold belong to category I or II, respectively. With link capacity reductions, passengers in categories I and II can travel on a URT network. Passengers who cannot travel on the URT due to insufficient capacities on their paths belong to category III. The fraction of acceptable trips (FAT) is proposed to measure the network reliability within a given time period. Only passengers in category I perceive acceptable service when link capacity reductions occur, since their GTC increases only slightly. Thus, the FAT equals the fraction of passengers in category I, which is defined as the metric of URT network reliability within a given time period.

Passengers' total GTC is used to reflect the network performance of a URT. The difference in passengers' total GTC with and without link capacity reductions measures the impact of link capacity reductions on the URT network. The total GTC increase ratio (TGCR) is defined as the metric of the unreliability of a URT network when link capacity reductions occur.

To identify lines' most critical links, the ratio of affected passenger volume (RAPT) is proposed. Finally, the proposed indicators and method are applied to Wuhan's URT network during evening peak hours. The reliability of Wuhan's URT network during evening peak hours is computed when the capacities of each line's most critical link or multiple lines' most critical links are reduced. The proposed indicators and method help to identify each line's most critical links, as well as the extent and the relations among the capacity reductions on lines' most critical link and network performance, thereby supporting infrastructure management and capacity allocation.

2. Model Formulation

2.1. URT Network and Passengers' Categories. A weighted and directed graph $G = (S, E)$ is used to represent a URT network, which consists of the set of URT stations S and the set of URT links $E \subseteq \{(i, j)|i, j \in S, i \neq j\}$ that connect stations directly.

Passengers in categories I and II can travel on the URT network. However, only passengers in category I get acceptable service and their trips are acceptable when link capacity reductions occur, since their GTC increases only slightly. (E_{redu}, x) represents that the capacities of link set $E_{\text{redu}}, E_{\text{redu}} \in E$ are reduced by reduction ratio $x, x \in (0, 100\%]$. If $x = 100\%$, then capacities of links in the set E_{redu} are reduced to zero. A passenger n traveling from station i to station j is represented as p_{ij}^n. The category of the passenger p_{ij}^n is determined by equations (1) and (2) when link capacity reductions occur:

$$\lambda = \frac{C_{\text{ij}}^n(E_{\text{redu}}, x) - C_{\text{ij}}^n(0, 0)}{C_{\text{ij}}^n(0, 0)}, \tag{1}$$

$$p_{\text{ij}}^n \in \begin{cases} P_1^{\text{ij}}(E_{\text{redu}}, x); & \lambda \leq \kappa \cap \left|pa_{\text{ij}}\right| > 0; \\ P_2^{\text{ij}}(E_{\text{redu}}, x); & \kappa > \lambda \cap \left|pa_{\text{ij}}\right| > 0; \\ P_3^{\text{ij}}(E_{\text{redu}}, x); & \left|pa_{\text{ij}}\right| = 0; \end{cases} \tag{2}$$

where λ is the relative increase in GTC for passenger p_{ij}^n when link capacity reductions occur, $C_{\text{ij}}^n(0, 0)$ and $C_{\text{ij}}^n(E_{\text{redu}}, x)$ are the GTCs of passengers p_{ij}^n from station i to station j when a URT network operates normally and when capacities of link set E_{redu} are reduced by x, respectively, $P_1^{\text{ij}}(E_{\text{redu}}, x)$, $P_2^{\text{ij}}(E_{\text{redu}}, x)$, and $P_3^{\text{ij}}(E_{\text{redu}}, x)$ are the sets of passengers from station i to station j and belonging to categories I, II, and III, respectively, when capacities of link set E_{redu} are reduced by x, $|\text{pa}_{\text{ij}}|$ is the number of paths with spare capacity from station i to station j, and κ is a preset threshold.

2.2. GTC Computation. The passengers' GTC is the sum of the monetary and nonmonetary costs of a trip [31–33]. Nonmonetary cost refers to the perceived travel time undertaking the journey. The perceived travel time is converted to a monetary value using a value of time, which usually varies according to the traveler's income, trip component, and trip purpose. The travel time from a station to another is separated into different components, i.e., walking time, waiting time, and in-vehicle time. Research shows that the different components of travel time are perceived by passengers differently, e.g., waiting time has a much higher perceived value than in-vehicle travel time when the vehicle is not crowded. In addition, seat availability and crowding in the vehicle also affect passengers' perceived in-vehicle time. Therefore, each part of the travel time has its own value and weighting [34]. When a URT network operates normally, passenger p_{ij}^n's GTC from station i to station j is computed with the following equation:

$$C_{ij}^{n}(0,0) = t_{ij}^{n}(0,0) \cdot \alpha + c_{ij}, \tag{3}$$

where $C_{ij}^{n}(0,0)$ is the passenger p_{ij}^{n}'s GTC from station i to station j when a URT network operates normally, $t_{ij}^{n}(0,0)$ is the passenger p_{ij}^{n}'s perceived travel time from station i to station j when a URT network operates normally, c_{ij} is fare from station i to station j, which is not affected by link capacity reductions, and α is the value of time figure that converts the perceived travel time into money.

The perceived travel time includes perceived waiting time (waiting time at origin and transfer stations), perceived walking time (access, egress, and transfer walking time), perceived in-vehicle time considering seat availability and crowding in the vehicle, and increased perceived travel time due to transfer [27, 35]. Each perceived travel time component equals time multiplied by time weights. Passenger p_{ij}^{n}'s perceived travel time from station i to station j when a URT network operates normally is computed with the following equation:

$$\begin{aligned} t_{ij}^{n}(0,0) = {} & \beta^{\text{wait}} \cdot t^{\text{wait}}(0,0) + \beta^{\text{walk}} \cdot t^{\text{walk}}(0,0) \\ & + \beta^{\text{crow}} \cdot t^{\text{inv}}(0,0) + \beta^{\text{trans}} \cdot n^{\text{trans}}(0,0), \end{aligned} \tag{4}$$

where $t^{\text{wait}}(0,0)$, $t^{\text{inv}}(0,0)$, $t^{\text{walk}}(0,0)$, and $n^{\text{trans}}(0,0)$ are the waiting time, walking time, in-vehicle time, and transfer times, respectively, when a URT network operates normally. The β values are the corresponding time weights.

Passengers' waiting time, transfer times, and crowding in vehicles increase while seat availability decreases when link capacity reductions occur. Therefore, passengers' GTC increases when the capacities of link set E_{redu} are reduced by x. If passenger p_{ij}^{n} belongs to categories I and II, then his GTC $C_{ij}^{n}(E_{\text{redu}}, x)$ can be computed with equations (3) and (4) when capacities of link set E_{redu} are reduced by x. However, if passenger p_{ij}^{n} belongs to category III, then his GTC is assumed to be the maximum GTC of passengers in categories I and II when the URT network operates normally.

2.3. Metrics of the Reliability and Unreliability of a URT Network. Passengers in categories I and II still can travel on the URT network when link capacity reductions occur. Nevertheless, passengers in category I are only slightly affected by link capacity reductions and their perceived service is acceptable. Thus, the FAT is the fraction of passengers in category I which is defined as the metric of the reliability of a URT network in a given time period, which is computed with the following equation:

$$R_{\text{re}}(E_{\text{redu}}, x) = \frac{\sum_{i \in S} \sum_{j \in S, j \neq i} \left| P_1^{ij}(E_{\text{redu}}, x) \right|}{\sum_{i \in S} \sum_{j \in S, j \neq i} v_{ij}}, \tag{5}$$

where $R_{\text{re}}(E_{\text{redu}}, x)$ is the metric of the reliability of a URT network when capacities of link set E_{redu} are reduced by x, $|P_1^{ij}(E_{\text{redu}}, x)|$ is the number of passengers trips who travel from station i to station j and belong to category I in a given time period, and v_{ij} is passenger trips traveling from station i to station j in a given time period.

The total GTC can be used to evaluate the performance of the URT network [27]. The reliability of a URT network cannot be measured comprehensively by only using as a metric of the fraction of passengers in category I (e.g., Case 1: the fractions of passengers in categories I, II, and III are 0.8, 0.1, and 0.1, respectively; Case 2: the fractions of passengers in categories I, II, and III are 0.8, 0.15, and 0.05, respectively.) The reliability in Cases 1 and 2 is the same. However, the network performance in Case 2 is better than Case 1, since passenger' GTC in category III is higher than passengers' GTC in category II. Therefore, the effects of link capacity reductions on the URT network are measured as the difference in passengers' total GTC with and without link capacity reductions. Passengers' total GTC in a given time period is computed with equation (6) when a URT network operates normally.

$$C(0,0) = \sum_{i \in S} \sum_{j \in S, j \neq i} \sum_{n=1}^{v_{ij}} C_{ij}^{n}(0,0), \tag{6}$$

where $C(0,0)$ is the passengers' total GTC when the URT network operates normally.

Equation (6) can be applied to computed passengers' total GTC $C(E_{\text{redu}}, x)$ when capacities of link set E_{redu} are reduced by x. The TGCR is defined as the metric of the unreliability of a URT network when capacities of link set E_{redu} are reduced by x, which is computed with the following equation:

$$R_{\text{un}}(E_{\text{redu}}, x) = \frac{C(E_{\text{redu}}, x) - C(0,0)}{C(0,0)}, \tag{7}$$

where $R_{\text{un}}(E_{\text{redu}}, x)$ is the metric of the unreliability of a URT network when capacities of link set E_{redu} are reduced by x.

2.4. Simulation Based on Traffic Assignment. In order to measure the reliability of a URT network when link capacity reductions occur, a logit-based stochastic user equilibrium model, which has the advantage of reflecting passengers' travel behavior and the congestion effect [36, 37], is used to compute passengers' GTC and determine passengers' categories. The model is formulated as equations (8)–(16) and solved by the successive weighted averages method [38]:

$$\min Z(f) = -\sum_{i \in S} \sum_{j \in S} \sum_{m \in M_{ij}} v_{ij}^{m} \cdot E\left(C_{ij}, \theta\right) + \sum_{e \in E} \text{fl}_e \cdot C_e(\text{fl}_e) - \sum_{e \in E} \int_0^{\text{fl}_e} C_e(\omega) \mathrm{d}\omega. \tag{8}$$

subject to

$$\sum_{m} v_{ij}^{m} = v_{ij}, \tag{9}$$

$$v_{ij}^{m} = \frac{\exp\left(-\theta \cdot C_{ij}^{m}\right)}{\sum_{m \in M_{ij}} \exp\left(-\theta \cdot C_{ij}^{m}\right)} \cdot v_{ij}, \tag{10}$$

$$E\left(C_{ij}, \theta\right) = -\frac{1}{\theta} \mathrm{In} \sum_{m \in M_{ij}} \exp\left(-\theta \cdot C_{ij}^{m}\right), \tag{11}$$

$$\mathrm{fl}_{e} = \sum_{i} \sum_{j} \sum_{m \in M_{ij}} v_{ij}^{m} \cdot \delta_{e,m}^{ij}, \tag{12}$$

$$\mathrm{Cap}_{e} = f_{e} \cdot n_{c}, \tag{13}$$

$$\mathrm{fl}_{e} \leq \mathrm{Cap}_{e}, \tag{14}$$

$$\psi_{e} = n_{s} \cdot f_{e}, \tag{15}$$

$$C_{e}(\omega) = \beta^{\mathrm{crow}} \cdot t_{e}^{\mathrm{inv}}, \tag{16}$$

where θ is a nonnegative parameter that represents the accuracy of passengers' perception of travel cost, v_{ij}^{m} is the volume of passenger trips from station i to station j on path m, which is computed with the logit model of equation (10), C_{ij}^{m} is the GTC of path m from station i to station j, v_{ij} is the passenger trip volume from station i to station j, $E(C_{ij}, \theta)$ is the mathematical expectation of passengers for GTC perception, $C_{ij} = \left\{\cdots, C_{ij}^{m}, \cdots\right\}$ is the GTC set corresponding to paths from station i to station j, fl_{e} is the passenger flow on link e, which is the sum of passenger flows on paths that contain link e, and $\delta_{e,m}^{ij}$ is a 0-1 binary variable. If the path m from station i to station j contains the link e, then $\delta_{e,m}^{ij} = 1$; otherwise, $\delta_{e,m}^{ij} = 0$. Cap_{e} is the capacity of the link e, which is the number of trains f_{e} traversing link e per hour multiplied by the maximum allowable number of passengers per train n_{c}. ψ_{e} is the number of seats on link e, which equals the seats of per train n_{s} multiplied by the number of trains running f_{e} on the link e per hour. $C_{e}(\omega)$ is the functional relation between travel cost in a link and the link's passenger flow ω, which is expressed in equation (16). t_{e}^{inv} is the in-vehicle time on link e.

2.5. Identification of Critical Links. The betweenness of links has been widely used to identify the critical links on a URT network [23]. The betweenness of a URT link is the ratio of the number of the shortest paths containing that link to the number of all shortest paths between OD pairs. It is computed with the following equation:

$$B_{e} = \sum_{i \in N} \sum_{j \in N, j \neq i} \frac{n_{ij}^{e}}{n_{ij}}, \quad \forall e \in E, \tag{17}$$

where B_{e} is the betweenness of link e, n_{ij} is the number of shortest paths from station i to station j, and n_{ij}^{e} is the number of shortest paths from station i to station j which contains link e.

The betweenness of links is proposed from the topology of a URT network. It measures the criticality of links from the shortest paths on a URT network. However, it cannot measure the effects of link disruptions on passengers' travel comprehensively, since it neglects the passenger flow on the URT network. The ratio of affected passenger trips (RAPT) is proposed to measure the passenger trips that a link can affect, i.e., the passenger trips whose travel paths include that link. It is used to identify critical links and specified with the following equation:

$$I_{e} = \frac{\sum_{i \in S} \sum_{j \in S, j \neq i} \sum_{m \in M_{ij}} v_{ij}^{m} \cdot \delta_{e,m}^{ij}}{\sum_{i \in S} \sum_{j \in S, j \neq i} v_{ij}}, \quad \forall e \in E, \tag{18}$$

where I_{e} is the ratio of affected passenger trips who is affected by link e, which is computed based on a logit-based stochastic user equilibrium model.

3. Application

3.1. Case Study Description. Wuhan's URT network is used in a case study to measure its reliability when link capacity reductions occur. Wuhan's URT network connects three parts of Wuhan (Wuchang, Hankou, and Hanyang) which are divided by Yangtze and Han rivers (the black dotted lines in Figure 1). Wuhan's URT included 7 lines and 149 stations (17 transfer stations) in September 2018. Figure 1 shows the operating lines, station numbers, and abbreviations of station names for Wuhan's URT in September 2018.

3.2. Data Preparation. Automatic Fare Collection (AFC) data records for 5 working days in September 2018 are obtained from Wuhan's URT operator to measure the reliability of that URT network during evening peak hours (5: 30–7:30). The average OD (origin station to destination station) trip distribution is computed using AFC data. The waiting time at origin station and transfer station is estimated as half of the lines' headway (Dixit et al. 2019 and De-Los-Santos et al. 2012) [23, 39]. The access walking time and egress walking time at different stations are different. Here, we assume that access walking time or egress walking time at a station is 4 minutes, according to our survey. The hourly capacities of lines and the number of seats per hour on lines (links have the same capacity and same seats along a line), the headway of lines, the in-vehicle travel time (including dwell time and train running time) of links, and the transfer walking time at transfer stations are obtained from an operator. The capacities, headway, and number of seats of lines during evening peak hours are shown in Table 1. Taking line 1 as an example, the transfer walking time at transfer stations and in-vehicle travel time of some links on line 1 are shown in Tables 2 and 3, respectively, during evening peak hours.

The preset threshold which is used to determine passengers' categories when link capacity reductions occur is obtained by conducting a questionnaire survey of passengers in Wuhan's URT network. A total of 1427 questionnaires are

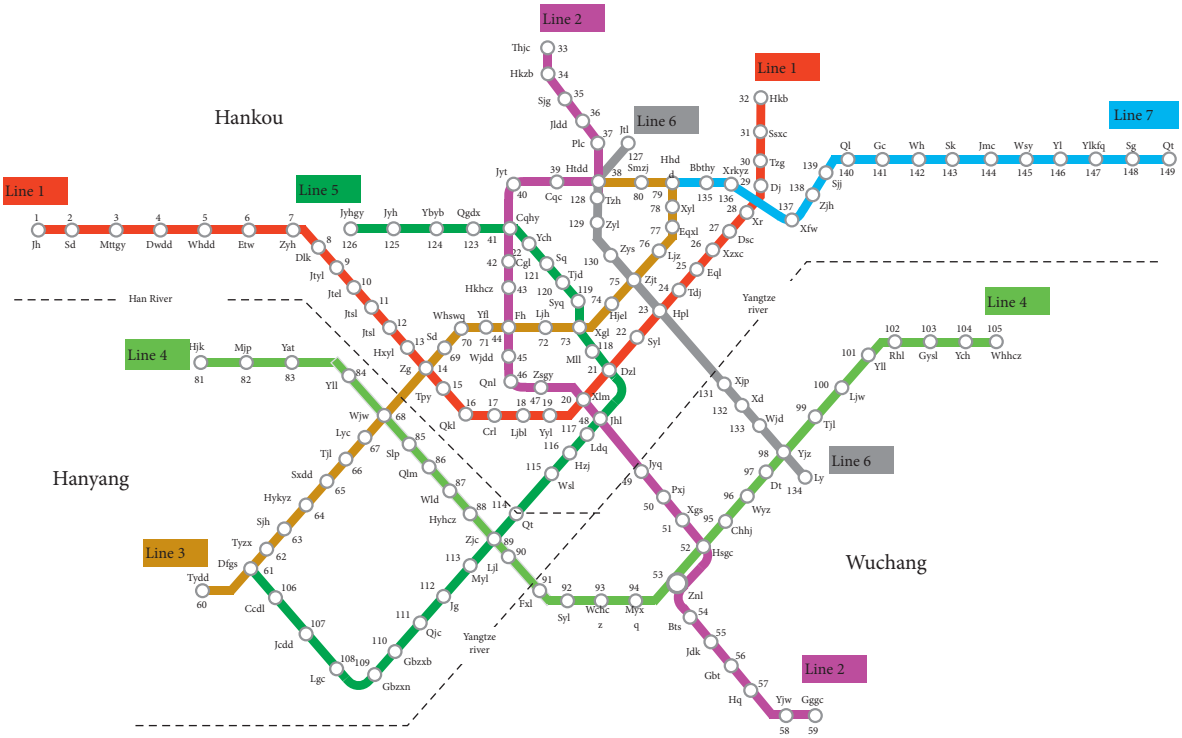


Figure 1: Wuhan's URT network in September 2018.

Table 1: Transport capacities and average headway of lines.

Lines	Downstream			Upstream		
	Headway (min)	Capacity (trips/h)	Seats (seats/h)	Headway (min)	Capacity (trips/h)	Seats (seats/h)
1	3.8	23 360	5 600	4.6	20 440	5 950
2	3.8	23 360	5 600	3.3	26 280	6 300
3	4.3	20 440	4 900	3.8	23 360	5 600
4	3.8	23 360	5 600	3.3	26 280	6 300
5	3.3	26 280	6 300	3.8	23 360	5 600
6	4.3	20 440	4 900	4.3	20 440	5 950
7	6.0	14 600	3 500	7.5	11 680	2 800

Table 2: Transfer walking time at transfer stations on line 1.

Station	Transfer direction	Time (min)	Transfer direction	Time (min)
14	Line 1 to line 3	2.9	Line 3 to line 1	2.6
20	Line 1 to line 2	2.5	Line 2 to line 1	2.0
21	Line 1 to line 5	2.7	Line 5 to line 1	2.7
23	Line 1 to line 7	2.9	Line 7 to line 1	2.7

Table 3: In-vehicle time of some links on line 1.

Link	In-vehicle time (min)		Link	In-vehicle time (min)	
	Downstream	Upstream		Downstream	Upstream
1-2	1.9	1.7	6-7	1.6	1.2
2-3	3.4	3.0	7-8	1.4	1.6
3-4	1.8	1.6	8-9	2.5	2.5
4-5	2.8	2.8	9-10	2.6	3.1
5-6	2.8	3.1	10-11	1.6	1.4

returned, including 1302 valid questionnaires. The effective sampling rate of the questionnaires is 91.24%, which meets the statistical requirements. The survey sample includes 697 men and 605 women. The surveyed passengers aged under 30 years old, between 30 and 60 years old, as well as over 60 years old account for 46.93%, 44.39%, and 8.68%, respectively, of total surveyed passengers. In addition, students (31.57%), officials (9.68%), businessmen (43.24%), and workers (15.51%) are surveyed. From the survey analysis, it is found that, during disturbances in Wuhan's URT network, passengers think their trips are acceptable only when their average relative increase in GTC is below 0.34. Therefore, the preset threshold to determine passengers' categories is 0.34. To compute passengers' GTC, the time weights, value of time, and fare are determined. The recommended default value for personal travel time should be 30% of household income per hour according to the economist Kenneth Gwilliam [40]. Therefore, the value of time α is estimated to

be 16.54 ¥/hour for passengers based on the household income of 161,000 ¥/per year in Wuhan. The time weights are $\beta^{\text{wait}} = 1.75$, $\beta^{\text{walk}} = 1.75$, and $\beta^{\text{trans}} = 8.35$ according to Cats and Jenelius [27]. β^{crow} is related to the seat availability (sitting or standing) and crowding in the trains, which is shown in Table 4 [35]. The load factor is the ratio of passenger flow on a link to the number of seats on that link. It can indicate the crowding in the trains. The fare between an OD pair in Wuhan's URT is determined based on the shortest distances between that OD pair. The relation of fares to the shortest distance between an OD pair is shown in Table 5.

3.3. Network Performance in Normal Operation. A logit-based stochastic user equilibrium model is applied to assign passenger OD trips to Wuhan's URT network during evening peak hours. The passenger volume on links is shown in Figure 2, which shows that the links with high passenger volume are located at the center of Wuhan. In addition, the links which connect three parts of Wuhan (Wuchang, Hankou, and Hanyang) have high passenger volumes.

The passengers' actual travel time and perceived travel time on Wuhan's URT network during evening peak hours are computed after passenger trip assignment. The average travel time per trip and average perceived travel time per trip on Wuhan's URT network are 32.57 minutes and 68.92 minutes, respectively, during evening peak hours. Figure 3 shows that the perceived in-vehicle time is 1.93 times of actual in-vehicle time per trip. In addition, the increased perceived travel time is 7.73 minutes per trip due to transfer, which cannot be measured fully with actual travel time. Therefore, the perceived travel time can reflect more passengers' travel information and travel quality than actual travel time.

The average GTC per trip is 23.43 ¥ which equals the fare (4.42 ¥) plus the monetary value of perceived travel time (19.01 ¥). Passengers' total GTC is 8,156,325.79 ¥ on Wuhan's URT network during evening peak hours according to 348,115 total passenger trips during evening peak hours. The components of the average GTC per trip are shown in Figure 4. It shows that the largest component of the GTC is the perceived in-vehicle time (53.97%). The monetary values of perceived walking time, waiting time, and increased perceived time due to transfer are, respectively, 10.48%, 7.58%, and 9.10% of the GTC on Wuhan's URT network during evening peak hours. The passengers' total GTC is computed with perceived travel time conveys service quality perceived by passengers (e.g., crowding in the vehicle, seat availability, perceived time components, and increased perceived travel time due to transfer).

3.4. Every Line's Most Critical Link and Its Simulated Capacity Reductions. The RAPT and betweenness of links are computed after the passenger OD trips assignment. Every line's most critical link on Wuhan's URT network during evening peak hours (shown in Table 6) is identified based on RAPT and betweenness.

The implicit assumptions of using betweenness of links to identify critical links in URT networks are that passenger trips between OD pairs are the same (all OD pairs are equally important) and all passengers choose the shortest paths to travel between OD pairs. However, the passenger trips between OD pairs vary greatly in the URT, and thus the weights of OD pairs are not the same. Passengers choose travel paths between OD pairs according to the GTC of paths. Therefore, not all passengers actually travel through the shortest paths. Critical links identification with RAPT considers the passenger trips between OD pairs and passengers' travel behavior. Therefore, Table 6 shows that the critical links identified with the betweenness and RAPT are quite different.

To simulate the link capacity reductions (E_{redu}, x) that often occur on a URT network, capacity reductions on each line's most critical link and multiple lines' most critical links are simulated, respectively, on Wuhan's URT network during evening peak hours. It is assumed that the links are independent of each other in a simulation [11, 41], since it is difficult to estimate the effect of link capacity reductions on other links' capacities.

The capacity reductions on each line's most critical link are simulated when its capacity is reduced by x (for x values of 10%, 20%, 30%, ..., 90%, 100%). The capacities of the most critical links are reduced by the same ratio x when capacity reductions on multiple lines' most critical links are simulated. The capacity reductions on the most critical links of two to seven lines, identified with RAPT and betweenness, respectively, are simulated. In addition, the capacity reduction links are determined by their rankings on RAPT and betweenness values, respectively, which are shown in Table 6. For example, if the capacity reductions on three lines' most critical links identified with RAPT are simulated, then the three lines are lines 5, 1, and 3. After the capacities of lines' most critical links are reduced, a logit-based stochastic user equilibrium model is used to compute the passengers' GTC using equations (3) and (4). Then, the reliability and unreliability metrics of Wuhan's URT network during evening peak hours are computed.

3.5. Results

3.5.1. Reliability of Wuhan's URT When Capacity Reductions Occur on the Lines' Most Critical Links

(1) Capacity Reductions on Each Line's Most Critical Link. The reliability metric of Wuhan's URT network (FAT) is computed during evening peak hours when the capacity of each line's most critical link is reduced and is shown in Figure 5. The results show that the capacity reduction on line 1 to 5's most critical link identified with RAPT decrease the FAT on Wuhan's URT network faster than for the most critical link identified with betweenness. This indicates that critical links can be identified more accurately with RAPT than with betweenness. The reason is that the most critical links identified with RAPT have high passenger volume and are located at Wuhan's center. If their capacities decrease,

Table 4: The value of β^{crow} in different load factors.

Load factor (%)	β^{crow}	
	Sitting	Standing
0–75	0.86	—
75–100	0.95	—
100–125	1.05	1.62
125–150	1.16	1.79
150–175	1.27	1.99
175–200	1.40	2.20
>200	1.55	2.44

Table 5: Fares corresponding to the shortest distance between an OD pair.

Distance (km)	(0, 4]	(4, 8]	(8, 12]	(12, 18]	(18, 24]	(24, 32]	(32, 40]	(40, 50]	>50
Price (¥)	2	3	4	5	6	7	8	9	10

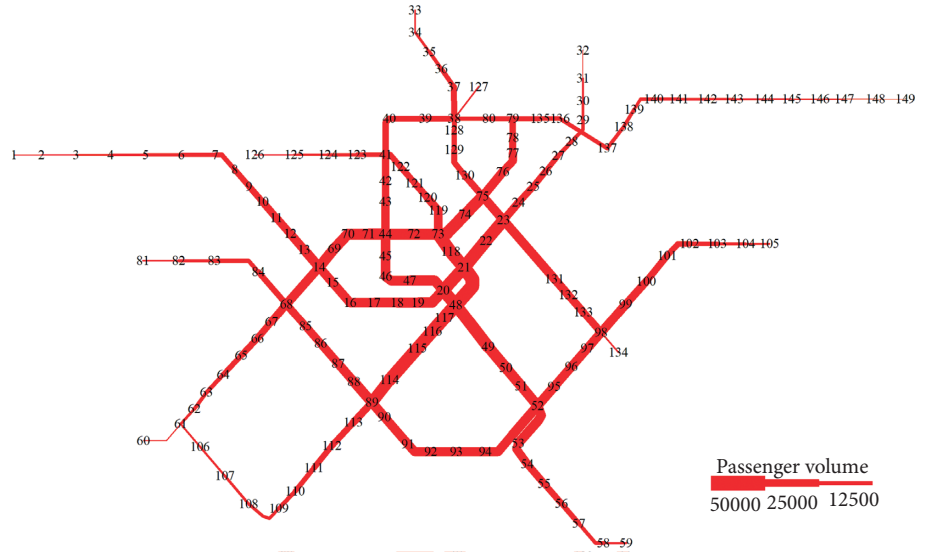


Figure 2: Passenger volume on Wuhan's URT links during evening peak hours.

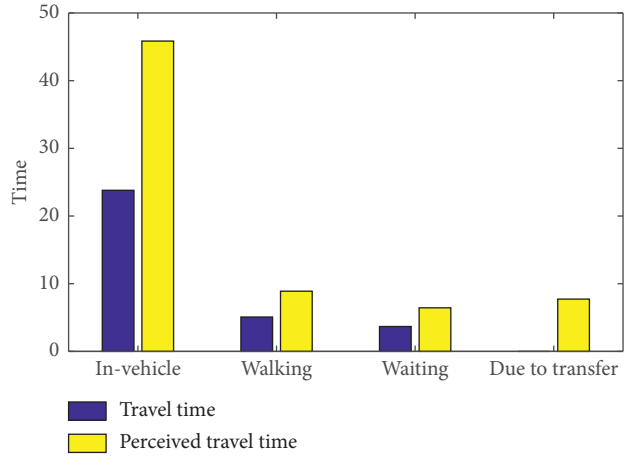


Figure 3: The components of the average travel time per trip and average perceived travel time per trip.

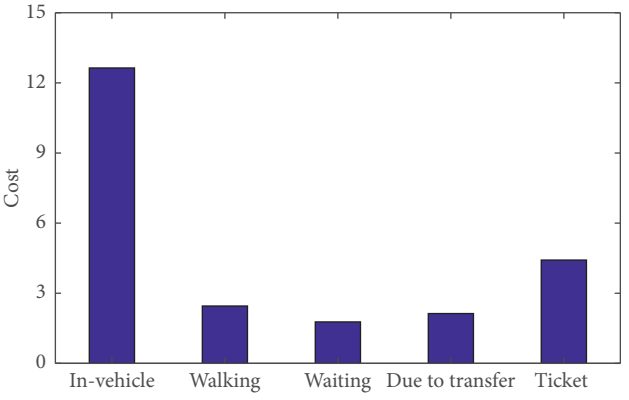


Figure 4: The components of the average GTC per trip.

Table 6: Every line's most critical link on Wuhan's URT network during evening peak hours.

Critical link	Identified with RAPT			Identified with betweenness		
	Link (station-station)	RAPT (%)	Rank	Link (station-station)	Betweenness (%)	Rank
On line 1	21-22	12.60	2	13-14	9.06	5
On line 2	20–48	12.32	4	53-54	6.24	7
On line 3	73-72	12.50	3	73-74	16.38	1
On line 4	88-89	9.83	5	98-99	6.57	6
On line 5	48–117	14.95	1	21–48	12.04	3
On line 6	23–75	8.98	6	132-133	11.22	4
On line 7	79–135	4.79	7	135-136	12.18	2

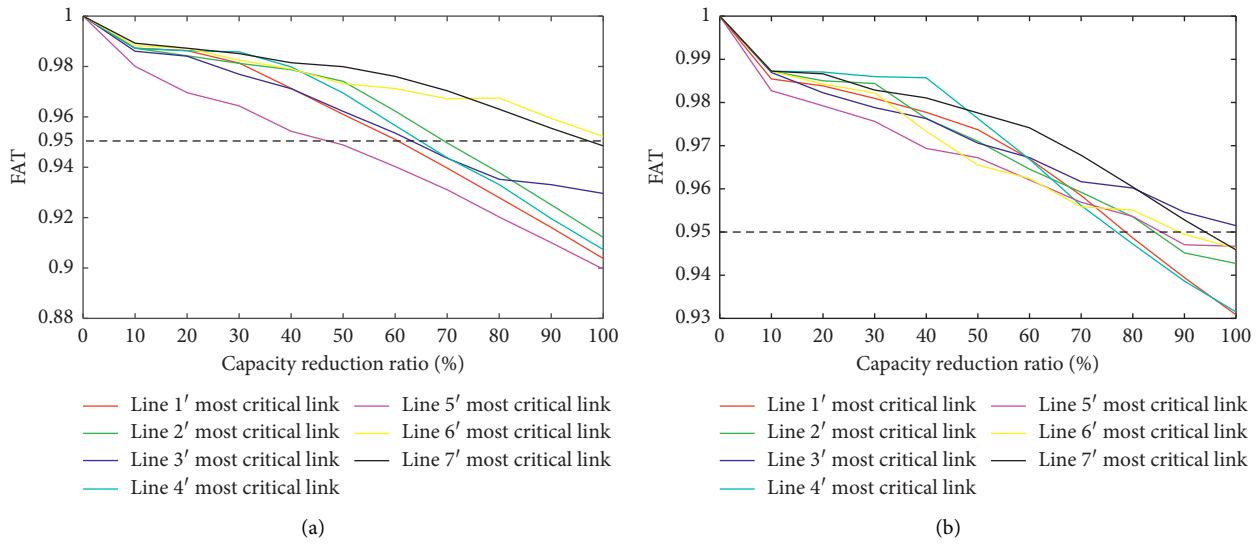


Figure 5: Reliability of Wuhan's URT network (a) when the capacity of each line's most critical link identified with RAPT is reduced and (b) when the capacity of each line's most critical link identified with betweenness is reduced.

then more passengers are shifted to categories II and III. Figure 5 illustrates that, as the capacity decreases, the FAT on Wuhan's URT network decreases slowly at first and then rapidly. The reason is that capacity reductions on a line's most critical link increase slowly the number of passengers who belong to category II when the capacity reduction ratio is below 50%. However, if the capacity of a line's most critical link decreases by over 50%, the passengers in category III increase rapidly due to their path flow exceeding path capacities.

The guidance for dealing with capacity reductions on each line's most critical link is obtained by analyzing Figure 5. For example, to maintain the high reliability of Wuhan's URT network during evening peak hours, i.e., guarantee that the FAT exceeds 0.95, the operators should avoid capacity reductions below 50% in line 5's most critical link which is identified with RAPT. Similarly, the capacity reduction ratio for another line's most critical link which corresponds to the networks' FAT of 0.95 can be determined. It can be used as a monitoring capacity reduction ratio to ensure the high reliability of Wuhan's URT network.

The effect of capacity reductions on a line's most critical link identified with RAPT on another line's reliability in Wuhan's URT network during evening peak hours is then analyzed. The FAT on a line is the fraction of passengers departing from that line and belonging to category I. The FAT on lines is computed when the capacity of a line's most critical link identified with RAPT is reduced by a ratio corresponding to the network's FAT value of 0.95. The result, shown in Table 7, indicates that capacity reductions on a line's most critical link reduce the FAT not only on that line but also on other lines. Taking line 1's most critical link as an example, if its capacity is reduced by 60% (corresponding to the network's FAT value of 0.95), then the FAT on lines 1 and 5 decreases to 0.91 and 0.93, respectively. The reason for the decreased reliability of line 5 is that passengers who ride line 5 from Hanyang to Hankou must pass through line 1's most critical link during evening peak hours. Thus capacity reductions on line 1's most critical link shift many passengers into categories II and III.

Some suggestions can be offered for mitigating the negative effect of capacity reductions on a line's most critical link on another line's reliability in Wuhan's URT network during evening peak hours. Taking line 1's most critical link as an example, if its capacity is reduced, then operators should emphasize prompt evacuation measures not only for line 1 but also for line 5. It means operators should take measures to recover the capacity of line 1's most critical link

Table 7: The FAT on lines when the capacity of a line's most critical link is reduced.

Critical link	Capacity reduction ratio (%)	FAT on lines						
		Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7
Line 1's	60	**0.91**	0.96	0.99	0.99	**0.93**	0.97	0.97
Line 2's	60	0.97	**0.92**	0.99	0.99	**0.95**	0.99	0.98
Line 3's	60	0.99	0.98	**0.87**	0.98	0.96	0.98	**0.92**
Line 4's	60	**0.96**	0.98	0.97	**0.92**	0.95	0.99	0.98
Line 5's	40	0.96	**0.95**	0.98	0.99	**0.79**	0.97	0.97
Line 6's	90	0.99	0.98	0.96	0.98	0.96	**0.70**	**0.93**
Line 7's	90	0.99	**0.95**	0.96	0.97	0.97	0.98	**0.71**

as soon as possible. Operators also need to transport passengers on line 5 rapidly by adding trains, enhancing passenger travel guidance and expediting security checks. Similarly, the priority line for which emergency measures should be taken can be determined from Table 7 when the capacity of a line's most critical link identified with RAPT is reduced.

(2) Capacity Reductions on Multiple Lines' Most Critical Links. The FAT on Wuhan's URT network during evening peak hours is computed when the capacities of multiple lines' most critical links are reduced, as shown in Figure 6. Figures 5 and 6 demonstrate that the capacity reductions on multiple lines' most critical links have a much stronger negative effect on the reliability of Wuhan's URT network than reductions on one line's most critical link. Figures 6(a) and 6(b) illustrate that the capacity reductions on multiple lines' most critical links identified with RAPT have a higher negative influence on the reliability of Wuhan's URT network than those caused by capacity reductions on multiple lines' most critical links identified with betweenness. The relation between network reliability, the number of a line's most critical links whose capacities are reduced, and the capacity reduction ratio shows that even if the number of a line's most critical links whose capacities are reduced increases to 7, the FAT on Wuhan's URT network can be maintained at about 0.9 when the capacity reduction ratio is below 20%. The reason for this result is that the ratio of passenger flow to capacities of lines' most critical links is nearly 80% on Wuhan's URT network during evening peak hours. If the capacities of lines' most critical links decrease by less than 20%, then some passengers are shifted into category II and few passengers are shifted into category III. Therefore, the FAT is nearly 0.9 in such a case. However, if the capacity reduction ratio exceeds 20%, many passengers should travel on new alternative paths because there is no spare capacity in their original paths. This shifts many passengers into category II and greatly decreases the FAT on Wuhan's URT network.

Some suggestions can be provided to operators by analyzing the result when the capacities of multiple lines' most critical links are reduced: to maintain the FAT on Wuhan's URT network above 0.9, the capacities of lines' most critical links should not decrease by more than 20%. The high reliability of Wuhan's URT network (whose FAT is over 0.95) can still be maintained if the capacities of the any two lines' most critical links are reduced by 20% or the capacities of any three lines' most critical links are reduced by 10%.

3.5.2. Unreliability of Wuhan's URT When the Critical Link Capacity Reductions Occur

(1) Capacity Reductions on Each Line's Most Critical Link. The network's unreliability metric for Wuhan's URT network during evening peak hours is computed when the capacity of each line's most critical link is reduced, which is shown in Figure 7. The capacity reduction on line 1 to 5's most critical link is identified with RAPT increase TGCR faster than for the most critical link identified with betweenness. Therefore, PAPT identifies lines' most critical links which greatly affect Wuhan's URT performance more accurately than betweenness. Figure 7 illustrates that capacity reduction on line 5's most critical link identified with RAPT significantly increases the total GTC and degrades the network performance.

To mitigate the negative impact of capacity reductions in a lines' most critical link on network performance of Wuhan's URT during evening peak hours, operators should prevent capacity reductions occurring on line 5's most critical link (link 48-117 identified with RAPT). For example, operators should strengthen the protection and inspection of lines' most critical links. To maintain the high performance of Wuhan's URT network, i.e., guarantee the TGCR below 0.05, operators should avoid capacity reductions exceeding 10% on line 5's most critical link. Similarly, the capacity reduction ratio for another line's most critical link which corresponds to network's TGCR of 0.05 can be determined. It can serve as a monitoring capacity reduction ratio to ensure the good performance of Wuhan's URT network.

The effect of capacity reductions on a line's most critical link identified with RAPT on another line's performance in Wuhan's URT network during evening peak hours is then analyzed. The impact on the line's performance is measured with the TGCR of passengers departing from that line. The TGCR on every line is computed when the capacity of a line's most critical link identified with RAPT is reduced by the capacity reduction ratio corresponding to network's TGCR of 0.05. The result is shown in Table 8, which demonstrates that capacity reductions on a line's most critical link increase total GTC on that line and on another line. The capacity reduction ratio for line 7's most critical link corresponding to the network's TGCR of 0.05 is highest among the lines' most critical links. The reason is that line 7 is located in Wuhan's surrounding area which has low travel demand. The TGCR on line 6 is high when the capacity of line 7's most critical link is reduced by 60% (corresponding to network's

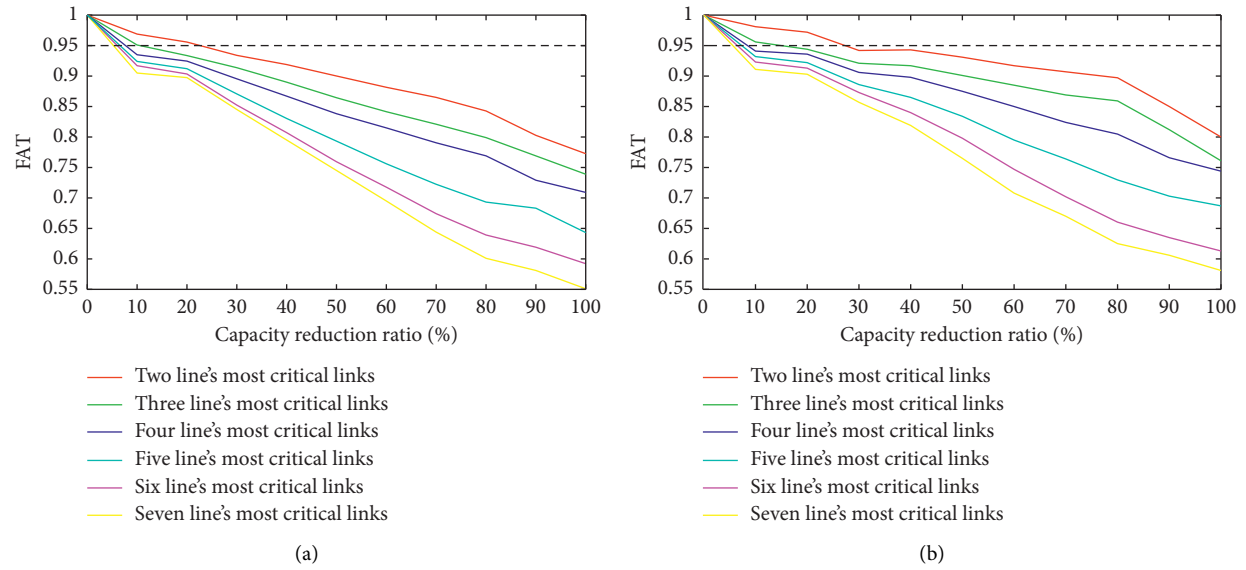


Figure 6: Reliability of Wuhan's URT network (a) when the capacities of multiple lines' most critical links identified with RAPT are reduced and (b) when the capacities of multiple lines' most critical links identified with betweenness are reduced.

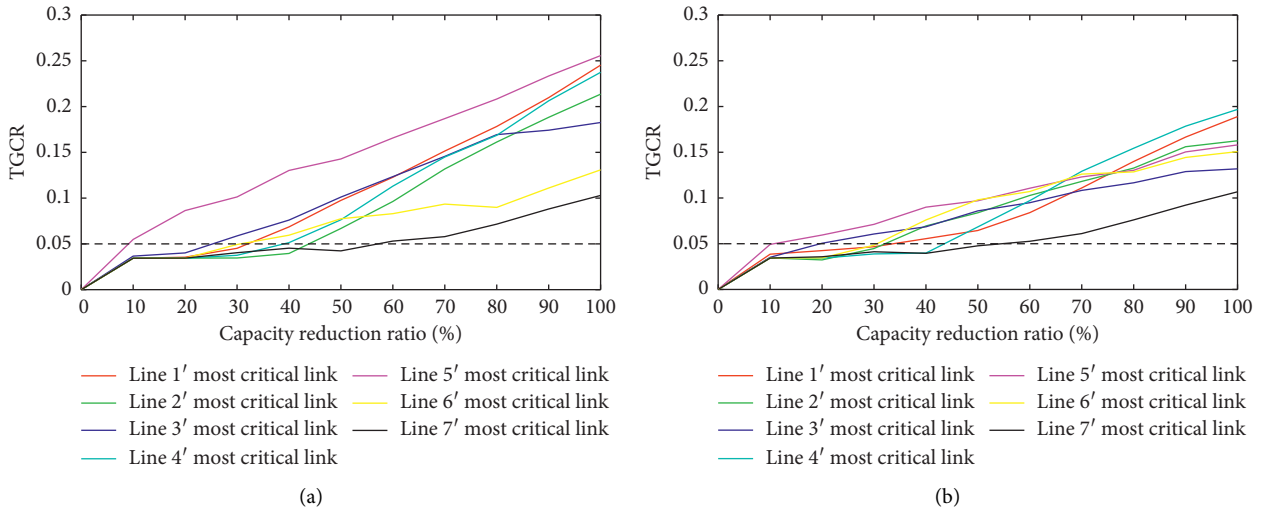


Figure 7: Unreliability of Wuhan's URT network (a) when the capacity of each line's most critical link identified with RAPT is reduced and (b) when the capacity of each line's most critical link identified with betweenness is reduced.

TGCR of 0.05), since it shifts many passengers into category III who depart from line 6 to line 7. Therefore, to maintain a high performance of Wuhan's URT network (with TGCR below 0.05), the capacity of line 7's most critical link should not decrease by over 60%. If the capacity reductions occur on line 7's most critical link (link 79-135), then the emergency measures should be taken for lines 6 and 7 to avoid poor network performance. These measures include running buses between stations 79 and 135 to transport passengers who need travel through capacity reduction link; adding trains on line 6, enhancing passenger travel guidance and expediting security checks on lines 6 and 7 to transport passengers on lines 6 and 7 as soon as possible. The priority line for which emergency measures should be taken can be determined from Table 8 when the capacity reductions occur on a line's most critical link.

(2) Capacity Reductions on Multiple Lines' Most Critical Links. The unreliability of Wuhan's URT network during evening peak hours is shown in Figure 8 when the capacities of multiple lines' most critical links are reduced. Figures 7 and 8 demonstrate that the capacity reductions on multiple lines' most critical links have a much stronger negative impact on the performance of Wuhan's URT network than that caused by capacity reductions on a line's most critical link. TGCR is high, and the network performance is poor when capacity reductions occur on multiple lines' most critical links. Figure 8 indicates that capacity reductions on

Table 8: The TGCR on every line when the capacity of a line's most critical link is reduced.

Critical link	Capacity reduction ratio (%)	TGCR on every line						
		Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7
Line 1	30	**0.07**	0.03	0.02	0.02	**0.04**	0.02	0.02
Line 2	40	0.02	**0.05**	0.01	0.02	**0.03**	0.01	0.01
Line 3	30	0.01	0.01	**0.05**	0.02	0.03	**0.06**	0.05
Line 4	30	0.02	**0.04**	0.01	**0.06**	0.02	0.02	0.01
Line 5	10	0.01	**0.04**	0.01	0.02	**0.05**	0.02	0.01
Line 6	40	0.01	0.01	0.01	0.01	0.02	**0.11**	**0.06**
Line 7	60	0.01	0.01	0.03	0.01	0.01	**0.07**	**0.12**

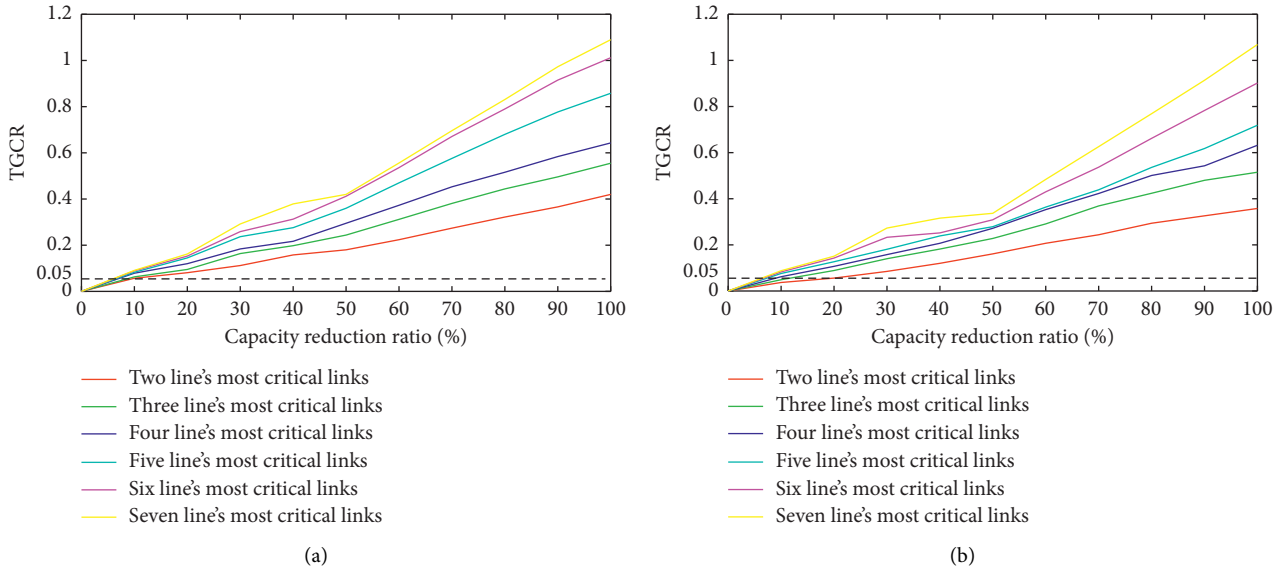


Figure 8: Unreliability of Wuhan's URT network (a) when the capacities of multiple lines' most critical link identified with RAPT are reduced and (b) when the capacities of multiple lines' most critical link identified with betweenness are reduced.

multiple lines' most critical links identified with RAPT have a higher negative influence on the performance of Wuhan's URT network than those identified with betweenness. The relation between network unreliability and capacity reduction ratio shows two inflection points when the capacity reduction ratio is 20% and 40%. Two reasons can account for this result. One reason is that the ratio of passenger flow to capacities of lines' most critical links is nearly 80% on Wuhan's URT network during evening peak hours. The TGCR is low because the number of passengers in categories II and III is low when the capacities of multiple lines' most critical links decrease by less than 20%. The other reason is that alternative paths exist on Wuhan's URT network when the capacity reduction ratio is between 20% and 40%. Most passengers can travel on alternative paths when the capacities of multiple lines' most critical links decrease between 20% and 40%. If the capacity reduction ratio exceeds 40%, then the TGCR increases rapidly since many passengers are shifted into category III due to the passenger flows on alternative paths exceeding capacities.

Two important suggestions for operators may be stated based on the above analysis: to avoid total GTC increases exceeding 15% on Wuhan's URT network during evening peak hours, the capacity of multiple lines' most critical links should not decrease by more than 20%. To avoid the high increase in total GTC and poor network performance of Wuhan's URT, operators should avoid having capacities of multiple lines' most critical links decrease by over 40%.

4. Conclusions

This paper proposed measures for evaluating the reliability and unreliability of a URT network when the link capacity reductions occur. The most critical link on a line is identified with the proposed RAPT indicator. The passengers' GTC and total GTC are used to reflect service perceived by passengers and network performance, respectively. To measure the reliability of a URT network when link capacity reductions occur, the FAT and the TGCR are defined as the reliability and unreliability metrics, respectively. A logit-based stochastic user equilibrium model is applied to compute passengers' GTC and determine the passengers' categories.

Fully and partially reduced link capacities are both considered in analyzing the effect of capacity reductions in

the lines' most critical links on the reliability of Wuhan's URT network during evening peak hours. The following conclusions may be drawn:

(1) The proposed RAPT indicator is effective in identifying lines' most critical links that greatly affect a URT's network reliability and performance.

(2) The capacity reduction ratio for each line's most critical link corresponding to networks' FAT of 0.95 and TGCR of 0.05 is determined. It can be used as a monitoring capacity reduction ratio to ensure high reliability and good performance of Wuhan's URT network.

(3) The priority line for which emergency measures should be taken can be determined after simulating capacity reductions on each line's most critical link. It can mitigate the negative effect of the capacity reductions in a line's most critical link on the reliability and performance of another line.

(4) In order to maintain the network reliability and network performance of Wuhan's URT above a certain level, the number of most critical links whose capacities are reduced in each line and the capacity reduction ratio that the network can withstand are determined. This can help operators in preparing emergency plans and assessing the harm caused by capacity reductions on lines' most critical links.

The reliability analysis of a URT network when link capacity reductions occur can facilitate the integration of measures for improving network reliability into URT network planning, operations, and real-time management. The identification of lines' most critical links also can provide guidance for infrastructure investment decisions and daily protection. Future studies may consider the probability of link capacity reductions and capacity recovery time. It is also desirable to further analyze how the impact of link capacity reductions on the reliability of a URT network can be mitigated.

Abbreviations

(1) Sets

S: The set of URT stations

E: The set of URT links

E_{redu}: The capacity reduction link set

$P_1^{\text{ij}}(E_{\text{redu}}, x)$, $P_2^{\text{ij}}(E_{\text{redu}}, x)$, and $P_3^{\text{ij}}(E_{\text{redu}}, x)$: The sets of passengers from station i to station j who belong to categories I, II, and III, respectively, when capacities of link set E_{redu} are reduced by x

C_{ij}: GTC set corresponding to paths from station i to station j

(2) Elements

i and j: Stations belong to S

e: Link belongs to L

C_{ij}^m: The GTC corresponding to the path m from station i to station j, $C_{\text{ij}}^m \in C_{\text{ij}}$

(3) Parameters

κ: The preset threshold

α: The value of time figure that converts perceived journey time into money

β^{wait}, β^{walk}, β^{crow}, and β^{trans}: Values of time weights for waiting time, walking time, in-vehicle time, and transfer times, respectively

θ: A nonnegative parameter represents the accuracy of passengers' perception of travel cost

$\delta_{e,m}^{\text{ij}}$: A 0-1 binary variable; if the path m from station i to station j contains the link e, then $\delta_{e,m}^{\text{ij}} = 1$; otherwise, $\delta_{e,m}^{\text{ij}} = 0$

(4) Variables

B_e: Betweenness of link e

$C(0,0)$: Passengers' total GTC when the URT network operates normally

$C(E_{\text{redu}}, x)$: Passengers' total GTC when capacities of link set E_{redu} are reduced by x

$C_{\text{ij}}^n(0,0)$: Passenger p_{ij}^n's GTC from station i to station j when the URT network operates normally

$C_{\text{ij}}^n(E_{\text{redu}}, x)$: Passenger p_{ij}^n's GTC from station i to station j when capacities of link set E_{redu} are reduced by x

$C_e(\omega)$: The functional relation between travel cost on a link and the link's passenger flow ω

C_{ij}^m: The GTC of the path m from station i to station j

Cap_e: The capacity of the link e

c_{ij}: Fare from station i to station j

$E(C_{\text{ij}}, \theta)$: The mathematical expectation of passengers for GTC perception

f_e: The number of trains that traverse link e in a given time period

fl_e: The passenger flow on link e

I_e: The ratio of affected passenger volume who is affected by link e

$n^{\text{trans}}(0,0)$: A passenger's transfer times when the URT system operates normally

n_c: The number of passengers that a train can carry

n_s: The number of seats per train

n_{ij}: The number of shortest paths from station i to station j

n_{ij}^e: The number of shortest paths from station i to station j which contains link e

$|P_1^{\text{ij}}(E_{\text{redu}}, x)|$: The number of passengers trips who travel from station i to station j and belong to category 1

$|\text{pa}_{\text{ij}}|$: The number of paths with spare capacity from station i to station j

$R_{\text{re}}(E_{\text{redu}}, x)$: Metric of reliability of a URT system when capacities of link set E_{redu} are reduced by x

$R_{un}(E_{redu}, x)$: Metric of unreliability of a URT system when capacities of link set E_{redu} are reduced by x

$t^{wait}(0,0)$, $t^{walk}(0,0)$, and $t^{inv}(0,0)$: Waiting time, walking time, and in-vehicle time, respectively, when the URT system operates normally

$t_{ij}^{n}(0,0)$: Passenger p_{ij}^{n}'s perceived travel time from station i to station j when a URT network operates normally

t_{e}^{inv}: The in-vehicle time on link e

v_{ij}: The volume of passenger trip from station i to station j

v_{ij}^{m}: The volume of passenger trips from station i to station j on the path m

x: Link capacity reduction ratio (the unit is %)

ψ_e: The number of seats on link e in a given time period

λ: Relative increase in GTC for passenger p_{ij}^{n} when link capacity reductions occur.

Data Availability

The data used to support the findings of this study are supplied by Yong Yin under license and so cannot be made freely available. The data can be made available from the corresponding author upon request (Yong Yin, yinyong@swjtu.edu.cn).

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFB1200700). The first author was supported by the China Scholarship Council (201907000071).

References

[1] T. Litman, "Impacts of rail transit on the performance of a transportation system," *Transportation Research Record*, vol. 1930, no. 1, pp. 21–29, 2005.

[2] G. Beirão and J. A. Sarsfield Cabral, "Understanding attitudes towards public transport and private car: a qualitative study," *Transport Policy*, vol. 14, no. 6, pp. 478–489, 2007.

[3] M. Cantwell, B. Caulfield, and M. O'Mahony, "Examining the factors that impact public transport commuting satisfaction," *Journal of Public Transportation*, vol. 12, no. 2, pp. 1–21, 2009.

[4] Y. Gu, X. Fu, Z. Liu, X. Xu, and A. Chen, "Performance of transportation network under perturbations: reliability, vulnerability, and resilience," *Transportation Research Part E: Logistics and Transportation Review*, vol. 133, Article ID 101809, 2020.

[5] M. A. P. Taylor, *Vulnerability Analysis for Transportation Networks*, Elsvier, Oxford, UK, 2017.

[6] L.-G. Mattsson and E. Jenelius, "Vulnerability and resilience of transport systems-a discussion of recent research," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 16–34, 2015.

[7] M. G. H. Bell, "A game theory approach to measuring the performance reliability of transport networks," *Transportation Research Part B: Methodological*, vol. 34, no. 6, pp. 533–545, 2000.

[8] H. Kim, C. Kim, and Y. Chun, "Network reliability and resilience of rapid transit systems," *The Professional Geographer*, vol. 68, no. 1, pp. 1–13, 2016.

[9] A. Chen, P. Kasikitwiwat, and C. Yang, "Alternate capacity reliability measures for transportation networks," *Journal of Advanced Transportation*, vol. 47, no. 1, pp. 79–104, 2013.

[10] H. Kim., Y. Song, and Y. Song, "Examining accessibility and reliability in the evolution of subway systems," *Journal of Public Transportation*, vol. 18, no. 3, pp. 89–106, 2015.

[11] Y. Yang, Y. Liu, M. Zhou, F. Li, and C. Sun, "Robustness assessment of urban rail transit based on complex network theory: a case study of the Beijing Subway," *Safety Science*, vol. 79, pp. 149–162, 2015.

[12] C. Jiang, L. Wu, F. Xu, and J. Yuan, "Characteristics and reliability analysis of the complex network in Guangzhou rail transit," *Intelligent Automation & Soft Computing*, vol. 19, no. 2, pp. 217–225, 2013.

[13] Z. Q. Wang and R. H. Xu, "Reliability simulation analysis of urban rail transit networks based on complex network," *Journal of System Simulation*, vol. 21, no. 20, pp. 6670–6674, 2009.

[14] X. Zhang, E. Miller-Hooks, and K. Denny, "Assessing the role of network topology in transportation network resilience," *Journal of Transport Geography*, vol. 46, pp. 35–45, 2015.

[15] J. Liu, H. Lu, H. Ma, and W. Liu, "Network vulnerability analysis of rail transit plans in Beijng-Tianjin-Hebei region considering connectivity reliability," *Sustainability*, vol. 9, no. 8, pp. 2–17, 2017.

[16] Z. Zou, Y. Xiao, and J. Gao, "Robustness analysis of urban transit network based on complex networks theory," *Kybernetes*, vol. 42, no. 3, pp. 383–399, 2013.

[17] Z.-P. Du and A. Nicholson, "Degradable transportation systems: sensitivity and reliability analysis," *Transportation Research Part B: Methodological*, vol. 31, no. 3, pp. 225–237, 1997.

[18] D. M. Scott, D. C. Novak, L. Aultman-Hall, and F. Guo, "Network robustness index: a new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006.

[19] J. Liu, Q. Peng, J. Chen, and Y. Yin, "Connectivity reliability on an urban rail transit network from the perspective of passenger travel," *Urban Rail Transit*, vol. 6, no. 1, pp. 1–14, 2020.

[20] D. Sun., Y. Zhao, and Q.-C. Lu, "Vulnerability analysis of urban rail transit networks: a case study of Shanghai, China," *Sustainability*, vol. 7, no. 6, pp. 6919–6936, 2015.

[21] H. Kim. and Y. Song, "An integrated measure of accessibility and reliability of mass transit systems," *Transportation*, vol. 45, no. 4, pp. 1075–1100, 2018.

[22] Q.-C. Lu, "Modeling network resilience of rail transit under operational incidents," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 227–237, 2018.

[23] A. De-Los-Santos, G. Laporte, J. A. Mesa, and F. Perea, "Evaluating passenger robustness in a rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 20, no. 1, pp. 34–46, 2012.

[24] J. Liu, P. M. Schonfeld, Q. Peng, and Y. Yin, "Measures of travel reliability on an urban rail transit network," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 6, 2020.
[25] M. Snelder, H. J. van Zuylen, and L. H. Immers, "A framework for robustness analysis of road networks for short term variations in supply," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 828–842, 2012.
[26] Y. Jiang, Y. Wang, W. Y. Szeto, A. H. F. Chow, and A. Nagurney, "Probabilistic assessment of transport network vulnerability with equilibrium flows," *International Journal of Sustainable Transportation*, 2020.
[27] O. Cats and E. Jenelius, "Beyond a complete failure: the impact of partial capacity degradation on public transport network vulnerability," *Transportmetrica B: Transport Dynamics*, vol. 6, no. 2, pp. 77–96, 2016.
[28] E. Jenelius, "Public transport experienced service reliability: integrating travel time and travel conditions," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 275–291, 2018.
[29] N. van Oort, "Incorporating service reliability in public transport design and performance requirements: international survey results and recommendations," *Research in Transportation Economics*, vol. 48, pp. 92–100, 2014.
[30] E. I. Diab, M. G. Badami, and A. M. El-Geneidy, "Bus transit service reliability and improvement strategies: integrating the perspectives of passengers and transit agencies in north America," *Transport Reviews*, vol. 35, no. 3, pp. 292–328, 2015.
[31] N. A. Bruzelius, "Microeconomic theory and generalised cost," *Transportation*, vol. 10, no. 3, pp. 233–245, 1981.
[32] F. J. Cesario, "Value of time in recreation benefit studies," *Land Economics*, vol. 52, no. 1, 1976.
[33] L. Lesley, "Generalised costs and the value of time as a method of patronage forecasting," *Acta Technica Jaurinensis*, vol. 2, no. 1, pp. 57–68, 2009.
[34] L. Todd, "Valuing transit service quality improvements," *Journal of Public Transportation*, vol. 11, no. 2, pp. 43–63, 2008.
[35] M. Wardman and G. Whelan, "Twenty years of rail crowding valuation studies: evidence and lessons from British experience," *Transport Reviews*, vol. 31, no. 3, pp. 379–398, 2011.
[36] A. Chen, S. Pravinvongvuth, X. Xu, S. Ryu, and P. Chootinan, "Examining the scaling effect and overlapping problem in logit-based stochastic user equilibrium models," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1343–1358, 2012.
[37] H. Kato, Y. Kaneko, and M. Inoue, "Comparative analysis of transit assignment: evidence from urban railway system in the Tokyo metropolitan area," *Transportation*, vol. 37, no. 5, pp. 775–799, 2010.
[38] H. X. Liu, X. He, and B. He, "Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem," *Networks and Spatial Economics*, vol. 9, no. 4, pp. 485–503, 2009.
[39] M. Dixit, T. Brands, N. Van Oort, O. Cats, and S. Hoogendoorn, "Passenger travel time reliability for multimodal public transport journeys transportation research record," *Journal of the Transportation Research Board*, vol. 2673, no. 2, 2019.
[40] T. Litman, *Transportation Cost and Benefit Analysis*, p. 31, Victoria Transport Policy Institute, Victoria, Canada, 2009.
[41] O. Cats and E. Jenelius, "Vulnerability analysis of public transport networks: a dynamic approach and case study for Stockholm," in *Proceedings of the 5th International Symposium on Transportation Network Reliability (INSTR2012)*, pp. 18-19, Hong Kong, China, December 2012.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 1796132, 18 pages
https://doi.org/10.1155/2020/1796132

Research Article

Evaluating Railway Operation Safety Situation in China Based on an Improved TOPSIS Method: A Regional Perspective

Xu Yan[1,2,3], **Qiyuan Peng**[1,2,3], **Yong Yin**[1,2,3], **Yongxiang Zhang**[1,2,3] and **Qingwei Zhong**[1,2,3]

[1]*School of Transportation & Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China*
[3]*National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 610031, China*

Correspondence should be addressed to Yong Yin; yinyong@swjtu.edu.cn

Received 7 February 2020; Revised 18 September 2020; Accepted 7 October 2020; Published 17 October 2020

Academic Editor: Prakash Ranjitkar

The evaluation of the railway operation safety situation is important for managers to ensure transportation safety and make control decisions. In this study, first, six situation indicators are designed from a regional perspective based on the characteristics of railway operation accidents, and the quantitative methods of these indicators are determined. Second, an improved technique for order of preference by similarity to ideal solution (TOPSIS) method is developed to evaluate the railway operation safety situation against the situation indicators. Based on the set-pair analysis (SPA) theory and cosine similarity measure (CSM), the comprehensive evaluation values and rankings of the safety situation are first calculated from the distance and trend levels for each period. Game theory is then employed to determine a more reasonable combined weight, and the values of the parameters involved in the situation indicators are also estimated. The real-life statistical accident data in a regional area of China from 2016 to 2018 are chosen as a case study to verify the proposed method. A brief analysis is conducted, resulting in suggestions for the evaluation results. Two groups of comparative experiments are designed to demonstrate the feasibility and effectiveness of the method. Finally, the quality of the evaluation results is verified through actual conditions and expert scoring. Some extensions and potential practical applications of this work are discussed.

1. Introduction

Safety is always the first priority of railway operation. By the end of 2018, the railway operating mileage in China reached 131,000 km and the railway network density was 0.01369 km/km^2. In addition, the total converted turnover was as high as 3,986.495 billion ton-km [1]. As a result, China Railway needs to undertake heavy and large-scale transportation tasks each day. At the same time, the complex operation characteristics and environment, such as continuous and dynamic train running processes, require operators to perceive and control the operational safety situation at the macroscopic level. Moreover, the goal of controlling the safety situation is to constantly strengthen railway operation safety and improve the security system. Railway operation management in China is divided by region. Therefore, railway operation safety situation evaluation (ROSSE) from a regional perspective is of great significance for operators to prevent and control operation risks, assist the decision-making process, improve the level of safety management, and ensure continuous and normal activities.

Safety situation evaluation fuses data and information at a high level, including time and space dimensions. In particular, safety situation evaluation has been widely used in military [2], road traffic [3, 4], energy [5], and other fields involving safety management [6]. With the rapid growth of passenger flow in China, the railway system, as a medium- and long-distance public transport mode, faces new challenges in safety

management, especially considering the heterogeneous transportation activities and complicated network structure on a regional scale. At present, research on railway operation safety management mainly covers (i) multifactor integrated management and (ii) accident analysis and modeling. The former mostly analyzes the influencing factors of safety situations by integrating multiple elements, such as humans, equipment, organizations, and the environment. Several theoretical frameworks and indicator systems have been established [7–10]. By analyzing and modeling accident data, the latter are mostly used to design indicators of safety situations, determine the causes of accidents, and predict the state [11–14]. With the fast-paced development of data analytic technology, railway managers in China are paying more attention to the applications of accident statistics, which provide useful, high-level views of safety situations [15]. Therefore, in this study, situation indicators were designed starting from the analysis of the multiattribute characteristics of accidents, where the ROSSE is performed based on these indicators through appropriate methods.

In terms of indicator design, many Chinese researchers select the number of accidents, the monthly average accident rates, and the accident-correlated coefficients as the main indicators to measure the safety situation [16–18]. The design of these indicators is mainly aimed at a certain type of operating line. However, only a few studies simultaneously integrated high-speed lines, normal-speed lines, and passenger and freight transportation within a certain region for holistic research. Therefore, the previous studies have limited practical meaning for managers when there is need to understand the overall safety situation. The number of situation indicators also needs to be supplemented.

In terms of evaluation methods, the multicriteria decision-making (MCDM) approach is widely used. It considers the assessment and ranking of multi-indicator solutions in a certain time series [19, 20]. The most well-known method in the MCDM problem is the technique for order of preference by similarity to ideal solution (TOPSIS) method developed by Hwang and Yoon [21]. According to this technique, the relative distances close to the ideal solution can be seen as the comprehensive evaluation results. The method has achieved good results in practical applications [22–24]. However, because the traditional TOPSIS method calculates Euclidean distance, it is easy to encounter a rank reversal problem (RRP) [25], which can lead to unstable evaluation results. Furthermore, most existing TOPSIS methods do not consider the similarities of trends between alternatives and the ideal solution. Therefore, the main motivation of this study is to develop an improved TOPSIS method to compensate for these two shortcomings.

The TOPSIS method needs to be applied in conjunction with the method of calculating indicator weights. The rationality of weights is critical to the validity and applicability of the evaluation results. To obtain the appropriate relative importance of indicators, the calculation of weights should combine the benefits of subjective and objective judgment methods. Hence, this is another focus of this study. In view of the above discussion, the contributions of this research can be summarized as follows:

(1) Six situation indicators for evaluation were designed considering the multiattribute characteristics of railway accidents for evaluating the railway safety situation from a regional perspective. Given the situation indicators, an improved TOPSIS method is proposed based on the distance and trend levels. Meanwhile, game theory (GT) is applied to determine the combined weight.

(2) A real-life case study based on the actual data of railway accidents in China is introduced, where the situation indicators and the ROSSE are calculated. Comparative experiments and analyses were performed to verify the effectiveness and evaluation quality of the proposed evaluation method.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive literature review. The situation indicators for ROSSE are designed and quantified from a regional perspective, and this is discussed in Section 3. In Section 4, an improved TOPSIS method along with preliminary estimation of the values of the parameters involved are discussed. Section 5 presents a case study that is based on the real-life accident statistics in a certain regional area of the China Railway, including the calculation of indicators, results, and ranking of the ROSSE, model comparison, and performance. In Section 6, the quality of the evaluation work is further verified, and the extensions and potential practical applications of the proposed method are discussed. Conclusions and possible further research directions are provided in Section 7.

2. Literature Review

Following the research ideas of the ROSSE and the process of approach formulation, in this section, the literature is reviewed from three aspects: (i) railway safety management and evaluation, (ii) the TOPSIS method and its improvement, and (iii) combined weight.

2.1. Railway Safety Management and Evaluation. Li and Guldenmund [6] provided a broad overview of safety management systems in 2018. They studied the literature, focusing on five core aspects: definition, evolution, models, purpose, and common elements. Based on this research, the methods of safety management can be carried out using (i) organizational models and (ii) accident-related models. Moreover, railway safety organizations mainly manage multiple factors, such as human behavior and the operating environment. Therefore, a brief review of these two directions was conducted.

For multifactor integrated management, a framework was presented in Morant et al. [8] to evaluate the safety and availability of the railway operation. They quantified the probability that railway traffic is not supervised by the signalling system. Hu et al. [10] established an index system for evaluating the high-speed railway of the environmental safety situation in China by analyzing the impact mechanism of severe weather. Wang et al. [26] applied the cusp catastrophe model to describe the dynamic changing process

of railway system safety, and a framework for the system risk was constructed. Crawford et al. [27] aimed to raise awareness of potential health and safety risks emerging in the railway industry. They found two mechanisms of risk that can impact the success of an integrated railway: human-automation design and progressive integration efforts.

In terms of accident analysis and modeling, Kyriakidis et al. [7] presented a framework to identify the most-significant human performance factors, which were derived from an analysis of 479 railway operational incidents. Ouyang et al. [11] applied an approach called "system-theoretic accident models and process" to analyze railway accidents and discussed the accident spreading processes. According to Li and Wang [12], the causal factors of accidents in a railway system are divided into several error types, and a model was proposed based on a complex network for risk monitoring. The risks of accident causal factors were quantified. Mirabadi and Sharifian [13] analyzed the data from past accident data of the Iranian Railway by applying association rule data-mining techniques to discover and reveal unknown relationships and patterns in data. This research considered accident conditions and relationships discovered among the most common accident factors. Using a Bayesian network model, Bearfield et al. [28] made local risk estimates from boarding and alighting incidents at railway stations, and they analyzed the data with expert judgments about causal factors. Liu et al. [29] focused on employing the fault tree analysis method combined with quantitative analysis to present a more comprehensive view of high-speed railway accidents. In China, 407 railway accident/incident reports were collected and studied by Zhou and Lei [30] using the human factor analysis and classification system framework. The results showed that the four most common errors in the railway system were "organization process," "inadequate supervision," "personal readiness," and "skill-based errors."

Accident data provide a wealth of resources to discover the causes, correlations, spreading mechanisms, and comprehensive performance to measure the railway safety situation or other characteristics. Combining big data analytics to handle this moderate size of data, one can better understand the safety situation. In terms of problems specific to China, the railway operation safety situation has attracted the attention of some researchers. For example, Wei et al. [16] divided different operational safety levels of normal-speed railway lines using the system clustering method. Railway operation safety situation indicators and their spatiotemporal distribution characteristics were analyzed by Xu et al. [17]. They proposed a safety situation prediction model for a high-speed railway based on the back-propagation neural network (BPNN). Gao et al. [18] introduced the industrial data classification method to explore the adaptability of BPNN and gray theory further in safety situation prediction.

For evaluation from a regional perspective, comprehensiveness and integrity are emphasized. Yang et al. [4] designed a series of qualitative and quantitative indices to determine the overall safety level of two highways in China. Laureshyn et al. [31] proposed a framework for organizing all traffic encounters into a severity hierarchy to describe the safety situation and trade-off between safety and efficiency in the traffic system. Relatively few studies have addressed this aspect in the field of railway safety evaluation.

2.2. TOPSIS Method and Its Improvement. In the past few decades, the application of the MCDM approach has dramatically increased in the areas of evaluation theories and assessment methodologies. According to Zavadskas et al. [19], the ROSSE of each period based on multiple situation indicators is a discrete MCDM problem. The ranking of decision-making choices can also be treated as a result of comprehensive evaluation. The TOPSIS method is a classic MCDM approach and a privileged technique in multiple solutions. It is created by selecting the best choice with the shortest Euclidean distance to the positive ideal solution (PIS) and the farthest Euclidean distance from the negative ideal solution (NIS), respectively. The traditional TOPSIS method can be used directly for evaluation purposes. For instance, Huang et al. [23] applied the entropy-TOPSIS method to evaluate urban rail transit system operation performance through an index system with eight indicators and 41 subindicators. Despite the great growth and evolution of the TOPSIS method, it has an RRP, for which De Farias Aires et al. [25] have given an interpretation. The RRP means that a decision-making preference ordering between two alternatives changes when an alternative is added or removed, which clearly contradicts the principle of independence from irrelevant alternatives. The RRP in the TOPSIS method results from the normalization procedure as well as the modifications in the PIS and NIS. Therefore, many researchers have improved the traditional TOPSIS method.

The most effective way to solve the RRP is to replace the calculation of Euclidean distance, such as the Mahalanobis distance used by Wang and Wang [32]. More studies have introduced other relatively complex theories to improve the TOPSIS method. Mousavi-Nasab and Sotoudeh-Anvari [33] integrated TOPSIS and data envelopment analysis (DEA) methods as an auxiliary tool to solve the material selection problem. Liu et al. [34] improved the risk evaluation in failure mode and effect analysis combining a cloud model with the TOPSIS method. Considering that the fuzzy-MCDM is a crucial topic in expert system and operations research, Salih et al. [22] conducted a survey on the state of the art of the fuzzy TOPSIS methods between 2007 and 2017. They further clarified the applicability and limitations of the method. Li et al. [24] applied the fuzzy TOPSIS method to evaluate the service quality of the Beijing metro system and analyzed the stability of the results by the ranking change for a line from different comparison sets of metro lines.

The frequently used normalization methods for TOPSIS were summarized [35]. It was proved that vector normalization does not change the diversity of the attribute data, and it is suitable for improving TOPSIS. This is also the reason for choosing the set-pair analysis (SPA) theory to construct the connection vectors. Herein, the application of SPA theory in the TOPSIS method is reviewed further. Kumar

and Garg [36] attempted to rate the different preferences of an object using the connection number, a major component of the SPA. An extension of the TOPSIS method was developed, based on the proposed connection number, to calculate the relative closeness of sets of alternatives. They further discussed the application of the SPA-TOPSIS method under an interval-valued intuitionistic fuzzy set environment. It was verified through a real-life numerical example in the literature [37]. Yuan and Luo [38] developed a comprehensive system of evaluation criteria and an improved SPA-TOPSIS method to assess the regional energy security performance in China from 2013 to 2017.

The above research demonstrates that improvement is mainly concentrated on the relative distance. Therefore, the cosine similarity measure (CSM) is introduced in this work to study the trend characteristics of solutions. The CSM is a simple and effective metric for learning similarity, and it has been widely used in the fields of machine learning, pattern recognition, and fuzzy strategic decision-making [39]. Ye [40] proposed a weighted CSM based on intuitionistic fuzzy sets, and Wei [41] applied weighted cosine function similarity measures in picture fuzzy sets to select the optimal production strategy. The applications in fuzzy decisions guided the formulation of the approach used in this study.

2.3. Combined Weight. For the TOPSIS approach used for the ROSSE, the weights of the situation indicators reflect the relative importance of different indicators in the decision-making process. In general, there are two categories of weighting methods: subjective and objective. The subjective methods determine weights solely according to the preferences or judgments of decision-makers. By contrast, the objective methods determine weights by solving mathematical models automatically without any consideration of subjective preferences, such as natural weight and Shannon entropy weight. As discussed by Wang [42], there are certain limitations in considering a single weighting method in the MCDM approach. The subjective weight is strongly influenced by expert experiences that could also contain some prejudices. In contrast, the objective weight neglects the knowledge of the decision-makers and the actual situation. Therefore, the comprehensive weight, combining the subjective and objective weights with an effective algorithm, is more reasonable in the approach formulation.

GT, focusing on the research of strategic interaction, is used to obtain the optimal equilibrium solution. In many studies on MCDM methods, GT has been chosen as the preferred method to determine the combined weight. Lai et al. [43] applied a combined weight integrating subjective and objective weights based on GT in a fuzzy comprehensive evaluation. They also showed the rationality of the results. Analogously, the fuzzy analytic hierarchy process (AHP) weight and entropy weight were combined by Sun et al. [44]. In addition, Liu et al. [45] improved the method for determining the combination weight using GT together with the decision-making trial and evaluation laboratory approach, which adjusted the weights of criteria to make the result more reasonable. Although the combined weight has been used frequently, it is generally composed of one subjective weight and one objective weight, and less research has been performed on the combinations of multiple weights.

In order to increase the comprehensiveness and readability of the literature review, we have summarized some of the publications mentioned above and divided them into two parts: (i) publications that are relevant on the railway safety management in Table 1 and (ii) publications that are relevant on the TOPSIS method in Table 2.

3. Design and Quantification of Situation Indicators from a Regional Perspective

Forming a series of evaluation indicators after analyzing the characteristics of railway operation accidents is one of the ways to measure the safety situation effectively. For the current status of organizing railway operations by region in China, the design of indicators should have the following principles: (i) Situation indicators should be universal within an operational region, covering all routes and transportation of passengers and freight. (ii) Situation indicators should make full use of the multiattribute characteristics of the accident data to reflect the safety situation and avoid duplication between indicator functions. (iii) Situation indicators should exclude the consideration of some extreme accidents, i.e., accidents that rarely occur but can lead to serious consequences. Principle (iii) is to consider that the goal is to develop an evaluation method for the safety situation that does not consider extreme cases and is more meaningful and practical. Such extreme accidents would make it difficult to measure and take advantage of the multiattribute characteristics of accidents, except for the severity. In addition, the laws and features of these accidents have obvious uncertainties [46]. Following the above principles and based on the most commonly used methods in the literature, six situation indicators were designed for ROSSE from a regional perspective and corresponding quantitative formulas were obtained. Each indicator was designed for the same period, such as a month.

3.1. Number of Accidents: **A**. In this study, the term "railway accidents" refers to all accidents that affect the normal operations of the locomotive and rolling stocks or other processes in a certain region. It can be directly adapted to reflect the safety situation. In China, railway accidents are divided into three categories, i.e., traffic, off-rail, and labor accidents. In particular, off-rail accidents involve collisions between trains and other types of vehicles or pedestrians. Therefore, A_t can be quantified as follows:

$$A = A_{\text{traffic}} + A_{\text{off-rail}} + A_{\text{labor}}. \quad (1)$$

3.2. Number of Accidents per Unit Turnover: **T**. The amount of turnover is generally used to reflect the transportation workload comprehensively because railway operational products include both transportation objects and distance. It

Table 1: Summary of relevant publications on railway safety management.

Subjects	Publications	Main contents	Methods	Weights	Data
Multifactor analysis and evaluation	[7]	Identify the most significant human performance shaping factors	Factors identification framework	—	Accident statistics
	[8]	Evaluate operational safety and the availability of signalling systems	Improved Markov model	—	State monitoring of signalling systems
	[10]	Evaluate environmental safety through establishing the impact index system of weather	Attribute recognition model	Natural attribute weight	Environmental statistics
Analysis and mining of accident characteristics	[11]	Analyze the accident and its spreading processes	System-theoretic accident models and process	—	Accident reports (4·28 China-Jiaoji)
	[12, 30]	Accident causation analysis	Factors identification, analysis, and classification model	—	Accident/incident reports
	[13]	Discover and reveal relationships and patterns among accidents	Association rules mining techniques	Subjective weight	Accident records
	[29]	Present a more comprehensive analysis of the accident	Fault tree and quantitative analysis	Expert scoring weight	Accident reports (7·23 China-Yongwen)
Safety state prediction	[14, 17]	State prediction based on the representative index system	Neural network	—	Accident statistics
Risk management	[26]	Describe the dynamic changing process of system safety	Cusp catastrophe model	—	Accident records
	[27]	Raise awareness of potential safety risks	Risk mechanism analysis	—	Related literature
	[28]	Local risk estimation	Bayesian network	—	Incidents records
Safety situation evaluation	This paper	Evaluate operational safety situation from a regional perspective	SPA-TOPSIS-CSM	Combined weight (subjective, natural, and entropy)	Accident statistics

is also used by operators to create future operation plans and economic assessments. Therefore, it is important to analyze the relationship between turnover and accidents from a workload perspective. Furthermore, T can be an effective supplement when the number of accidents is only accessed from the time dimension. Considering that passenger and freight transportation are both included in the regional area, the turnover here is the total weighted-sum turnover of those two modes of transportation with a unit of 100 million ton-km. The quantification method is as follows:

$$T = \frac{A}{\left(T_{\text{freight}} + \alpha T_{\text{passenger}}\right)}, \tag{2}$$

where T_{freight} and $T_{\text{passenger}}$ represent freight and passenger turnover, respectively, and α is the conversion weight and takes the value of 1 according to the current statistical system of railways in China.

3.3. Accident Grades: **G**. For reasonable ROSSE, in addition to the number of accidents, one should also consider the severity of different accidents, the scope of influence, and the ability of operators to control the situation which are comprehensively represented as the accident grades. The Railway Accidents Investigation and Handling Rules [47] in China (hereinafter referred to as the "Handling Rules") divide accidents into four categories: especially major accidents, major accidents, slightly major accidents, and general accidents. Among them, the especially major, major, and slightly major accidents are extreme accidents that should not be included during evaluation. For general accidents, the Handling Rules further classify them into four grades from high to low severity, abbreviated as A, B, C, and D. Therefore, only four types of general accidents are considered. The specific quantification method of G is given in

$$G = \sum \beta_o A_o, \quad o = 1, 2, 3, 4, \tag{3}$$

where o represents the types of general accident and takes values of 1 to 4 corresponding to A to D, respectively, A_o represents the number of accidents of type o, and β_o is the severity grade parameter of the type o accident and is yet to be estimated.

3.4. Structure of Accident Properties: **S**. Clarifying the accident properties is an important part of the investigation. Specifically, accidents with different properties can have

Table 2: Summary of relevant publications on the TOPSIS method.

Subjects	Publications	Main contents	Methods	Weights	Data
Overviews	[19]	Describe the situation with reviews of MCDM methods	—	—	Related literature
	[20, 22]	Development, classification, and comparison of TOPSIS methods	Clustering and comparative analysis	—	Related literature
Traditional TOPSIS	[23]	System operation performance evaluation	TOPSIS method based on the Euclidean distance	Entropy weight	Operational statistics
Fuzzy-TOPSIS	[24, 42]	Introduce fuzzy numbers to solve the uncertainty problem	TOPSIS method with fuzzy theory	Combined weight (subjective and entropy)	Surveys
SPA-TOPSIS	[36, 37]	Introduce connection number of SPA and IFS to handle the uncertainty problem	SPA-TOPSIS-IFS	Subjective weight	Numerical examples
	[38]	Evaluate energy security performance from a regional perspective	SPA-TOPSIS	MTGS weight	Environmental statistics
Comprehensive application	[33]	Provide a comprehensive framework for solving the material selection problem	TOPSIS-COPRAS-DEA	Multiple weights (subjective, entropy, and digital logic)	Numerical examples
	[34]	Develop a novel integrated model to improve the conventional FMEA technique	FMEA-TOPSIS-cloud model	Subjective weight	Numerical examples
Safety situation evaluation	This paper	Evaluate operational safety situation from a regional perspective	SPA-TOPSIS-CSM	Combined weight (subjective, natural, and entropy)	Accident statistics

IFS, intuitionistic fuzzy set; MTGS, Mahalanobis–Taguchi Gram–Schmidt method; COPRAS, complex proportional assessment; FMEA, failure mode and effect analysis.

different types of hazard sources, amounts of damage that can be caused, and identification or control measures. The severity of different accident properties can be distinguished from the efforts to control, the level of difficulty for early warning, and potentially dangerous hazards. For instance, more efforts are required to guarantee and control the safety of high-speed passenger train operations. Thus, for the structure of accident properties over a period of time, the smaller the proportion of accidents on high-speed passenger trains, the better the safety situation. Considering this, the structure of accident properties S was designed as a situation indicator, combining the composition and conversion parameters of different accident properties. Equation (4) specifies the calculation method of S:

$$S = \sum_{q=1}^{n} \left(\frac{\lambda_q A_q}{A} \right), \quad q = 1, 2, \ldots, n, \; n = 5, \tag{4}$$

where q represents the five categories of detailed accident properties according to the Rules, including high-speed train accidents, bullet train accidents, ordinary train accidents, operation line construction accidents, and operation standardization accidents; A_q is the number of accidents of property q; and λ_q is the conversion parameter to be estimated, that is, influence relation of the structure of accident properties on the regional safety situation.

3.5. Degree of Accident Correlations: **C**. Regional rail transportation can simultaneously involve several different operation activities, such as trains running at different speeds, heterogeneous operating lines, passenger boarding or alighting, and freight loading or unloading. Hence, the occurrence of one accident can easily affect normal operation activities and directly cause new accidents or become the inducement of subsequent accidents. For example, unexpected train accidents can lead to insufficient connection capacity of rolling stock, which may cause delays or disruptions if handled inappropriately [48]. Therefore, with the increase in accidents, it is possible to explore the degree of accident correlations to measure the stability and robustness of regional safety situations. The calculation method for this situation indicator C is shown as follows:

$$C = \sum_{p=1}^{A} \left(\frac{A_p^{\text{correlation}}}{A} \right), \tag{5}$$

where $A_p^{\text{correlation}}$ represents the number of accidents that are affected by the accident p.

3.6. Average Recovery Time of Accidents: **R**. The recovery time of an accident is the time taken from the beginning to the complete elimination of the adverse effects of the accident. It reflects the ability and efficiency of the railway

system to resume normal operations. On this basis, the average recovery time of all accidents can better indicate this situation from the perspective of the entire region. For instance, a shorter average recovery time means that the entire system has a more optimized maintenance scheduling or advanced incident handling technology [49]. As a result, R is used as the last situation indicator of the design and its quantification method is provided as follows:

$$R = \sum_{r=1}^{A} \left(\frac{T_r^{\text{recovery}}}{A} \right), \tag{6}$$

where T_r^{recovery} represents the time required to recover from accident r.

Thus far, six situation indicators have been designed for ROSSE from a regional perspective. For intuitive display, the information of the indicators is summarized in Table 3.

4. Approach Formulation and Parameter Estimation

As discussed previously, the set of designed situation indicators was used to create the ROSSE of each time period. An appropriate MCDM approach for evaluation is proposed in this section. More than 60 MCDM methods have been applied previously [19]. However, it is challenging to choose a suitable MCDM technique for a given problem. Mousavi-Nasab and Sotoudeh-Anvari [33] conducted a detailed literature review and analysis of this selection problem. They pointed out that an increasing number of studies have used at least two MCDM methods for decision-making-related content and emphasized the straightforwardness and understandability of technique application. Mulliner et al. [50] concluded that the choice of a method should be adaptable to the goal. In the present research, each indicator can only measure one aspect of the overall safety situation, and there may be a certain trade-off relationship between indicators. For example, an increase in the number of accidents may reduce the average recovery time of accidents because of the effective rescue measures. Therefore, the comprehensiveness and integrity of the evaluation results are of the highest value.

Fortunately, TOPSIS has an easier calculation procedure than other well-known MCDM techniques, such as AHP and the analytic network process (ANP). It is also a compensatory method and provides clear trade-offs among the criteria. This good compromise between various indicators can better serve the goals of this work. However, the shortcomings of the traditional TOPSIS method cannot be ignored, as mentioned in Section 2.2. To this end, an improvement strategy combining two different methods was formulated to make the work self-contained.

On the one hand, SPA theory was used to construct the connection vectors of the situation sets in all periods. Thus, one can replace the Euclidean distance calculated in the TOPSIS method by calculating the distance of the connection vectors between the alternatives and the ideal solution. On the other hand, the CSM was employed to estimate the degree of consistency at the trend level between the situation set vectors and the ideal solution vectors. The vectors are all transformed with the NIS as the origin. It is worth mentioning that CSM is also a commonly used metric to replace the Euclidean distance and usually provides good results; however, it is not always advantageous [39]. For instance, two sets with different attribute values may have a very high similarity measure. This outcome is undesirable. Therefore, these two metrics are applied together to compensate for their shortcomings, instead of completely replacing one of them.

Mousavi-Nasab and Sotoudeh-Anvari [33] summarized the commonly used techniques, such as AHP, Shannon's entropy, and the Simos approach, to determine the indicator weight. Expert scoring and entropy weighting are the most popular subjective and objective methods for combination. However, in determining the objective weight, the entropy weight method focuses on the discreteness or uncertainty of the information involved in an indicator, but it ignores the numerical value of the indicator. Therefore, natural weight that focuses more on the value of indicator was introduced to avoid this bias. GT was applied to determine the combined weight of the subjective, natural, and entropy weights.

The above processes of improvement were organized in steps, described as follows. The parameters involved in the situation indicators were preliminarily estimated. The implementation details are illustrated in Figure 1.

4.1. Evaluation Model Based on the Improved TOPSIS Method

Step 1. Construction and dimensionless processing of evaluation matrix

First, an initial evaluation matrix X is constructed that contains m time periods according to the six situation indicators described in Section 3, where the railway operation safety situation of any period can be described by the situation set $\{A, T, G, S, C, R\}$. For convenience, variable x_{ij} is introduced to represent the j-th indicator in the i-th period uniformly, where $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and $n = 6$:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{\text{mn}} \end{bmatrix} = \left[x_{ij} \right]_{\text{mn}}. \tag{7}$$

Then, the matrix X is dimensionless processed by the min-max method to obtain a standardized evaluation matrix Y. If the indicators are of the benefit type, the calculation for standardization can be expressed as equation (9). Otherwise, if the indicators are of the cost type, the calculation for standardization can be formulated as equation (10):

Table 3: Names, symbols, quantification methods, and units for all situation indicators.

Indicator	Symbol	Quantification method	Unit
Number of accidents	A	$A = A_{\text{traffic}} + A_{\text{off-rail}} + A_{\text{labor}}$	—
Number of accidents per unit turnover	T	$T = A/(T_{\text{freight}} + \alpha T_{\text{passenger}})$	—/hundred million ton-kilometers
Accident grades	G	$G = \sum \beta_o A_o$	—
Structure of accident properties	S	$S = \sum_{q=1}^{n} (\lambda_q A_q / A)$	—
Degree of accident correlation	C	$C = \sum_{p=1}^{A} A_p^{\text{correlation}} / A$	—
The average recovery time of accidents	R	$R = \sum_{r=1}^{A} (T_r^{\text{recovery}} / A)$	Hour

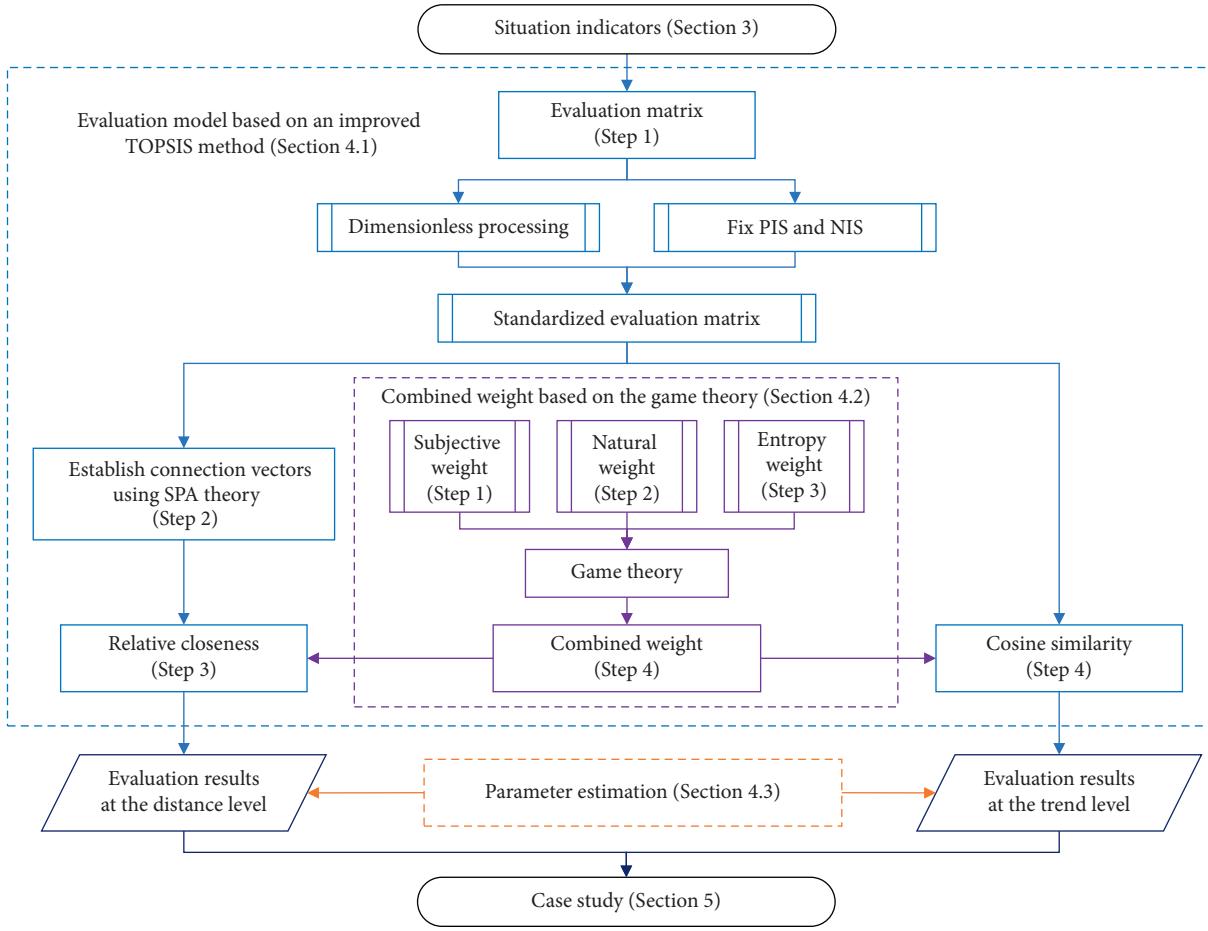


Figure 1: Implementation framework of the improved TOPSIS method.

$$Y = \left[y_{ij}\right]_{\text{mn}}, \tag{8}$$

$$y_{ij} = \frac{x_{ij} - \min_j\left(x_{ij}\right)}{\max_j\left(x_{ij}\right) - \min_j\left(x_{ij}\right)}, \tag{9}$$

$$y_{ij} = \frac{\max_j\left(x_{ij}\right) - x_{ij}}{\max_j\left(x_{ij}\right) - \min_j\left(x_{ij}\right)}, \tag{10}$$

where y_{ij} represents the dimensionless form of x_{ij} and $y_{ij} \in [0, 1]$ and $\max_j(x_{ij})$and $\min_j(x_{ij})$ are the maximum and minimum values for the j^{th} column indicators in X, respectively.

After dimensionless processing, the values of the PIS set Y^+ and NIS set Y^- of the standardized evaluation matrix are fixed, as shown in equation (11), which can be used to simplify the construction of subsequent models:

$$\begin{cases} Y^{+} = \{1, 1, 1, 1, 1, 1\}, \\ Y^{-} = \{0, 0, 0, 0, 0, 0\}. \end{cases} \tag{11}$$

Step 2. Establishment of connection vectors using SPA theory

The basic definition and description of the SPA theory can be found elsewhere [51]. In this study, the SPA theory mainly works on the set pairs between the situation set of each period and the corresponding PIS or NIS set, which are denoted as $Y_i \sim Y^{+}$ and $Y_i \sim Y^{-}$. Considering the "identity," "discrepancy," and "contrary" features of each set pair, the connection degrees are calculated as follows:

$$\begin{aligned} \eta_i^{+} &= a_i^{+} + b_i^{+} h + c_i^{+} k = \sum_{j=1}^{n} \omega_j \eta_{ij}^{+}, \eta_{ij}^{+} = a_{ij}^{+} + b_{ij}^{+} h + c_{ij}^{+} k, \\ \eta_i^{-} &= a_i^{-} + b_i^{-} h + c_i^{-} k = \sum_{j=1}^{n} \omega_j \eta_{ij}^{-}, \eta_{ij}^{-} = a_{ij}^{-} + b_{ij}^{-} h + c_{ij}^{-} k, \end{aligned} \tag{12}$$

where η_i^{+} and η_i^{-} are the connection degrees of $Y_i \sim Y^{+}$ and $Y_i \sim Y^{-}$ in the i-th period, respectively, and ω_j represents the combined weight of each indicator in the situation set, which is determined in Section 4.2. In addition, $\omega_j \in [0, 1]$ and $\sum_{j=1}^{n} \omega_j = 1$; η_{ij}^{+} and η_{ij}^{-} are the subconnection degrees between y_{ij} and the corresponding j-th column indicators in Y^{+} and Y^{-}. The specific calculation rules can be further elaborated as

$$\eta_{ij}^{+} \begin{cases} \text{when } y_{ij} = 0,\ a_{ij}^{+} = b_{ij}^{+} = 0,\ c_{ij}^{+} = 1, \\ \text{when } y_{ij} \in (0, 1],\ a_{ij}^{+} = y_{ij},\ b_{ij}^{+} = 1 - a_{ij}^{+},\ c_{ij}^{+} = 0, \end{cases}$$

$$\eta_{ij}^{-} \begin{cases} \text{when } y_{ij} = 0,\ a_{ij}^{-} = 1,\ b_{ij}^{-} = c_{ij}^{-} = 0, \\ \text{when } y_{ij} \in (0, 1),\ a_{ij}^{-} = 1 - y_{ij},\ b_{ij}^{-} = 1 - a_{ij}^{-},\ c_{ij}^{-} = 0, \\ \text{when } y_{ij} = 1,\ a_{ij}^{-} = b_{ij}^{-} = 0,\ c_{ij}^{-} = 1. \end{cases} \tag{13}$$

Thus, the connection vectors $\overrightarrow{\eta}_i^{+}$ and $\overrightarrow{\eta}_i^{-}$ of $Y_i \sim Y^{+}$ and $Y_i \sim Y^{-}$ in the i-th period can be obtained using equation (14). The connection vectors between the PIS or NIS set and itself are both equal to $(1, 0, 0)$:

$$\begin{cases} \overrightarrow{\eta}_i^{+} = (a_i^{+}, b_i^{+}, c_i^{+}), \\ \overrightarrow{\eta}_i^{-} = (a_i^{-}, b_i^{-}, c_i^{-}). \end{cases} \tag{14}$$

Step 3. Evaluation and ranking at the distance level: relative closeness

The distances D_i^{+} and D_i^{-} of connection vectors in the set pairs $Y_i \sim Y^{+}$ and $Y_i \sim Y^{-}$ are calculated, i.e., the distances between $\overrightarrow{\eta}_i^{+}$ or $\overrightarrow{\eta}_i^{-}$ and vector $(1, 0, 0)$. Based on the TOPSIS method, the relative closeness between the situation set and ideal solution is RC_i^{+} given by equation (16):

$$\begin{cases} D_i^{+} = \sqrt{(1 - a_i^{+})^2 + (b_i^{+})^2 + (c_i^{+})^2}, \\ D_i^{-} = \sqrt{(1 - a_i^{-})^2 + (b_i^{-})^2 + (c_i^{-})^2}, \end{cases} \tag{15}$$

$$\mathrm{RC}_i^{+} = \frac{D_i^{+}}{(D_i^{+} + D_i^{-})}, \quad \mathrm{RC}_i^{+} \in [0, 1]. \tag{16}$$

Thus, the results and ranking of railway operation safety situation can be achieved based on RC_i^{+} at the distance level. The larger the value of RC_i^{+}, the better the railway operation safety situation. It is not difficult to see that the closer the connection vector distance is between the situation set and the PIS, the farther away is the situation set from the NIS. Because the ideal solutions have fixed values, the results and ranking will not change because of the increase or decrease in the number of the situation sets to be evaluated. Therefore, it can effectively solve the RRP.

Step 4. Evaluation and ranking at the trend level: cosine similarity

The idealized vector $\overrightarrow{\rho}^{+}$ is constructed using Y^{+} and Y^{-} with the NIS Y^{-} as the origin. Then, the situation set in the i-th period is transformed into a space vector $\overrightarrow{\varepsilon}_i$ combining the combined weight ω_j. By calculating the cosine similarity CS_i^{+} of vectors $\overrightarrow{\varepsilon}_i$ and $\overrightarrow{\rho}^{+}$, one can measure the consistency between the situation set and idealized target at the trend level. As part of the improved TOPSIS method, it is necessary to sort the calculation results. The larger the value of CS_i^{+}, the smaller the angle between these two vectors, which means that the trend of the railway operation safety situation is better:

$$\begin{aligned} \overrightarrow{\rho}^{+} &= \left(\rho_j | j = 1, 2, \ldots, n\right) = Y^{+} - Y^{-}, \\ \overrightarrow{\varepsilon}_i &= \left(\omega_j \varepsilon_{ij} | j = 1, 2, \ldots, n\right) = \omega_j (Y_i - Y^{-}), \\ \mathrm{CS}_i^{+} &= \cos\left(\overrightarrow{\varepsilon}_i, \overrightarrow{\rho}^{+}\right) = \frac{\left(\sum_{j=1}^{n} \omega_j \varepsilon_{ij} \rho_j\right)}{\left(\sqrt{\sum_{j=1}^{n} \left(\omega_j \varepsilon_{ij}\right)^2} \sqrt{\sum_{j=1}^{n} \rho_j^2}\right)}, \quad \mathrm{CS}_i^{+} \in [0, 1]. \end{aligned} \tag{17}$$

The calculation process can be partially simplified by substituting the values of Y^{+} and Y^{-} as follows:

$$\begin{aligned} \overrightarrow{\rho}^{+} &= (1, 1, 1, 1, 1, 1), \\ \overrightarrow{\varepsilon}_i &= \omega_j Y_i. \end{aligned} \tag{18}$$

4.2. Combined Weight Based on GT

Step 1. Determine subjective weight ψ_j

The expert evaluation method is a common method for obtaining subjective weights. Several operators were

invited to score each situation indicator according to the importance of their influence on the railway operation safety situation, and the average value of each situation indicator was calculated to obtain the weight ψ_j. This weight is given directly in the case study, and it is stipulated that

$$\sum_{j=1}^{n} \psi_j = 1, \quad \psi_j \in (0,1). \tag{19}$$

Step 2. Determine natural weight μ_j

The natural weight objectively reflects the importance of the index in the evaluation to a certain extent. The variable y_{ij} represents the attribute value of the j^{th} situation indicator in the i^{th} period in the standardized evaluation matrix Y. Therefore, the quantification method for the natural weight μ_j of each indicator is as follows:

$$\mu_j = \frac{\sqrt{\sum_{i=1}^{m} y_{ij}^2}}{\sum_{j=1}^{n} \sqrt{\sum_{i=1}^{m} y_{ij}^2}}, \quad \mu_j \in (0,1). \tag{20}$$

Step 3. Determine entropy weight φ_j

The entropy weight method calculates the indicator weights based on the information entropy (IE) of indicators, which was originally developed by Shannon [52]. Generally, a smaller value of IE indicates a higher diversity of the corresponding data and a larger weight, and vice versa [53]. After normalizing y_{ij} into $y_{ij}^{'}$ in the standardized evaluation matrix Y, the IE E_j for each indicator can be defined as follows:

$$y_{ij}' = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}}, \tag{21}$$

$$E_j = 1 + \frac{1}{\ln m} \sum_{i=1}^{m} \left(y_{ij}' \ln y_{ij}' \right). \tag{22}$$

Moreover, the entropy weight φ_j obtained from IE is finally expressed as

$$\varphi_j = \frac{E_j}{\sum_{j=1}^{n} E_j}. \tag{23}$$

Step 4. Determine combined weight ω_j

In GT, multiple types of weight can reach a compromise through a process of dispersion minimization [44]. After obtaining the above three weights, one can calculate the combined coefficient δ of each weight, which decides the method of linear combination of different weights. If the combination coefficients of the three weights are δ_ψ, δ_μ, and δ_φ, the optimal equilibrium weight vector ω of the combined weight can be expressed as

$$\omega = \delta_\psi \psi + \delta_\mu \mu + \delta_\varphi \varphi, \quad \delta_\psi > 0,\ \delta_\mu > 0,\ \delta_\varphi > 0, \tag{24}$$

where ψ, μ, and φ are the weight vectors of each weight—for example, $\psi = (\psi_j | j = 1, 2, \ldots, n)$.

To minimize the deviation between ω and the other three weights, the optimized combined coefficients must be solved by

$$\begin{pmatrix} \psi\psi^T & \psi\mu & \psi\varphi \\ \mu\psi & \mu\mu^T & \mu\varphi \\ \varphi\psi & \varphi\mu & \varphi\varphi^T \end{pmatrix} \begin{pmatrix} \delta_\psi \\ \delta_\mu \\ \delta_\varphi \end{pmatrix} = \begin{pmatrix} \psi\psi^T \\ \mu\mu^T \\ \varphi\varphi^T \end{pmatrix}, \tag{25}$$

Therefore, the final combined weight ω_j is given in

$$\omega_j = \frac{\delta_\psi}{\delta_\psi + \delta_\mu + \delta_\varphi} \psi_j + \frac{\delta_\mu}{\delta_\psi + \delta_\mu + \delta_\varphi} \mu_j + \frac{\delta_\varphi}{\delta_\psi + \delta_\mu + \delta_\varphi} \varphi_j. \tag{26}$$

4.3. Estimation of Parameters Involved in Situation Indicators. The parameters β_o and λ_q involved in the situation indicators must be preliminarily estimated according to relevant laws, regulations, studies, and expert experience. Because the evaluation method adopts a sequential optimization technique based on the similarity to the ideal target, the values of the parameters focus on the ratio relationship rather than the numerical value. Therefore, the estimated parameters were normalized for the convenience of calculation. The specific estimation process of these two parameters is as follows:

(1) Severity grade parameter: β_o

The value of parameter β_o mainly depends on the relative relationship between the severity of different grades of accidents. To this end, some excerpts from the Handling Rules were selected for comparison and are listed in Table 4. There is a quantitative contrast relationship between grades A and B based on the fuzzy value range. This relationship can provide a basis for determining the ratio of the conversion factors between them. Grades C and D contain more descriptions of the accident scenarios without a clear quantitative relationship. Grade C is more inclined to passenger accidents and train operation accidents, whereas grade D is more targeted at freight accidents and shunting operation accidents. Therefore, quantitative and qualitative analyses were combined to estimate the parameters. The results are presented in

$$\beta_o = (\beta_1, \beta_2, \beta_3, \beta_4) = (0.5, 0.25, 0.15, 0.1). \tag{27}$$

(2) Property conversion parameter: λ_q

Table 4: Some useful criteria for parameter estimation in the Handling Rules.

Grades	Criteria
A	2 people died; 5–10 people were seriously injured; direct economic losses of 5–10 million yuan; 3–6 h of interruption of busy trunk traffic, etc.
B	1 person died; 5 people were seriously injured; 100–500 million yuan directly economically lost; busy trunk line was interrupted for 1–3 h, etc.
C	No casualties; direct economic losses below 1 million yuan; train conflicts, derailment, collisions, etc.; passenger train separation; trains sent to occupied section or line, arrival or departure route is not ready or mistakes are made handling train blocking
D	No casualties; direct economic losses of less than 1 million yuan; shunting conflicts, derailment, collisions, etc.; freight train separation; mishandling in operation due to construction, overhaul; delayed train violation of regulations, etc.

Parameter λ_q reflects the impact of different accident properties on the railway safety situation. Wang [54] conducted practical research on the accident properties in China through an association rule analysis of operation accident data. The correlation between the safety hazard and the accident properties was quantified using gray theory. The results of this study can be used as a basis for parameter estimation. It is necessary to make some appropriate adjustments to it with reference to the comments of operators. Finally, the values of λ_q are determined as follows:

$$\begin{aligned}\lambda_q &= (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \\ &= (0.34, 0.28, 0.20, 0.12, 0.06).\end{aligned} \tag{28}$$

5. Case Study

A real-life case study was conducted by taking a certain railway operating region in China as a background. The statistics of accidents in this region from 2016 to 2018 were taken as the initial input data, and the time unit for each period was set as one month. The original accident statistics are not presented in this study because of confidentiality reasons. Considering the limited size of the case, Microsoft Excel was used as a simple and effective tool to achieve the indicator calculations and model solving.

5.1. Calculation and Analysis of Situation Indicators. The various situation indicators for each evaluation month were calculated according to equations (1)–(6), and the calculation results are listed in Table 5. To understand and compare the trend of the situation indicators better, Figure 2 was drawn as a supplement considering only the value fluctuations.

Combining Table 4 and Figure 2 shows that (i) the fluctuations of indicators A, T, and G are relatively stable and exhibit a downward trend as a whole, although increasing slightly in the last few months, and (ii) compared with the fluctuating trends of indicators A, T, and G, the other three situation indicators have more obvious fluctuations, especially the indicator C. In addition, the indicators S and C show an increasing trend, while the indicator R declines steadily from the end of 2017. Judging from the changes in the above situation indicators, the frequency and severity of railway accidents in this region have been effectively controlled and the ability and efficiency of accident rescue have been improved. However, the structure of accident properties and the degree of accident correlations move in a direction that is not conducive to the safety situation. This may be related to the continuous increase in the density of high-speed trains and transportation tasks. Obviously, the scope of impacts and the degree of coeffects caused by accidents are constantly expanding. Thus, disruption management under uncertainty [55] is currently being researched extensively.

5.2. ROSSE Based on the Improved TOPSIS Method. Based on the calculation of the situation indicators, the initial evaluation matrix including 36 months was constructed. All the situation indicators were of the cost type. According to equations (19)–(26), multiple weights and the final combined weight were calculated and are presented in Table 6. The results show that the combined weights of different indicators are relatively close, with the indicator T reaching a maximum value of 0.1812. However, there are still some obvious differences between different types of weight rankings. Combining the weight and parameter values, Table 7 gives the comprehensive evaluation results and their ranks at the distance and trend levels using the improved TOPSIS method.

The overall evaluation results show that the railway operation safety situation in this region was poor in the first nine months of 2016; then, it gradually improved. At the distance level, the maximum relative closeness occurred in June 2018, which means that, during that month, the safety situation was the best. The changing trend of relative closeness was generally on the rise, but it was sometimes accompanied by a sudden increase or decrease. In terms of the trend level, the maximum evaluation value was reached in March 2017. The fluctuation range of the cosine similarity was reduced after October 2016; however, it slightly deteriorated after February 2018. In addition, the values of the situation indicators show that S and C are the main causes of such fluctuations. The overall trends of the indicators are reflected in Figure 3.

The above results suggest that the development of railway operation safety situations in regional areas is generally benign but unstable. With continuous expansion of the operation scale, especially the rapid increase in high-speed

Table 5: Calculation results of situation indicators for ROSSE in a certain region of China from 2016 to 2018.

Year. month	A	T	G	S	C	R
2016.01	52	0.3223	8.60	0.1047	0.32	1.95
2016.02	43	0.2742	7.25	0.1372	1.76	1.99
2016.03	42	0.2471	7.15	0.1427	1.71	1.96
2016.04	50	0.3074	6.85	0.1246	1.62	1.46
2016.05	45	0.2829	7.25	0.0892	0.94	1.85
2016.06	55	0.3653	7.95	0.1154	1.58	1.59
2016.07	57	0.3384	6.90	0.1250	1.65	1.18
2016.08	63	0.3598	7.75	0.1432	1.88	1.20
2016.09	57	0.3522	8.10	0.1600	1.73	1.51
2016.10	38	0.2271	5.95	0.1118	0.87	1.81
2016.11	37	0.2383	5.40	0.0964	0.32	1.59
2016.12	33	0.1982	6.00	0.1633	0.61	2.23
2017.01	45	0.2406	7.80	0.1345	0.46	1.71
2017.02	24	0.1321	4.65	0.1533	0.33	1.88
2017.03	33	0.1865	5.60	0.1514	0.62	1.58
2017.04	29	0.1688	5.60	0.1120	0.28	1.99
2017.05	27	0.1541	5.45	0.2000	0.78	2.10
2017.06	28	0.1690	5.80	0.1954	1.72	2.16
2017.07	16	0.0851	2.80	0.1440	1.06	1.92
2017.08	21	0.1090	4.25	0.1767	2.10	2.08
2017.09	31	0.1793	5.65	0.1440	0.71	1.80
2017.10	25	0.1366	4.95	0.1673	1.08	2.08
2017.11	21	0.1223	3.90	0.1280	1.28	1.86
2017.12	23	0.1352	4.20	0.1813	2.04	1.76
2018.01	26	0.1409	5.05	0.1860	1.42	1.47
2018.02	23	0.1240	4.25	0.1709	1.74	1.30
2018.03	25	0.1175	4.95	0.1963	2.76	1.38
2018.04	26	0.1354	5.25	0.1508	1.12	1.58
2018.05	21	0.1071	3.80	0.1109	1.19	1.55
2018.06	19	0.0971	3.75	0.1300	0.42	1.58
2018.07	25	0.1146	5.10	0.1900	1.96	1.60
2018.08	34	0.1507	6.85	0.1969	1.65	1.47
2018.09	30	0.1492	6.50	0.1618	1.20	1.68
2018.10	25	0.1165	5.20	0.2127	2.24	1.57
2018.11	27	0.1408	4.65	0.1294	1.63	1.45
2018.12	25	0.1322	4.75	0.1700	1.08	1.44

transportation, the causes and suddenness of accidents are becoming increasingly complicated, leading to a more sensitive safety situation of the railway system and resulting in increased difficulty in safety management. Therefore, railway operators should improve the methods of risk perception and causation analysis, focus on strengthening the prevention measures of passenger accidents, and promote the daily evaluation of safety situations. For example, pattern detection methods can be used to analyze the effectiveness of safety control policies when the results of ROSSE change significantly, such as in October 2016 in the case study. More suggestions are obtained by calculating the quartile of the evaluation results to classify the safety situation levels, as shown in Table 8. This can be applied as assistance for operator decision-making in actual operations.

5.3. Comparative Experiments. Two groups of comparative experiments were designed to verify the effectiveness of the proposed method. The experiments were based on the real-life data of the case study:

(1) To verify the elimination of the RRP, data for 16 months were randomly excluded, and the relative closeness was recalculated at the distance level. Table 9 compares the ranking results between the experiment and original case study. The ranking comparison includes 20 months of data, and the ranking size relationship of each month is not reversed.

(2) The traditional entropy-TOPSIS method was employed for the ROSSE as a comparison of the improved TOPSIS method proposed in this study, and the evaluation performance was analyzed. As shown in Figure 4, the improved method is the same as the traditional method for ROSSE. The changing trend of relative closeness is similar and Spearman's rank correlation index (SRCI) [33] of ranking is 0.98, which implies that the two types are closely related. However, the difference is that the value range of the relative closeness obtained by the improved method is nearly 20% larger than that obtained by the traditional method. This indicates that the improved method is more sensitive to situation awareness. Thus, the evaluation results are more practical for detecting slight changes in the safety situation.

The above two experiments were mainly aimed at comparing the relative closeness at the distance level. In addition, the improved method in this study considers the cosine similarity at the trend level. The SRCI of ranking between these two levels is 0.74, which means that there is a certain difference between them. Therefore, it makes sense to take advantage of the cosine similarity to improve the evaluation model. Together, they make the ROSSE more comprehensive and reasonable.

6. Discussion

6.1. Quality Verification. In general, it is challenging to measure the quality of comprehensive evaluation results because access to relevant information is limited. A small number of studies achieved this verification by comparing it with the outcomes in existing or similar assessment reports [43]. However, to the best of our knowledge, this is the first time that ROSSE has been performed from a regional perspective, and therefore, the effect of comparative verification is significantly weakened. Nevertheless, the Railway Safety Announcements from 2016 to 2018 [56] were collected. These reports were analyzed, and the main conclusions are as follows:

(1) The reports show that the entire safety situation of the national railway has gradually improved from 2016 through 2018, especially for the significant reduction of accident rates and accident mortality rates. This is consistent with the overall trend of the evaluation results.

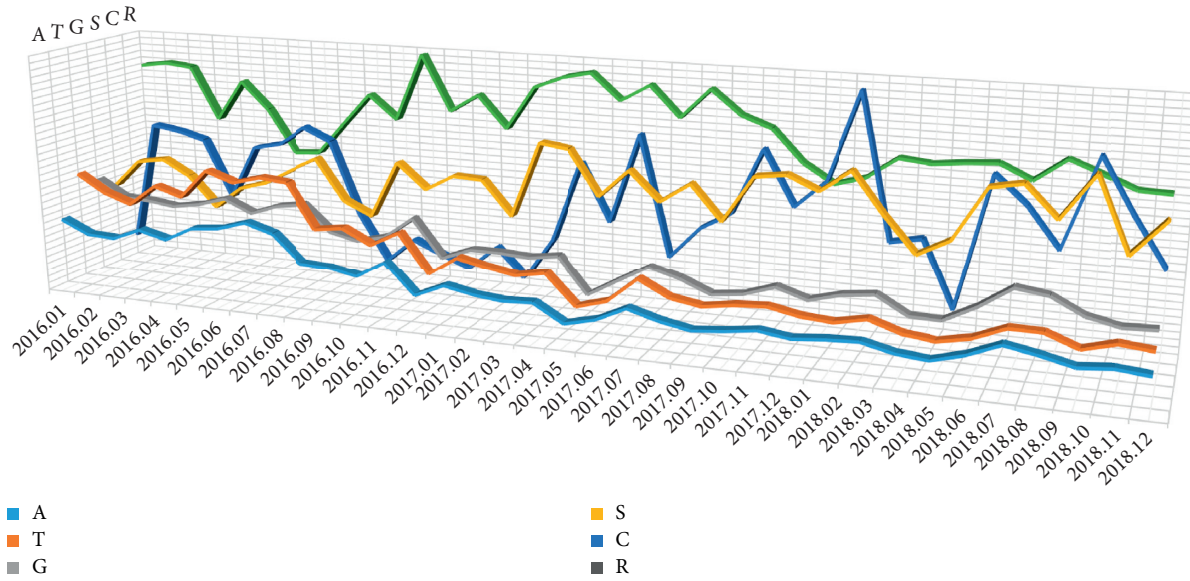


Figure 2: Value fluctuations of situation indicators.

Table 6: Values of weights for situation indicators.

Weight/indicator	A	T	G	S	C	R
Subjective weight ψ_j	0.1263	0.1614	0.2010	0.1859	0.1566	0.1688
Natural attribute weight μ_j	0.1842	0.1863	0.1490	0.1532	0.1752	0.1521
Entropy weight φ_j	0.1377	0.1900	0.1741	0.1855	0.1239	0.1888
Combined weight ω_j	0.1463	0.1812	0.1737	0.1765	0.1449	0.1774

Table 7: Results and ranking of ROSSE in a certain region of China from 2016 to 2018.

t	D_i^+	D_i^-	RC_i^+	Ranking	CS_i^+	Ranking
2016.01	0.7477	0.5745	0.4345	29	0.7544	35
2016.02	0.8931	0.5211	0.3685	35	0.9387	18
2016.03	0.8595	0.5547	0.3922	32	0.9595	13
2016.04	0.7737	0.6405	0.4529	28	0.8917	25
2016.05	0.7117	0.6169	0.4643	27	0.8712	27
2016.06	0.8074	0.5098	0.3870	33	0.7674	34
2016.07	0.7713	0.5612	0.4212	30	0.7940	32
2016.08	0.8300	0.5003	0.3761	34	0.6929	36
2016.09	0.9910	0.4232	0.2993	36	0.7768	33
2016.10	0.6051	0.8091	0.5721	16	0.9668	9
2016.11	0.4585	0.9557	0.6758	9	0.9627	11
2016.12	0.6535	0.6725	0.5072	23	0.8933	24
2017.01	0.7147	0.6995	0.4946	26	0.9171	20
2017.02	0.4575	0.9567	0.6765	8	0.9625	12
2017.03	0.5355	0.8787	0.6213	12	0.9909	1
2017.04	0.4920	0.8388	0.6303	11	0.9486	14
2017.05	0.7088	0.7054	0.4988	24	0.8738	26
2017.06	0.8244	0.5898	0.4171	31	0.8571	29
2017.07	0.3520	0.9369	0.7269	3	0.9389	17
2017.08	0.6476	0.7667	0.5421	22	0.8594	28
2017.09	0.5673	0.8469	0.5989	14	0.9862	3
2017.10	0.6168	0.7974	0.5639	19	0.9223	19
2017.11	0.4261	0.9881	0.6987	4	0.9635	10
2017.12	0.6061	0.8081	0.5714	17	0.9143	21
2018.01	0.5495	0.8647	0.6114	13	0.9455	15
2018.02	0.4422	0.9720	0.6873	6	0.9425	16
2018.03	0.5561	0.7847	0.5852	15	0.8396	31
2018.04	0.4833	0.9309	0.6583	10	0.9846	4
2018.05	0.2919	1.1223	0.7936	2	0.9839	5
2018.06	0.2540	1.1602	0.8204	1	0.9891	2
2018.07	0.6069	0.8073	0.5709	18	0.9083	22
2018.08	0.7109	0.7033	0.4973	25	0.8958	23
2018.09	0.6192	0.7950	0.5622	20	0.9675	8
2018.10	0.5909	0.7395	0.5558	21	0.8464	30
2018.11	0.4350	0.9792	0.6924	5	0.9781	6
2018.12	0.4568	0.9574	0.6770	7	0.9700	7

(2) Although the trend is optimistic, the reports indicate that the number of exposed security risks has increased exponentially. Deepening the environmental governance along the railway, especially for high-speed railways, will become the focus of transportation safety supervision. This also means that the safety situation is still unstable, which confirms the results of this study at the trend level.

In addition to the above macro conclusions, experts were invited to further score the quality of the entire evaluation work. Concerning the literature on performance evaluation [23, 38], the scoring items were designed based on mainly three aspects, that is, indicator, methodology, and result.

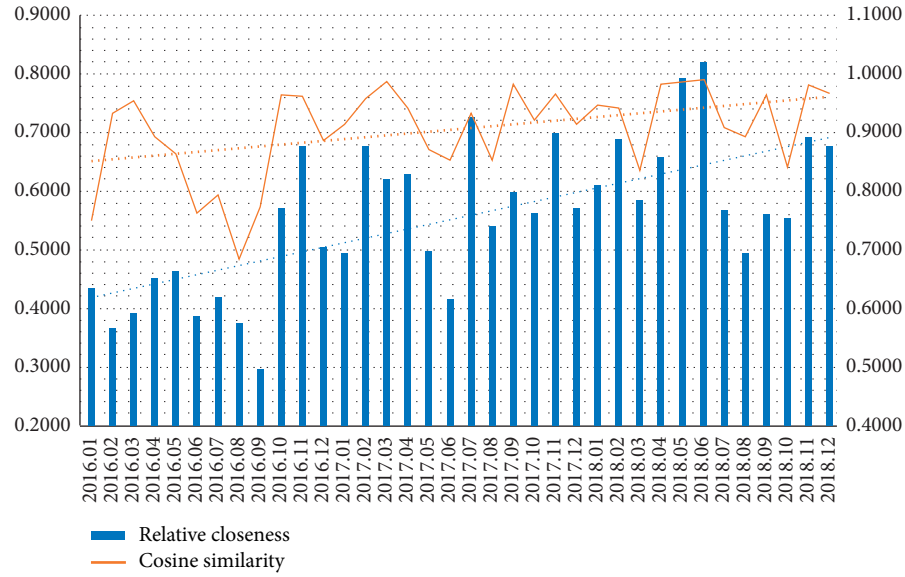


Figure 3: ROSSE results and overall trends.

Table 8: Divisions of safety situation levels.

Safety situation level	Distance: relative closeness	Trend: cosine similarity
A	$0.6627 \leq RC_i^+$	$0.9643 \leq CS_i^+$
B	$0.5674 \leq RC_i^+ < 0.6627$	$0.9305 \leq CS_i^+ < 0.9643$
C	$0.4615 \leq RC_i^+ < 0.5674$	$0.8682 \leq CS_i^+ < 0.9305$
D	$RC_i^+ < 0.4615$	$CS_i^+ < 0.8682$

Table 9: Ranking comparison of relative closeness at distance level.

Serial number	1	2	3	4	5	6	7	8	9	10
Original ranking	24	31	3	22	14	19	4	17	13	6
Comparative ranking	18	20	3	17	10	14	4	12	9	6
Serial number	11	12	13	14	15	16	17	18	19	20
Original ranking	15	10	2	1	18	25	20	21	5	7
Comparative ranking	11	8	2	1	13	19	15	16	5	7

In terms of the indicator, the main consideration is the rationality of indicator design (RID) and the completeness of the indicator system (CIS). In addition, the effectiveness of the indicator system in reflecting the regional perspective (ERP) is required. The effectiveness of the methodology was verified. Hence, the straightforwardness (STR) and understandability (UND) described in Section 4 are important goals. Furthermore, the scalability of the method (SCM) is one of the key capabilities to respond to the changes in practical applications. Finally, the rationalities of weight calculation results (RWC), evaluation results at the distance level (RED), and evaluation results at the trend level (RET) are included in the scoring item design.

Eight experts engaged in railway operation management scored each item from 1 to 10. The higher the score, the higher the recognition of the corresponding item. All scores of eight experts and the overall average score are illustrated in Figures 5(a) and 5(b), respectively.

The scores show that experts are positive about the quality of the evaluation work. The average scores of RWC, RED, and RET are all close to 8, which indicates good performance. However, CIS and SCM still need to be strengthened. These two scoring items are both closely related to the extensions of the method; therefore, they are discussed further in the next section.

6.2. Method Extensions. The safety of railway operations depends on various factors, including infrastructure and rolling stock reliability, emergency rescue rules, natural environment monitoring, and human factors. If these elements are to be considered fully, the evaluation indicator system required must be huge and complex. Because this work is from a regional perspective, it has been restricted in the design of accident-based indicators. Hence, the extensions of the indicator system mainly consider the inclusion of other relatively independent elements, such as natural environment monitoring and human factors.

Although railway transportation is less affected by the natural environment than other modes, up to nearly 80% of passenger transport accidents are caused by bad weather or sudden natural disasters [10]. Environmental factors that are likely to cause accidents and equipment failure include heavy rainfall, lateral wind, lightning, earthquakes, ice disasters, and dramatic changes in temperature differences. However, unlike accident-based indicator design, the impact of environmental factors is uncertain. Therefore, this category of indicators should reflect the changes brought about by the environment to safety risks—especially the impact of sudden natural disasters on the safety situation.

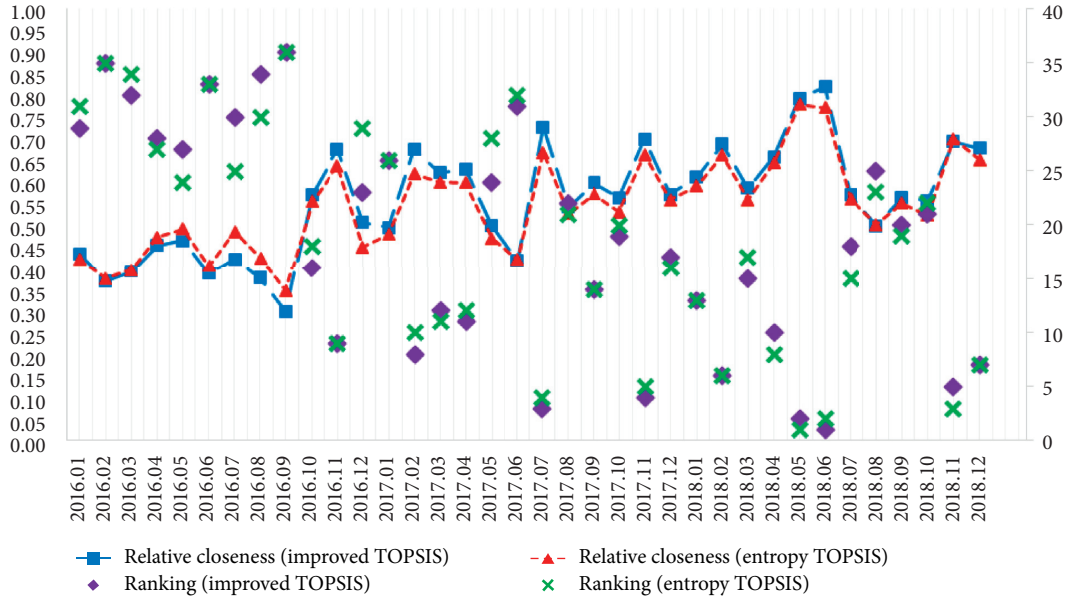


Figure 4: Comparison of evaluation results at the distance level between the improved TOPSIS method and entropy-TOPSIS method.

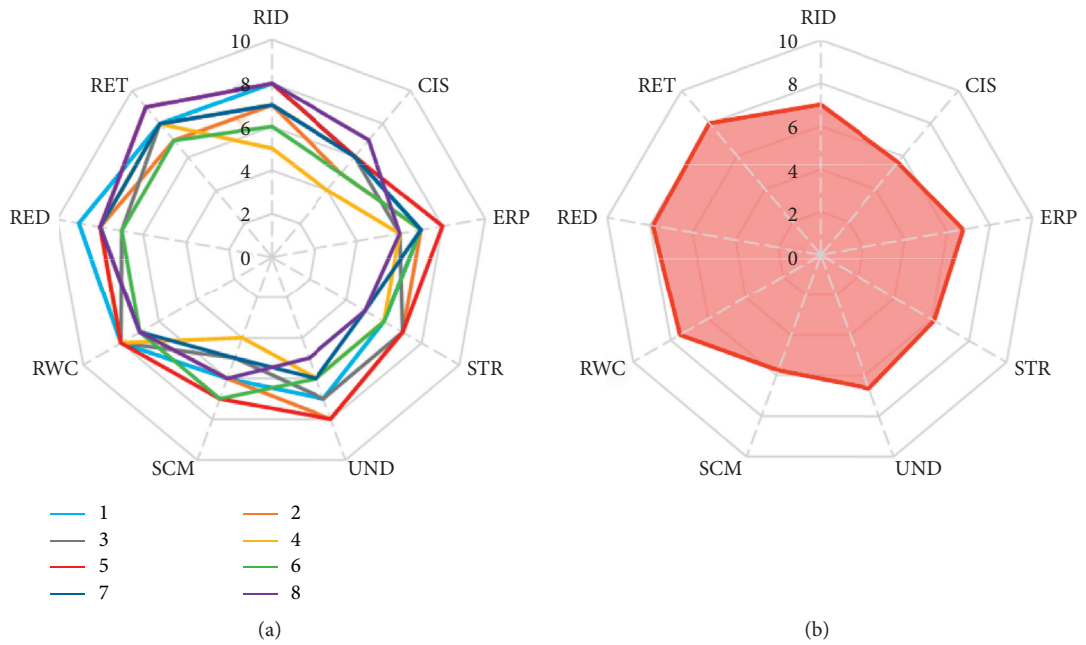


Figure 5: Radar charts of scoring by experts: (a) all scores of eight experts and (b) average score.

The extension of the indicator system based on human factors is also similar to that based on environmental factors. According to Zhou et al. [30], four errors with the largest percentage of occurrence in the railway system are organization process, inadequate supervision, personal readiness, and skill-based errors. These human factors indirectly affect the overall safety situation through state changes at the supervisory and organizational levels. Therefore, human errors cannot be directly applied to the TOPSIS method, as in the accident-based approach, to deal with safety situations. Thus, a reasonable conversion process is necessary.

Considering the above possible extension direction of the indicator system, operation managers can encounter more incomplete or unobtainable information. For example, the influence on the safety situation of decision mistakes in human factors cannot be accurately evaluated using real numbers. To deal with these uncertainties and ambiguities, fuzzy theory is suitable for integration as an extension

direction. This combined method can solve imprecise information by converting it into linguistic variables while retaining the effectiveness in decision making.

6.3. Practical Applications. With the increased availability of data, railway operation safety analysis and management techniques based on accidents or incidents play an important role. Operation managers can better perceive safety situations and discover potential risks through accident data mining. However, because of different purposes, different levels of managers have different needs for safety information acquisition. Generally, grassroots managers are more concerned about the specific safety performance and changes in each link of the system, whereas senior managers pay more attention to the overall safety situation of the system and the key features exposed. The comprehensive evaluation from a regional perspective in this study mainly serves the latter.

In actual safety management, railway senior managers are usually required to make comprehensive and guided decisions. For example, in China, railway managers in charge of a regional area are responsible for coordinating and supervising dozens of operation lines and hundreds of transportation resources. Faced with a large amount of available information, they need to avoid falling into data traps and make top-level decisions directly and efficiently. Therefore, the basis for making such decisions must also be regionally comprehensive and capable of judging trends.

The improved method proposed herein can appropriately meet the requirements in this regard. Distance-level evaluation results can help managers intuitively judge the performance of the safety situation throughout the entire historical period. Trend-level evaluation results can enable managers to make better judgments about decision-making directions. The TOPSIS method also has advantages in avoiding paying too much attention to the original situation sets. The division of the safety situation levels, as shown in Section 5.2, is also an effective practical application.

7. Conclusions

In this paper, six situation indicators were designed from a regional perspective and an improved TOPSIS method was proposed for the ROSSE. Starting from the multiattribute analysis of railway accidents, the safety situation can be characterized by the situation set of indicators, including the number of accidents, number of accidents per unit turnover, accident grades, structure of accident properties, degree of accident correlations, and average recovery time of accidents. Corresponding quantification methods were given, and the parameters involved in the indicators were preliminarily estimated.

In terms of the evaluation model, the traditional TOPSIS method was improved by combining SPA theory and the CSM. However, the SPA theory was used to introduce the connection vector to calculate the relative closeness as the evaluation result at the distance level. Through the vectorization of the situation set, the cosine similarity between the situation vector and ideal vector was calculated as the evaluation result at the trend level. Moreover, a combined weight based on GT was applied in the evaluation process. Then, a case study using the proposed method was performed on railway accident statistics in a certain regional area of China from 2016 to 2018. The evaluation results show that the overall safety situation has undergone benign development, but it is still unstable. It is necessary to strengthen the measures of safety management by improving the relative closeness and limiting the fluctuations of cosine similarity. The control measures of high-speed railway passenger transportation should also be considered.

The effectiveness of the improved TOPSIS method was verified through two groups of comparative experiments. The results illustrate that the RRP was effectively solved and that the improved method is more sensitive to situation awareness. The measurement of cosine similarity can also help to better understand and evaluate the railway operation safety situation. A quality verification of the entire evaluation work was conducted through the analysis of actual reports and expert scorings. Method extensions and practical applications were also discussed. In addition, evaluation work can be further combined with other transportation organization or disruption management research to formulate collaborative optimization strategies, such as robust train timetabling [57], train platforming [58], rolling stock scheduling [59], and energy saving [60]. The method of parameter estimation needs to be improved.

Data Availability

The original accident data used to support the findings of this study are currently under embargo. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (U1834209) and the National Key Research and Development Plan of China (2017YFB1200700). The authors would like to thank Editage (http://www.editage.cn) for English language editing.

References

[1] China State Railway Group Co., Ltd., *China Railway Corporation 2018 Statistical Bulletin*, China-Railway Net, Beijing, China, 2019.

[2] M. Martínez-Córcoles and K. Stephanou, "Linking active transactional leadership and safety performance in military operations," *Safety Science*, vol. 96, pp. 93–101, 2017.

[3] C. Wang, C. Xu, J. Xia, Z. Qian, and L. Lu, "A combined use of microscopic traffic simulation and extreme value methods for

traffic safety evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 281–291, 2018.

[4] Y. Yang, S. M. Easa, Z. Lin et al., "Evaluating highway traffic safety: an integrated approach," *Journal of Advanced Transportation*, vol. 90, p. 11, Article ID 4598985, 2018.

[5] F. H. Gandoman, J. Jaguemont, S. Goutam et al., "Concept of reliability and safety assessment of lithium-ion batteries in electric vehicles: basics, progress, and challenges," *Applied Energy*, vol. 251, 2019.

[6] Y. Li and F. W. Guldenmund, "Safety management systems: a broad overview of the literature," *Safety Science*, vol. 103, pp. 94–123, 2018.

[7] M. Kyriakidis, A. Majumdar, and W. Y. Ochieng, "Data based framework to identify the most significant performance shaping factors in railway operations," *Safety Science*, vol. 78, pp. 60–76, 2015.

[8] A. Morant, A. Gustafson, P. Söderholm, P.-O. Larsson-Kråik, and U. Kumar, "Safety and availability evaluation of railway operation based on the state of signalling systems," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 231, no. 2, pp. 226–238, 2017.

[9] H. Song and E. Schnieder, "Evaluating Fault Tree by means of Colored Petri nets to analyze the railway system dependability," *Safety Science*, vol. 110, pp. 313–323, 2018.

[10] Q. Hu, N. Gao, and B. Zhang, "High speed railway environment safety evaluation based on measurement attribute recognition model," *Computational Intelligence and Neuroscience*, vol. 2014, p. 10, Article ID 470758, 2014.

[11] M. Ouyang, L. Hong, M.-H. Yu, and Q. Fei, "STAMP-based analysis on the railway accident and accident spreading: taking the China-Jiaoji railway accident for example," *Safety Science*, vol. 48, no. 5, pp. 544–555, 2010.

[12] K. Li and S. Wang, "A network accident causation model for monitoring railway safety," *Safety Science*, vol. 109, pp. 398–402, 2018.

[13] A. Mirabadi and S. Sharifian, "Application of association rules in Iranian Railways (RAI) accident data analysis," *Safety Science*, vol. 48, no. 10, pp. 1427–1435, 2010.

[14] S. Nefti and M. Oussalah, "A neural network approach for railway safety prediction," in *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, vol. 4, pp. 3915–3920, New York, NY, USA, 2004.

[15] F. Ghofrani, Q. He, R. M. P. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: a survey," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 226–246, 2018.

[16] N. Wei, D. Liu, and Z. Chen, "Operational safety analysis of general speed railway based on system clustering," *China Safety Science Journal*, vol. 28, no. S1, pp. 88–92, 2018.

[17] L. Xu, F. Cao, and N. Gao, "On the safety operation state prediction of high-speed railway based on neural network," *Technology & Economy in Areas of Communications*, vol. 18, no. 6, pp. 35–38, 2016.

[18] N. Gao, Q. Hu, B. Zhang et al., "An accident prediction model of high-speed railway operation based on data classification," *Journal of Transport Information and Safety*, vol. 33, no. 1, pp. 71–78, 2015.

[19] E. K. Zavadskas, Z. Turskis, and S. Kildienė, "State of art surveys of overviews on MCDM/MADM methods," *Technological and Economic Development of Economy*, vol. 20, no. 1, pp. 165–179, 2014.

[20] S. Opricovic and G.-H. Tzeng, "Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, no. 2, pp. 445–455, 2004.

[21] C. L. Hwang and K. P. Yoon, *Multiple Attributes Decision Making Methods and Applications*, Springer, Berlin, Germany, 1981.

[22] M. M. Salih, B. B. Zaidan, A. A. Zaidan, and M. A. Ahmed, "Survey on fuzzy TOPSIS state-of-the-art between 2007 and 2017," *Computers & Operations Research*, vol. 104, pp. 207–227, 2019.

[23] W. Huang, B. Shuai, Y. Sun, Y. Wang, and E. Antwi, "Using entropy-TOPSIS method to evaluate urban rail transit system operation performance: the China case," *Transportation Research Part A: Policy and Practice*, vol. 111, pp. 292–303, 2018.

[24] J. Li, X. Xu, Z. Yao, and Y. Lu, "Improving service quality with the fuzzy TOPSIS method: a case study of the beijing rail transit system," *IEEE Access*, vol. 7, pp. 114271–114284, 2019.

[25] R. F. D. F. Aires and L. Ferreira, "A new approach to avoid rank reversal cases in the TOPSIS method," *Computers & Industrial Engineering*, vol. 132, pp. 84–97, 2019.

[26] Y. Wang, U. A. Weidmann, and H. Wang, "Using catastrophe theory to describe railway system safety and discuss system risk concept," *Safety Science*, vol. 91, pp. 269–285, 2017.

[27] E. G. C. Crawford and R. L. Kift, "Keeping track of railway safety and the mechanisms for risk," *Safety Science*, vol. 110, pp. 195–205, 2018.

[28] G. Bearfield, A. Holloway, and W. Marsh, "Change and safety: decision-making from data," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 227, no. 6, pp. 704–714, 2013.

[29] P. Liu, L. Yang, Z. Gao, S. Li, and Y. Gao, "Fault tree analysis combined with quantitative analysis for high-speed railway accidents," *Safety Science*, vol. 79, pp. 344–357, 2015.

[30] J.-L. Zhou and Y. Lei, "Paths between latent and active errors: analysis of 407 railway accidents/incidents' causes in China," *Safety Science*, vol. 110, pp. 47–58, 2018.

[31] A. Laureshyn, Å. Svensson, and C. Hydén, "Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation," *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1637–1646, 2010.

[32] Z.-X. Wang and Y.-Y. Wang, "Evaluation of the provincial competitiveness of the Chinese high-tech industry using an improved TOPSIS method," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2824–2831, 2014.

[33] S. H. Mousavi-Nasab and A. Sotoudeh-Anvari, "A comprehensive MCDM-based approach using TOPSIS, COPRAS and DEA as an auxiliary tool for material selection problems," *Materials & Design*, vol. 121, pp. 237–253, 2017.

[34] H.-C. Liu, L.-E. Wang, Z. Li, and Y.-P. Hu, "Improving risk evaluation in FMEA with cloud model and hierarchical TOPSIS method," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 84–95, 2019.

[35] P. Chen, "Effects of normalization on the entropy-based TOPSIS method," *Expert Systems with Applications*, vol. 136, pp. 33–41, 2019.

[36] K. Kumar and H. Garg, "Connection number of set pair analysis based TOPSIS method on intuitionistic fuzzy sets and their application to decision making," *Applied Intelligence*, vol. 48, no. 8, pp. 2112–2119, 2018.

[37] H. Garg and K. Kumar, "A novel exponential distance and its based TOPSIS method for interval-valued intuitionistic fuzzy sets using connection number of SPA theory," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 595–624, 2020.

[38] J. Yuan and X. Luo, "Regional energy security performance evaluation in China using MTGS and SPA-TOPSIS," *Science of the Total Environment*, vol. 696, p. 133817, 2019.
[39] P. Xia, L. Zhang, and F. Li, "Learning similarity with cosine similarity ensemble," *Information Sciences*, vol. 307, pp. 39–52, 2015.
[40] J. Ye, "Cosine similarity measures for intuitionistic fuzzy sets and their applications," *Mathematical and Computer Modelling*, vol. 53, no. 1-2, pp. 91–97, 2011.
[41] G. Wei, "Some cosine similarity measures for picture fuzzy sets and their applications to strategic decision making," *Informatica*, vol. 28, no. 3, pp. 547–564, 2017.
[42] T.-C. Wang and H.-D. Lee, "Developing a fuzzy TOPSIS approach based on subjective weights and objective weights," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8980–8985, 2009.
[43] C. Lai, X. Chen, X. Chen, Z. Wang, X. Wu, and S. Zhao, "A fuzzy comprehensive evaluation model for flood risk based on the combination weight of game theory," *Natural Hazards*, vol. 77, no. 2, pp. 1243–1259, 2015.
[44] L. Sun, Y. Liu, B. Zhang, Y. Shang, H. Yuan, and Z. Ma, "An integrated decision-making model for transformer condition assessment using game theory and modified evidence combination extended by D numbers," *Energies*, vol. 9, no. 9, p. 697, 2016.
[45] T. Liu, Y. Deng, and F. Chan, "Evidential supplier selection based on DEMATEL and game theory," *International Journal of Fuzzy Systems*, vol. 20, no. 4, pp. 1321–1333, 2018.
[46] Z. Fan, "Cycle analysis and characteristics for the occurrence of major railway accidents in China," *China Safety Science Journal*, vol. 28, no. S1, pp. 135–140, 2018.
[47] China State Railway Group Co., Ltd., "China's railway accidents investigation and handling rules," China State Railway Group Co., Ltd., Berlin, Germany, 2007.
[48] Q. Zhong, R. M. Lusby, J. Larsen, Y. Zhang, and Q. Peng, "Rolling stock scheduling with maintenance requirements at the Chinese High-speed Railway," *Transportation Research Part B: Methodological*, vol. 126, pp. 24–44, 2019.
[49] Y. Zhang, A. D'Ariano, B. He, and Q. Peng, "Microscopic optimization model and algorithm for integrating train timetabling and track maintenance task scheduling," *Transportation Research Part B: Methodological*, vol. 127, pp. 237–278, 2019b.
[50] E. Mulliner, N. Malys, and V. Maliene, "Comparative analysis of MCDM methods for the assessment of sustainable housing affordability," *Omega*, vol. 59, pp. 146–156, 2016.
[51] K. Zhao, "Set pair and set pair analysis: a new concept and systematic analysis method," *Proceedings of the National Conference on System Theory and Regional Planning*, vol. 12, pp. 87–91, 1989.
[52] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, 1947.
[53] P. Chen, "On the diversity-based weighting method for risk assessment and decision-making about natural hazards," *Entropy*, vol. 21, no. 3, p. 269, 2019.
[54] K. Wang, *Research on Railway Accident Data Mining and Forecasting and Early Warning Method Based on the Correlation analysis*, Beijing Jiaotong University, China, 2016.
[55] E. Van der Hurk, L. Kroon, and G. Maróti, "Passenger advice and rolling stock rescheduling under uncertainty for disruption management," *Transportation Science*, vol. 52, no. 6, pp. 1391–1411, 2018.
[56] "National Railway Administration of the People's Republic of China, ," , 2018.
[57] Y. Zhang, Q. Peng, Y. Yao, X. Zhang, and X. Zhou, "Solving cyclic train timetabling problem through model reformulation: extended time-space network construct and Alternating Direction Method of Multipliers methods," *Transportation Research Part B: Methodological*, vol. 128, pp. 344–379, 2019.
[58] Y. Zhang, Q. Zhong, Y. Yin et al., "A fast approach for reoptimization of railway train platforming in case of train delays," *Journal of Advanced Transportation*, vol. 12, p. 20, Article ID 5609524, 2020.
[59] Q. Zhong, Y. Zhang, D. Wang et al., "A mixed integer linear programming model for rolling stock deadhead routing before the operation period in an urban rail transit line," *Journal of Advanced Transportation*, vol. 12, p. 18, Article ID 3809734, 2020.
[60] W. Li, Q. Peng, C. Wen et al., "Integrated optimization on energy saving and quality of service of urban rail transit system," *Journal of Advanced Transportation*, vol. 12, p. 22, Article ID 3474020, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 9853164, 13 pages
https://doi.org/10.1155/2020/9853164

Research Article

A Green Demand-Responsive Airport Shuttle Service Problem with Time-Varying Speeds

Ming Wei,[1,2] Binbin Jing,[3] Jian Yin,[3,4] and Yang Zang[3]

[1]*School of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China*
[2]*CAAC Key Laboratory of General Aviation Operation (Civil Aviation Management Institute of China), Beijing 102202, China*
[3]*School of Transportation, Nantong University, Nantong 226019, China*
[4]*School of Urban Management, Yuncheng Kindergarten Teachers College, Yuncheng 224005, China*

Correspondence should be addressed to Binbin Jing; jingbin19@126.com

Received 6 February 2020; Revised 22 July 2020; Accepted 21 August 2020; Published 2 September 2020

Academic Editor: Yu Jiang

This study proposes a multiobjective mixed integer linear programming (MOMILP) model for a demand-responsive airport shuttle service. The approach aims to assign a set of alternative fuel vehicles (AFVs) located at different depots to visit each demand point within the specified time and transport all of them to the airport. The proposed model effectively captures the interactions between path selection and environmental protection. Moreover, users with flexible pick-up time windows, the time-varying speed of vehicles on the road network, and the limited fuel for the route duration are also fully considered in this model. The work aims at simultaneously minimizing the operating cost, vehicle fuel consumption, and CO_2 emissions. Since this task is an NP-hard problem, a heuristic-based nondominated sorting genetic algorithm (NSGA-II) is also presented to find Pareto optimal solutions in a reasonable amount of time. Finally, a real-world example is provided to illustrate the proposed methodology. The results demonstrate that the model not only selects an optimal depot for each AFV but also determines its route and timetable plan. A sensitivity analysis is also given to assess the effect of early/late arrival penalty weights and the number of AFVs on the model performance, and the difference in quality between the proposed and traditional models is compared.

1. Introduction

Airports are generally located in the suburbs of a city; therefore, passengers living in urban areas need to be transported to the airport by shuttle buses. To increase the flexibility and accessibility of airport services, an increasing number of traffic authorities in different countries have realized airport shuttle services in a regular bus or subway system as one of the most effective strategies to meet passengers' accessibility needs. Most of the existing shuttle services are fixed-route regular bus or subway systems, which are suitable for high-density residential areas. However, this can often have high operating costs and low service levels in some remote areas with few people. Currently, with the rapid expansion of mobile internet, the demand-responsive airport shuttle services (DRASS) using app-based ride services is a future trend in designing a customized supply-to-demand transit route. Some of the benefits of this may include reduced operational cost and carbon emission and improved mobility [1, 2].

A DRASS transit system, in which passengers place their travel orders through a mobile app to obtain a door-to-door-type airport shuttle service and many vehicle routes located at different depots are assigned to visit these demand points and transport them to the airport, is an important component of a demand-responsive transit system (DRT) [3]. As with DRT, the increase in consciousness about environmental impact has made green DRASS a critical issue. In this case, DRASS route decisions determine operating cost, vehicle fuel consumption, and carbon emissions [4, 5]. Therefore, green DRASS is different from traditional systems, and a study considering the environmental impact on AFVs route design is timely.

The carbon emissions of airport shuttle buses are mainly related to their travel speed and load [6]. Recent research on minimizing emissions in DRASS models can be divided into two main categories. The first is the set of models where vehicle speed is assumed to be constant [1, 2, 7], and the second set includes models where the vehicle speed is time-varying based on different road conditions [5, 8]. Although the first models are not accurate, they are suitable where there is a lack of traffic data. Recently, remote vehicle tracking techniques have been used to collect detailed traffic data on transit times for different roads by time of day and day of the week, which provides the possibility for the second models to accurately estimate carbon emissions. Therefore, it is important to find the optimal relationship between time-varying networks, DRASS route design, and carbon emissions in order to reduce the negative impact on the environment.

The main contribution of this paper is to investigate a multiobjective green DRASS with a time-varying network in order to minimize operating cost, vehicle fuel consumption, and CO_2 emissions. The main research tasks are summarized as follows: (1) coordination of a DRASS transit routing and departure time guidance process based on a time-varying network to balance path selection and environmental protection; (2) development of a heuristic-based NSGA-II algorithm to efficiently obtain a set of Pareto optimal solutions. Finally, a numerical simulation example in the real world is provided to illustrate the optimization approach.

The remainder of this paper is organized as follows: Section 2 reviews the related literature concerning green DRASS; Section 3 details the framework of green DRASS with a time-varying network; Section 4 presents a heuristic method based on NSGA-II for resolving MOMILP; a real-world case study is used to demonstrate the applicability of the proposed approach in Section 5; and some conclusions and possible future work are discussed in Section 6.

2. Literature

Both DRT and DRASS are extensions of the vehicle routing problem (VRP) and the pick-up and delivery problem (PDP), which aim at assigning all customers in demand points to vehicles located at different bus depots and designing routes to transport them from their home or workplace to destinations [1, 2, 5, 8]. The only difference between them is that DRT strengthens the connectivity between residential areas and rail stations, while DRASS transports air passengers to airports [1, 7]. They have similar objective functions (i.e., operating cost, carbon emissions, and passenger satisfaction) and constraints (i.e., time windows, vehicle capacity, and mileage). Fortunately, these factors normally do not change the properties of the problem, and thus, their models and algorithms are universal [8–11]. Because rail transit is more popular than civil aviation, the amount of research on DRT is greater than that of DRASS.

In general, DRT and DRASS often involve customers being picked up within specific time windows, which is related to the customers' satisfaction level in the airport shuttle service [9]. A variant of vehicle routing with time windows has been studied by researchers and can be divided into hard time windows [3, 4, 7, 9] and soft time windows [12–14]. In the former, time windows cannot be violated, where the vehicles must arrive at the earliest arrival time and leave the customers before the latest departure time. However, a little tardiness is acceptable to a person if they can still catch their trip. To deal with the latter issues pertaining to a small deviation from time windows, an early or late penalty cost is calculated once a predetermined time window is unmet. Obviously, DRT and DRASS with soft time windows can reduce fleet size and improve operating efficiency, compared to that of hard time windows. Most existing studies only allow the vehicle to visit customers in a single period of time. Sometimes, passengers may provide multiple time windows, and a vehicle can arrive at the pick-up locations within one of the specified periods. DRT and DRASS with multiple time windows can reduce the operating cost compared to the single-time window problem [9, 15, 16].

Currently, there is an increasing amount of interest in a new variant of DRT and DRASS with time-varying speeds. Due to road traffic congestion depending on the time of day, the speed is not constant. This differs significantly by the hour of the day, by the day of the week, and by the season of the year [5, 17]. Generally speaking, historical data are used to obtain the hypothetical distribution of fuzzy, grey, and random variables [18–20]. However, a piecewise function can describe time-varying speeds more precisely, in the absence of statistics [21]. When vehicle speeds are replaced by traffic volumes, these models are further expanded to time-dependent DRT and DRASS [22, 23]. Also, the vehicle speed may also be used as an additional decision variable in time-dependent models, which is useful in reducing the waiting time for all customers [24, 25]. To avoid the situation where a vehicle with a later departure time arrives before a vehicle with an earlier departure time, these models are constructed based on the first in first out (FIFO) or queuing approach [26].

In recent years, another variant of DRT and DRASS has attracted widespread attention regarding the environment. Here, not only the operational cost is considered but also the energy consumption/CO_2 emissions. Compared to the traditional problem, this would save energy consumption and CO_2 emissions but increase some operating costs. Most research has only focused on energy minimizing [27] or CO_2 emission minimizing vehicle routing [28]. Few have investigated an integrated routing, energy consumption, and CO_2 emission model [21, 29]. These studies were also extended by taking driver wages into account [21, 29]. Furthermore, this research can be divided into time-independent and time-dependent models: in the former, the energy consumption and CO_2 emissions are only related to routes, not schedules; in the latter, carbon emissions also depend on the departure time of the vehicles. Due to the fuel consumption and CO_2 emissions of delivery vehicles being closely related to the driving speed, load and road conditions, and other factors, the result of time-dependent models

is more accurate and reliable than that of time-independent models [5, 21].

As DRT and DRASS belongs to the class of NP-hard problems, two solution methods, i.e., route-building heuristics and evolutionary algorithms, are often used to resolve large-scale problem instances. For route-building heuristics, a set of feasible solution is quickly found, and better solutions can be obtained by further search based on the fine-tuned initial solutions [30–32]. There are two ways to get the initial solution: (1) all customers are grouped into several vehicles, and then, the order in which the vehicles will visit the customers is determined; (2) each route is gradually constructed to visit some customers until the travel demand is met. Some widely used evolutionary algorithms that are often used include the column generation metaheuristic [12], ant colony algorithms [3, 15, 19], Tabu search [12, 16, 17], particle swarm optimization [20], and genetic algorithms [7, 10, 33–35].

To the best of our knowledge, the following important issues deserve further investigation:

(1) Most studies have neglected the integrated operation of depot selection, vehicle scheduling, and routing to transport passengers from home to the airport. Since different depots for each vehicle route influences the time the vehicle arrives and leaves the demand point, the ignorance of assigning optimal depots to the vehicle route often results in increasing the operational costs [36, 37].

(2) Few studies have analyzed the impact of time-varying road conditions and soft time windows on AFV scheduling and routing, which leads to a lack of an integrated solution that balances fuel consumption, CO_2 emissions, and operating costs [32].

(3) DRASS is an NP-hard problem. Although some methods have been explicitly used to solve the two relevant objectives between operating costs and fuel consumption/CO_2 emissions, a heuristic algorithm should be used to efficiently find a set of Pareto solutions to balance all of the three objectives [34, 35].

3. Methodology

3.1. Research Framework. In the DRASS model, passengers will be transported by several AFVs from their workplaces or home addresses to the airport. The location, maximum expected riding time, preferred pick-up time windows, and number of persons on board for each demand point is easily obtained using the mobile app in advance. If an AFV arrives at the passenger's location early or late, there will be a penalty charge. Each AFV, initially located at a depot, has its own limited fuel for the duration of the route. The fuel consumption of an AFV is only related to its speed, vehicle load weight, and mileage [6]. It is obvious that the change in traffic volume in different periods determines the time-varying speed of vehicles on the road, and its rules can be depicted by a piecewise function. The main aim of this study is to assign all demand points to several AFVs, route each AFV starting from a selected depot, visit the points and end at the airport, and determine the time this vehicle arrives and leaves each demand point. A MOMILP model for DRASS is presented to simultaneously minimize operational cost, vehicle fuel consumption, and CO_2 emissions. The key inputs of the model are the travel information of each demand point, endurance and energy consumption of AFVs, and the real travel speed/distance matrix between these vehicle nodes.

To better understand the underlying principles of the DRASS model, a small example consisting of one airport (M), eight demand points ($C1$–$C8$), and two depots ($D1$-$D2$) is shown in Figure 1. The change curves of the average speeds on different roads are shown in Figure 2. The numbers in brackets around each demand point denotes time windows and maximum expected riding time. The red numbers on top of each demand point denotes the number of passengers to be picked up. The number in-line linking two adjacent demand points is the travel distance. All customers in each demand point spend one minute getting on the bus. There are three generated AFV routes in the optimization process, i.e., ($D1$ (7 : 20)–$C5$ (7 : 35, 7 : 36)–$C4$ (7 : 56, 7 : 57)–M (8 : 12)), ($D2$ (7 : 15)–$C2$ (7 : 35, 7 : 36)–$C6$ (8 : 01, 8 : 02)–M (8 : 17)), and ($D2$ (7 : 25)–$C3$ (7 : 40, 7 : 41)–$C1$ (8 : 01, 8 : 02)–M (8 : 12)). For example, vehicle 3 departs from $D2$ at 7 : 25. Due to the speed of the vehicle on road section $D2$–$C3$ during this period being 20 km/h, this vehicle arrives and leaves at $C3$ at 7 : 40 and 7 : 41 with a speed of 15 km/h and loads three persons. Obviously, the vehicle does not arrive at the location within the time window and is about five minutes late. Similarly, the arrival and departure time for the vehicle at $C1$ are (8 : 01, 8 : 02) (nine minutes early), and the vehicle picks up four passengers. Finally, it returns back to M at 8 : 12, and a total of seven passengers are unloaded. In this case, each objective function of all AFV routes can be calculated easily.

To deal with real-life situations, this DRASS model will be based on the following assumptions:

(1) All demand points must be visited by AFVs, and each of them can only be covered once by one AFV.

(2) The unit cost of fuel, the unit cost of carbon emissions, and the per-unit-time reward of vehicles arriving at the demand point normally, early, or late can be estimated.

(3) Biggest continue voyage course of each AFV can be predicted using an intelligent car-carried terminal in advance.

(4) The travel distance matrix between these vehicle nodes as well as time-varying speed on the road can be obtained from an Open GIS tool.

3.2. Model Formulation

3.2.1. Notation. To facilitate model presentation, all definitions and notations used hereafter are summarized in Table 1.

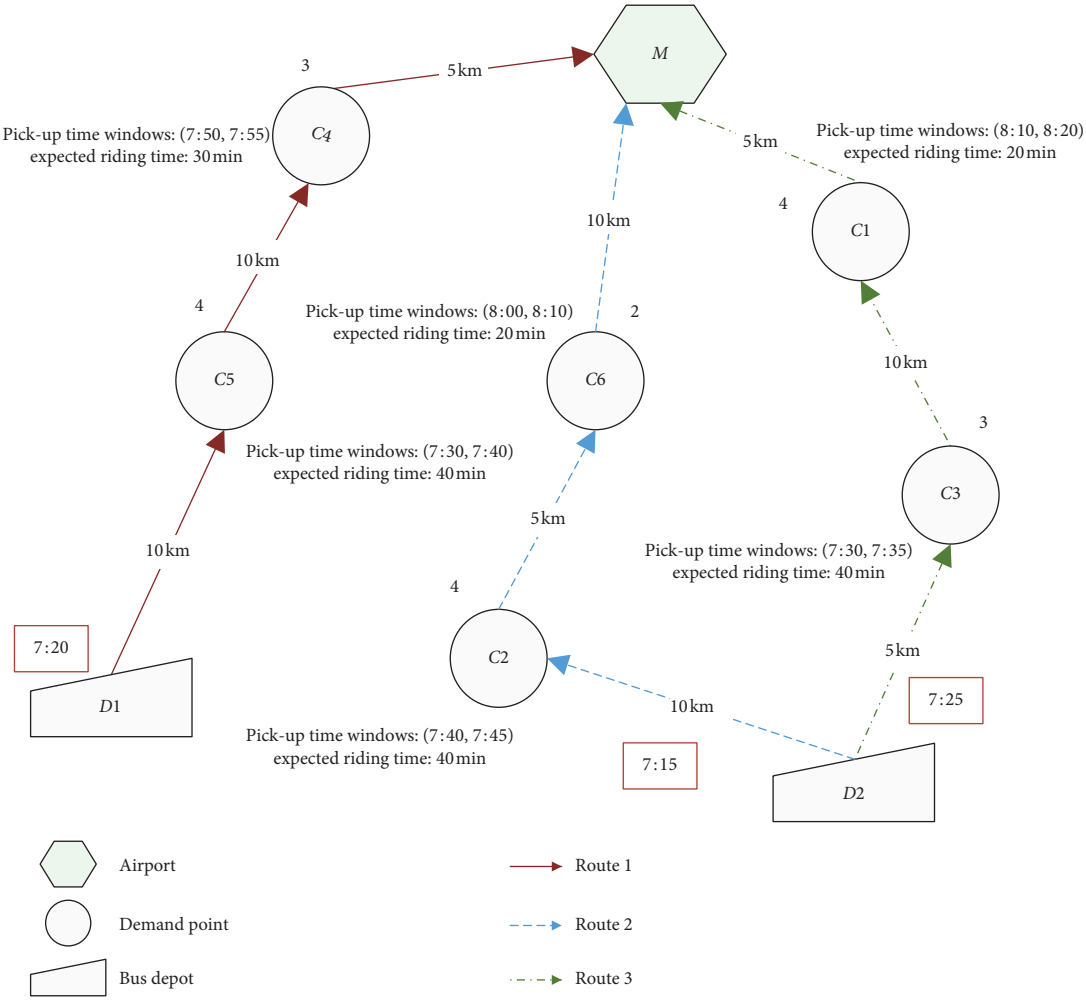


Figure 1: Graphical representation of the DRASS problem.

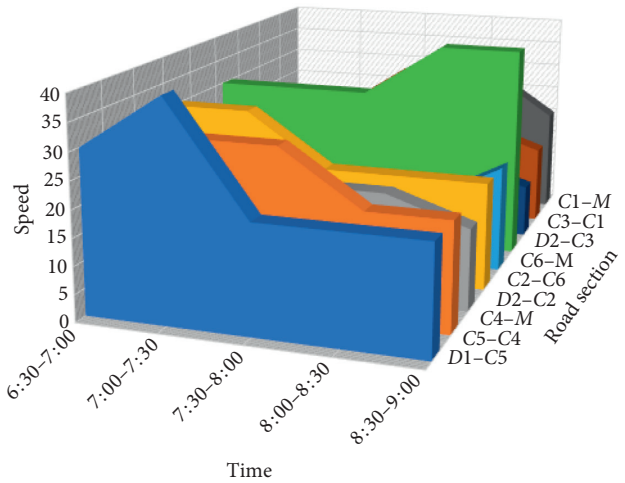


Figure 2: Vehicle average speed change curve on roads in small example.

Table 1: Parameters and variables in the mathematical model.

Indices	
i, j	Vehicular node (demand point, depot, and airport) index
k	AFV index
0	Airport index
Sets	
I	Set of demand points
K	Set of AFVs
D	Set of depots
Parameters	
β	The boarding time per passenger
γ	The average weight of each person
$[a_i, l_i]$	Pick-up time window of demand point i; $\forall i \in I$
Lt_i	The maximum expected ride time of the demand point i; $\forall i \in I$
p_i	Number of persons to be picked up at demand point i; $\forall i \in I$
T_i	Pick-up time of the demand point i, i.e., $T_i = p_i \cdot \beta$; $\forall i \in I$
Q_k	Maximum capacity of the AFV
$D^k_{\max}$	Biggest continue voyage course of each AFV k; $\forall k \in K$
$T_{\min}$	Minimum total travel time of the AFV route
d_{ij}	Travel distance between the vehicular node i and j; $\forall i, j \in I \cup D \cup 0$
$v_{ij}(t^k_i)$	The piecewise function of travel speed between the vehicular node i and j, related to the time of the AFV k visiting demand point i; $\forall k \in K\ \forall i, j \in I \cup D \cup 0$
t^k_0	The time of the AFV k arriving the airport;$\forall k \in K$
w_k	The weight of the AFV k;$\forall k \in K$
c_k	Fixed use cost of the AFV k;$\forall k \in K$
c_e	Penalty fee for early arrival
c_l	Penalty fee for late arrival
c_f	Unit cost of the fuel
c_p	Unit CO_2 emission cost
f_p	CO_2 emission factor
H	A very large fixed value
Decision variables	
x^k_{ij}	Whether the vehicular node i precedes vehicular node j on the AFV k or not; $\forall k \in K\ \forall i, j \in I \cup D \cup 0$
y^k_i	Whether the vehicular node i is visited by the AFV k or not; $\forall k \in K\ \forall i \in I \cup D \cup 0$
t^k_i	The time of the AFV k arriving at demand point i; $\forall k \in K\ \forall i \in I$
q^k_i	Number of passengers at demand point i assigned to the AFV k; $\forall k \in K\ \forall i \in I$
U_{ik}	An auxiliary (real) variable for subtour elimination constraint in the AFV k; $\forall k \in K$

3.2.2. Formulation. The proposed problem can be formulated as the following mixed integer program (MIP), which requires minimization:

$$\min f_1 = \sum_{\forall i \in I} \sum_{\forall k \in K} y^k_i \cdot p_i \cdot \left[c_e \cdot \max\left(a_i - t^k_i, 0\right) + c_l \cdot \max\left(t^k_i - l_i, 0\right)\right], \tag{1}$$

$$\min f_2 = \sum_{\forall i \in I} \sum_{\forall k \in K} x^k_{i0} \cdot c_k + c_f \cdot \sum_{\forall i, j \in I \cup D \cup 0} \sum_{\forall k \in K} x^k_{ij} \cdot F \cdot \left(w_k + q^k_i, v_{ij}\left(t^k_i\right), d_{ij}\right), \tag{2}$$

$$\min f_3 = c_p \cdot f_p \cdot \sum_{\forall i, j \in I \cup D \cup 0} \sum_{\forall k \in K} x^k_{ij} \cdot F \cdot \left(w_k + q^k_i, v_{ij}\left(t^k_i\right), d_{ij}\right), \tag{3}$$

which is subject to

$$\sum_{\forall i \in I} y^k_i \geq 1, \quad \forall k \in K, \tag{4}$$

$$\sum_{\forall i \in D} y^k_i = 1, \quad \forall k \in K, \tag{5}$$

$$\sum_{\forall k \in K} y^k_i = 1, \quad \forall i \in I, \tag{6}$$

$$\sum_{\forall j \in I} x^k_{ji} = 0, \quad \forall k \in K,\ \forall i \in D, \tag{7}$$

$$\sum_{\forall j \in I} x^k_{ij} \leq y^k_i, \quad \forall k \in K,\ \forall i \in D, \tag{8}$$

$$\sum_{\forall i \in I} x^k_{0i} = 0, \quad \forall k \in K, \tag{9}$$

$$\sum_{\forall i \in I} x_{i0}^k = 1, \quad \forall k \in K, \tag{10}$$

$$\sum_{\forall j \in I \cup 0} x_{ij}^k = \sum_{\forall j \in I \cup D} x_{ji}^k = y_i^k, \quad \forall k \in K,\ \forall i \in I, \tag{11}$$

$$U_{ik} - U_{jk} + |I \cup D \cup M| \cdot x_{ij}^k \geq |I \cup D \cup M| - 1, \quad \forall k \in K,\ \forall i \in I \cup D \cup 0, \tag{12}$$

$$\frac{t_i^k + T_i + d_{ij}}{v_{ij}\left(t_i^k\right) - \left(1 - x_{ij}^k\right)} \cdot H \leq t_j^k, \quad \forall k \in K,\ \forall i \in I \cup D \cup 0, \tag{13}$$

$$\frac{t_i^k + T_i + d_{ij}}{v_{ij}\left(t_i^k\right) + \left(1 - x_{ij}^k\right)} \cdot H \geq t_j^k, \quad \forall k \in K,\ \forall i \in I \cup D \cup 0, \tag{14}$$

$$t_0^k - t_i^k \leq Lt_i, \quad \forall k \in K,\ \forall i \in I, \tag{15}$$

$$q_i^k + p_i - \left(1 - x_{ij}^k\right) \cdot H \leq q_j^k, \quad \forall k \in K,\ \forall i \in I \cup D \cup 0, \tag{16}$$

$$q_i^k + p_i + \left(1 - x_{ij}^k\right) \cdot H \geq q_j^k, \quad \forall k \in K,\ \forall i \in I \cup D \cup 0, \tag{17}$$

$$q_i^k \leq Q_k, \quad \forall k \in K,\ \forall i \in I, \tag{18}$$

$$\sum_{\forall i,j \in I \cup D \cup 0} x_{ij}^k d_{ij} \leq D_{\max}^k, \quad \forall k \in K, \tag{19}$$

$$\sum_{\forall i,j \in I \cup D \cup 0} x_{ij}^k t_{ij} \geq T_{\min}, \quad \forall k \in K, \tag{20}$$

where $F(w_k + q_i^k, v_{ij}(t_i^k), d_{ij})$ is a function used to calculate the fuel consumption of vehicles passing through a given arc (i, j) when considering time-varying speeds. The speed $v_{ij}(t_i^k)$ is a piecewise function.

$v_{ij}(t_i^k)$ and its value depend on t_i^k. $w_k + q_i^k$ denotes the total weight, including the vehicle weight w_k and the load weight for the vehicle leaving the demand point i. Using the formulation proposed by Demir et al. [6, 34, 35, 38], this can be expressed as follows:

$$F\left(w_k + q_i^k, v_{ij}\left(t_i^k\right), d_{ij}\right) = 0.0308 d_{ij} \left\{ \frac{33}{v_{ij}\left(t_i^k\right)} + 0.8175 + 0.2725\left(w_k + q_i^k \cdot \gamma\right) + 0.0035 v_{ij}\left(t_i^k\right)^2 \right\}. \tag{21}$$

In this formulation, the objective function (1) aims at minimizing the penalty cost of time-window violation. The objective functions (2) and (3) simultaneously aim at minimizing the operating cost, including the fixed cost and cost of fuel consumption and cost of CO_2 emissions. Constraint (4) guarantees that each AFV visits at least one demand point. Constraint (5) is used to assign one depot to each AFV. Constraint (6) guarantees that all demand points are visited by AFVs. Constraints (7) and (8) guarantee that each AFV starts at the selected bus depot. Constraints (9) and (10) guarantee that each AFV eventually ends at the airport. Constraint (11) sets all demand points (except the airport and bus depot) being served by each AFV to have the same incoming and outgoing arcs. Constraint (12) is used for the subtour elimination in the vehicle routing. Constraints (13) and (14) are used for calculating the time for each AFV passing through adjacent vehicular points i and j. Constraint (15) guarantees that the riding time of each demand point must be less than the expected value. Constraints (16) and (17) are used for calculating the load weight for each AFV passing through adjacent vehicular points i and j. Constraint (18) guarantees that the number of passengers in each AFV must be less than the vehicle capacity. Constraint (19) guarantees that the total mileage of each AFV should not exceed its maximum range. Constraint (20) guarantees that the travel time of each AFV must meet its lower limits.

4. Heuristic-Based NSGA-II to Resolve Green DRASS

In this study, three possibly conflicting objectives require optimization. To avoid the disadvantages of the weighted sum approach to set weights to characterize the decision-makers preferences, NSGA-II is used to find a set of Pareto optimal solutions. As mentioned above, our model involves three core variables, including y_i^k, x_{ij}^k, and t_i^k , where y_i^k decides x_{ij}^k. Obviously, once several demand points and one depot can be assigned to an AFV (i.e., y_i^k), its route (i.e., x_{ij}^k) is also easily obtained according to the shortest distance principle. Hence, a two-stage NSGA-II-based heuristic approach is developed to solve this problem [39–41]. In the first stage, NSGA-II is used to determine the assignment of demand points and depots to different AFVs as well as departure time of each AFV. In the second stage, the A* algorithm is embedded in NSGA-II to construct the route of each AFV: leaving the depot, visiting the selected demand points, and returning to the airport. Through these two operations, the speed, weight, and time for an AFV arriving and leaving the location of demand points in the objective function are also determined.

4.1. NSGA-II in the First Stage. A three-dimensional vector $U = (U_1, U_2, U_3)$ represents the chromosome of a feasible solution in green DRASS, where each element u_i in $U_1 = (u_1, u_2, \ldots, u_I)$, being a natural number in 1, 2, ... ,K, denotes an AFV u_i visiting demand point $i\,(i = 1, 2, \ldots, I)$. Each element u_i in $U_2 = (u_{I+1}, u_2, \ldots, u_{I+K})$, being a real number, denotes the departure time of the AFV $k\,(k = 1, 2, \ldots, K)$. Each element u_i in $U_3 = (u_{I+K+1}, u_2, \ldots, u_{I+2K})$, being a natural number in 1, 2, ..., D, denotes the

departure depot of the AFV k ($k = 1, 2, \ldots, K$). For example, a chromosome vector $U = \{1\ 1\ 2\ 2\ 1\ 2\ 6{:}00\ 6{:}10\ 3\ 4\}$ of two AFVs, four depots, and six demand points can be understood as follows: vehicle 1 leaves depot 3 at 6:00 and visits demand points 1, 2, and 5; vehicle 2 leaves depot 4 at 6:10 and visits demand points 3, 4, and 6.

Figure 3 describes the optimization procedure of NSGA-II. An initial population consisting of individuals is randomly generated. At each iteration, each individual is firstly decoded as an assignment of demand points and depots to vehicles as well as the determination of departure time for the vehicle, and the A* algorithm is used to search for the shortest distance route of each vehicle. After this, all objective functions are calculated for fitness evaluation. The selected individuals in the parent population exchange genes to generate new individuals by using crossover and mutation operators. The current population with older and new individuals is sorted and selected again based on nondomination to obtain offspring. The selection depends on rank and crowding distance for an individual. The crowding distance, related to averaging the distances between individuals in a front, is used to find a consistent spread of solutions along the Pareto front. If the solutions are in the same nondominated front, those with a higher crowding distance are chosen first; otherwise, those with the lowest rank are chosen first. The algorithm continues until the maximum number of iterations is reached.

4.2. A Algorithm in the Second Stage.* In the first stage, all demand points, the depots, and the airport are assigned to different AFVs. Based on the output of the first stage algorithm, a route for each AFV would be found based on the shortest distance principle. The A* algorithm is one of the most efficient algorithms for finding the shortest path [42]. The detailed process for this algorithm in solving the problem is given as follows:

(i) Step 1: define three sets, close_list, open_list, and parent_node, where close_list and open_list are used to record visited and unvisited demand points and *parent_node* is used to record the route.

(ii) Step 2: put the depot into open_list and set lose_list = ∅ and parent_node = ∅.

(iii) Step 3: in open_list, select a demand point with the minimum travel distance as the current node. Remove the current node from open_list, and put it into close_list.

(iv) Step 4: if the current node is the airport, the search ends; otherwise, search the adjacent nodes of the current node. Put these adjacent nodes into open_list, and the current node is set as a node in parent_node.

(v) Step 5: if the current node is in open_list, its total travel distance is recalculated and updated when the result is less than the previous one; parent_node is also updated. If the current node is in close_list, go to step 3.

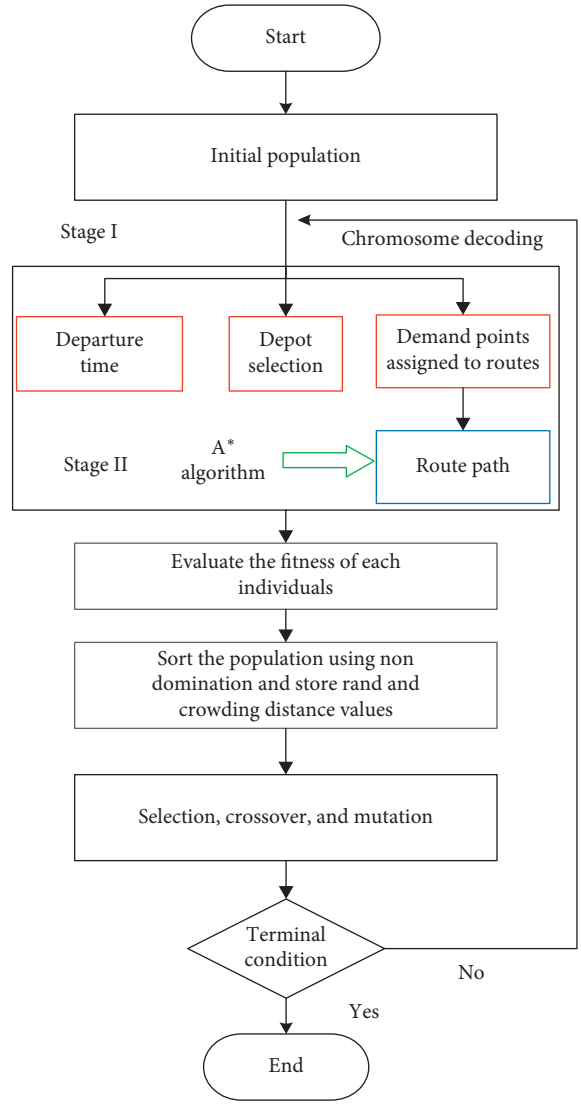


Figure 3: Flow chart of the NSGA-II-based two-stage algorithm.

5. Numerical Example

5.1. Example Description. In this section, a case study of designing green DRASS for Nantong City in China is given to demonstrate the applicability of the proposed model. There is one airport (M), five depot centers ($D1$–$D5$), and 33 demand points ($P1$–$P33$). Figure 4 is used to map the spatial distribution of these vehicular nodes, in which time-varying speeds on the road are obtained from a Baidu Open GIS tool using http://api.map.baidu.com. Moreover, the number of passengers and their preferred time windows and expected ride time in demand points are shown in Table 2. The parameters of NSGA-II are given in [34]. The proposed algorithm is run on Matlab 2017 B using a Dell Inspiron N5040 laptop with Core i3 processor and 4 GB memory. The key parameters used in the case study are as follows:

(i) Number of AFVs: 6.

(ii) Maximum capacity of each AFV: $Q_k = 12$ per.

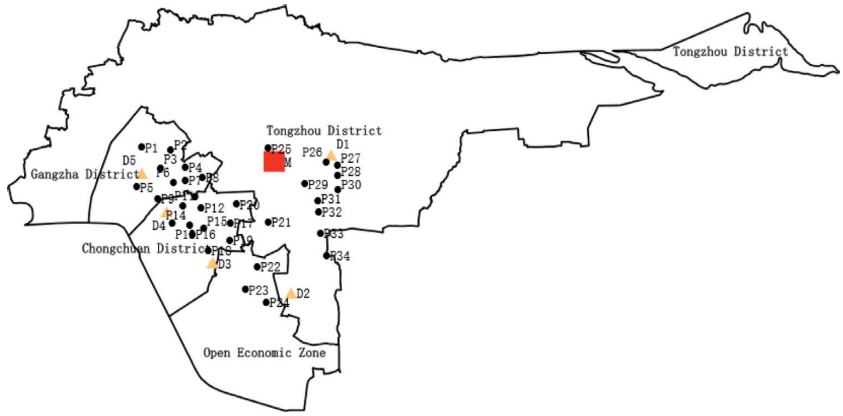


Figure 4: Spatial distribution of vehicles visiting points in Nantong City in China.

Table 2: Basic information of demand points.

No.	$[a_i, l_i]$	p_i	$Lt_i(h)$	No.	$[a_i, l_i]$	p_i	$Lt_i(h)$	No.	$[a_i, l_i]$	p_i	$Lt_i(h)$
$P1$	8:30–8:40	2	2.5	$P12$	7:30–8:00	1	4	$P23$	8:10–8:30	2	3.5
$P2$	8:20–8:30	1	2.5	$P13$	6:20–6:40	2	2.5	$P24$	8:00–8:30	2	3
$P3$	8:00–8:10	2	2.5	$P14$	6:20–6:50	2	2	$P25$	7:10–7:30	2	1
$P4$	6:40–7:10	2	4	$P15$	6:20–6:40	2	2	$P26$	7:40–7:50	2	0.5
$P5$	10:00–10:10	2	1.5	$P16$	6:40–7:00	2	2	$P27$	6:40–7:10	1	5
$P6$	8:00–8:20	2	1	$P17$	8:00–8:30	1	3.5	$P28$	10:00–10:10	2	1
$P7$	6:40–7:10	1	1.5	$P18$	6:30–7:00	2	2.5	$P29$	7:50–8:00	2	1.5
$P8$	6:40–7:10	2	1.5	$P19$	7:00–7:10	2	2.5	$P30$	7:50–8:10	2	1
$P9$	6:10–6:30	2	2.5	$P20$	6:50–7:10	2	4.5	$P31$	6:40–7:10	1	5
$P10$	7:40–8:00	2	1.5	$P21$	6:30–7:00	2	2	$P32$	8:00–8:20	1	1
$P11$	6:30–7:00	2	4.5	$P22$	8:00–8:20	1	3.5	$P33$	8:00–8:10	2	0.5

(iii) Biggest continue voyage course of each AFV: $D^k_{\max} = 100$ km.

(iv) Weight of each AFV: $w_k = 2300$ kg.

(v) Fixed use cost of each AFV: $c_k = 120$ CNY.

(vi) Minimum travel time of vehicle route: $T_{\min} = 3$ min.

(vii) Boarding time of per passenger: $\beta = 0.5$ min.

(viii) Average weight of each person: $\gamma = 60$ kg.

(ix) Penalty fee for early arrival: $c_e = 1$ CNY/min.

(x) Penalty fee for late arrival: $c_l = 3$ CNY/min.

(xi) Unit cost of the fuel: $c_f = 7.59$ CNY/liter.

(xii) Unit CO_2 emission cost: $c_p = 80$ CNY/ton.

(xiii) CO_2 emission factor: $f_p = 0.785$ kg/liter.

5.2. Results. The CPU time for finding a set of Pareto optimal solutions is usually less than 1 minute. As explained previously, the proposed model could yield 10 feasible metaoptimal solutions in three dimensions, including the assignment of each demand point to an AFV, its route, and timetable. The upper and lower bounds of objective function 1 are 1305.8 RMB and 693.7 RMB, that of objective function 2 are 1187.2 RMB and 1156.7 RMB, and that of objective function 3 are 3.9 RMB and 3.7 RMB, respectively. Figures 5 and 6 reveal the changing relationship between any two goals. As the value of objective function 1 becomes larger, that of objective function 2 becomes smaller. This is because the reduction in the total penalty charges for violating the passenger's time window leads to the need for more mileage for vehicles covering all demand points, thus increasing energy consumption costs. Objective 2 and objective 3 are positively linearly correlated, which determines that the change in trend of objective 1 and objective 2 is consistent with the change in trend of objective 1 and objective 3. This is because energy consumption determines carbon emissions.

Tables 3 and 4 are used to provide the routing and scheduling results related to the metaoptimal solution (693.7, 1187.2, 3.9). The pick-up time, riding time, and early/late arrival time of the AFV at each pick-up location are detailed in Table 3, in which zero, positive, and negative numbers in the fourth column of the table represent normal, early, and late arrival, respectively. The arrival and departure time for all AFVs visiting their demand points as well as fuel consumption and CO_2 emissions of each AFV are detailed in Table 4. Taking the AFV of $V1$ as an example, the vehicle leaves the depot $D5$ at 06 : 28 and arrives at the demand point $P29$ at 7 : 36. Since the client's time window is (7 : 50, 8 : 00), the early arrival time is 13.9 minutes. After 1 minute to pick up 2 persons, it leaves this location at 7 : 37. Similarly, the arrival and departure time of $P30$ and $P33$ are (7 : 53, 7 : 54)

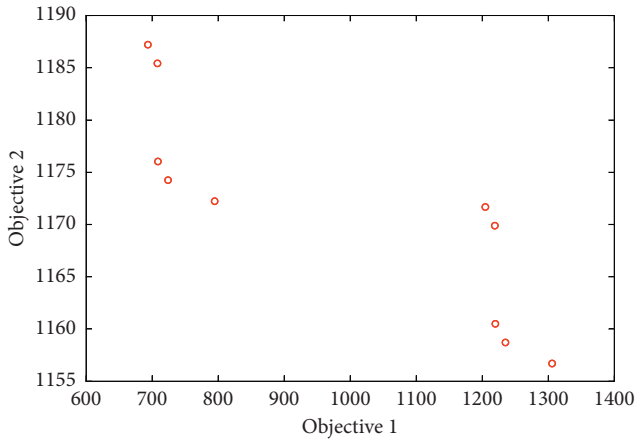


Figure 5: Negative nonlinear relationship between objective 1 and objective 2.

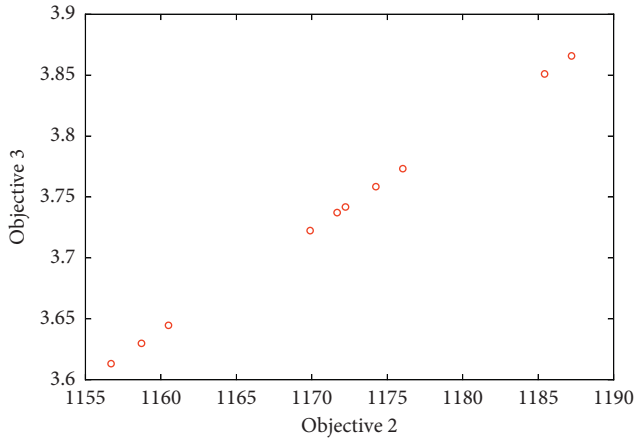


Figure 6: Positive linear relationship between objective 2 and objective 3.

and (8 : 07, 8 : 08), and their early or late arrival times are 0 minutes and 0 minutes. Finally, this vehicle reaches its airport of M at 8 : 37 to unload 6 persons. As above mentioned, the riding times of $P29$, $P30$, and $P33$ for $V4$ are calculated as 61.1 minutes, 44.2 minutes, and 29.6 minutes. The vehicle travels a total of 52.6 km, consuming 9 liters of petrol and emitting 7.1 kg of CO_2. The costs in terms of energy consumption and carbon emissions are 68.3 CNY and 0.5 CNY, and the penalty for violating the time window is 27.8 CNY.

Furthermore, the proposed model has unique features compared to traditional DRASS with fixed speed. As shown in Figure 7, mileage of the proposed model is 0.36% more than for the traditional one, but the carbon emissions and fuel consumption of the proposed model are 0.87% and 0.66% less than those of the traditional one. In traditional DRASS with fixed speed, the more the mileage, the more the carbon emissions and fuel consumption. However, the flexible departure time of each AFV in time-varying traffic conditions can avoid vehicles driving on congested roads, thus increasing energy consumption costs.

5.3. Comparative Analysis and Parameter Sensitivity. In this section, sensitivity analyses are also given to investigate the impact of the number of AFVs and the changes in weight coefficient c_l/c_e on the model performance. Comparison of the results is shown in Tables 5 and 6, from which the following holds:

(1) As the weight coefficient gradually increases, early arrival time will be increased while late arrival time will be reduced. Besides, total mileage, fuel consumption, and CO_2 emissions remain the same, except for the boundary conditions being met. When it happens, DRASS plan with an increase in mileage would lead to more fuel consumption and CO_2 emissions. This is because increased weight coefficient c_l/c_e may cause an increase in early arrival time and a reduction in late arrival time, which leads to more mileage.

(2) As the number of AFVs increases, total mileages slightly increase. This can be attributed to an increase in invalid mileage from vehicles starting from the depot and ending at the airport. Similarly, a reduced number of demand points visited by those AFVs may causes vehicles to reach the pick-up locations faster, thus increasing total early and late arrival time. When mileage of two DRASS plans varies widely, DRASS plan with greater mileage means more energy consumption and emissions. However, DRASS plan with small mileage in congested road conditions may consume more energy and emit more emissions, as the mileage difference in them is small.

5.4. Validation of Algorithm. In order to verify the validity of proposed algorithm, the multiobjective DRASS could be transformed into a single objective function with minimization of the weighted sum of two objective functions (1), (2), and (3). Several instances with different problem scales are randomly generated. Globally optimal solutions are gained by using CPLEX 12.6 to solve small size instances. We also implement the ACO-based heuristic algorithm to solve large-scale instances. Table 7 shows their difference in terms of computation efficiency and solution quality. As more demand points are covered by all routes, more solved time is required to obtain the solutions, where the computation time of CPLEX increases geometrically, while that of heuristic algorithm increases gently. We can observe that CPLEX takes more than 1 h to solve optimal solutions in case of more than 30 demand points, while the heuristic spends 5.1 seconds to obtain locally optimal solutions. The difference of best solutions is controlled under 10%, but that of solved time will grow at a geometric rate as well.

Table 3: Assignment results of passengers picked up by each AFV.

No.	Pick-up time (min)	Ride time (min)	Early/late time (min)	Vehicle
*P*8	1	78.8	0	
*P*7	0.5	68.3	−4.2	*V*1
*P*25	1	38.2	−14.3	
*P*26	1	20.4	−12.1	
*P*18	1	138.4	0	
*P*19	1	131.8	15.0	
*P*21	1	110.1	−6.7	*V*2
*P*10	1	62.3	0	
*P*6	1	48.1	0	
*P*27	0.5	287.5	0	
*P*31	1	265.7	−14.7	
*P*20	1	243.2	−17.2	
*P*12	0.5	232.4	0	*V*3
*P*22	1	204.5	0	
*P*23	1	195.6	0	
*P*24	1	187.6	0	
*P*5	1	83.8	0	
*P*29	1	61.1	13.9	
*P*30	1	44.2	0	*V*4
*P*33	1	29.6	0	
*P*11	1	253.3	0	
*P*4	1	234.5	−7.7	
*P*17	0.5	207.5	35.3	
*P*3	1	148.3	−14	*V*5
*P*2	0.5	141.3	−1	
*P*1	1	134.3	0	
*P*28	1	43.6	0	
*P*9	1	135.1	0	
*P*13	1	122.1	0	
*P*14	1	116.9	0	*V*6
*P*16	1	112	0	
*P*15	1	100.2	−17.5	
*P*32	1	41.9	4.3	
—	—	—	—	—

Table 4: Routing and scheduling plan of each AFV.

Vehicle	Sequence of demand points visited by each AFV	Mileage (km)	Fuel consumption (liter)	CO_2 emissions (kg)	Number of passengers
*V*1	*D*4 (6 : 45)–*P*8 (7 : 03–7:04)–*P*7 (7 : 14, 7 : 14)–*P*25 (7 : 44, 7 : 45)–*P*26 (8 : 02, 8 : 03)–*M* (8 : 22)	48.4	9.1	7.1	7
*V*2	*D*3 (6 : 35)–*P*18 (6 : 38, 6 : 39)–*P*19 (6 : 44, 6 : 45)–*P*21 (7 : 06, 7 : 07)–*P*10 (7 : 54, 7 : 55)–*P*6 (8 : 08, 8 : 09)– *M* (9 : 56)	44.4	6.8	5.3	10
*V*3	*D*1 (6 : 36)–*P*27 (6 : 42, 6 : 43)–*P*31 (7 : 04, 7 : 05)–*P*20 (7 : 27, 7 : 28)–*P*12 (7 : 37, 7 : 38)–*P*22 (8 : 05, 8 : 06) –*P*23 (8 : 14, 8 : 15)– *P*24 (8 : 22, 8 : 23)– *P*5 (10 : 06, 10 : 07)–*M* (11 : 30)	94.5	15.0	11.8	12
*V*4	*D*5 (6 : 28)–*P*29 (7 : 36, 7 : 37)–*P*30 (7 : 53, 7 : 54)–*P*33 (8 : 07, 8 : 08)–*M* (8 : 37)	52.6	9.0	7.1	6
*V*5	*D*4 (6 : 21)–*P*11 (6 : 39, 6 : 40)–*P*4 (6 : 57, 6 : 58)–*P*17 (7 : 24, 7 : 25)–*P*3 (8 : 24, 8 : 25)–*P*2 (8 : 31, 8 : 31)– *P*1 (8 : 37, 8 : 38)–*P*28 (10 : 08, 10 : 09)–*M* (10 : 52)	90.5	14.2	11.1	12
*V*6	*D*4 (6 : 17)–*P*9 (8 : 22, 8 : 23)–*P*13 (6 : 35, 6 : 36)–*P*14 (6 : 40, 6 : 41)–*P*16 (6 : 45, 6 : 46)–*P*15 (6 : 57, 6 : 58)–*P*32 (7 : 55, 7 : 56)–*M* (8 : 37)	47.4	7.2	5.8	11

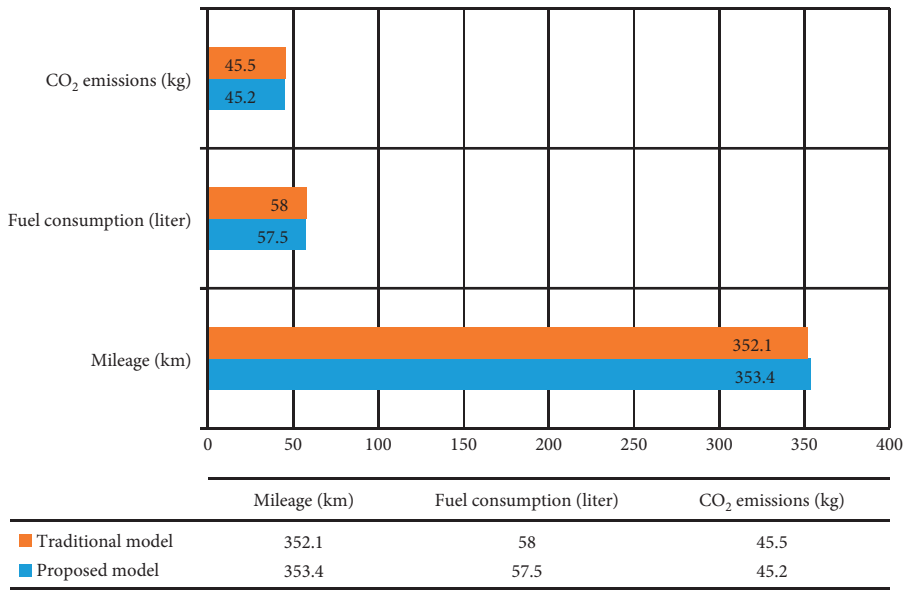


Figure 7: Comparison of the results between proposed and traditional model.

Table 5: Comparison of the results of different weight coefficients c_l/c_e

Scenario	Mileage (km)	Fuel consumption (liter)	CO_2 emissions (kg)	Early arrival time (min)	Late arrival time (min)
1 : 3	377.8	61.5	48.3	68.4	119.3
2 : 3	374	60.0	47.1	77.1	111.3
3 : 3	374	60.0	47.1	77.1	109.3
4 : 3	366.1	59.2	46.5	86.0	78.0

Table 6: Comparison of the results of different numbers of AFVs.

Scenario	Mileage (km)	Fuel consumption (liter)	CO_2 emissions (kg)	Early and late arrival time (min)
6 AFVs	353.4	57.5	45.2	363.4
7 AFVs	444.7	74.3	58.3	891.1
8 AFVs	451	73.8	57.9	1000.6

Table 7: Comparison of CPLEX solution and heuristic solution.

	CPLEX results		Heuristic results			Difference	
Scenario $\|I\|$	Solved time	Best (CNY)	Solved time	Average (CNY)	Best (CNY)	Solved time	Best (CNY)
10	389 s	2496.2	2.3 s	2693.1	2496.2	0.59%	0%
30	3912 s	6794.9	5.1 s	7651.6	6885.5	0.13%	1.3%
60	>24 h	10214.6	9.6 s	12268.4	11003.9	0.01%	7.7%
120	—	—	21.8 s	29663.2	28088.5	—	—
180	—	—	35.6 s	52069.6	50149.7	—	—

6. Conclusions

In this paper, a novel MOMILP model was proposed for green DRASS with time-varying speeds to balance the operating cost, vehicle fuel consumption, and CO_2 emissions. Furthermore, customer's soft time windows and limited fuel capacity for an AFV are incorporated into the proposed model in order to represent real-world conditions. The metaoptimal solutions are obtained using NSGA-II. The main findings can be summarized as follows:

(1) Carbon emissions cost and fuel consumption cost of the proposed model are positively linearly correlated, and both these and the total penalty charges for violating the passenger's time window are negatively nonlinearly correlated.

(2) When mileage of two DRASS plans varies widely, DRASS plan with a greater mileage means more energy consumption and emissions. However, DRASS plan with small mileage in congested road conditions may consume more energy and emit

more emissions than DRASS plan with big mileage in good road conditions, as the mileage difference in them is small.

(3) The carbon emissions and fuel consumption of the proposed model are less than those of DRASS with fixed speeds. This is because the flexible departure time of each AFV in time-varying traffic conditions can avoid vehicles driving on congested roads, thus increasing energy consumption costs.

Note that the vehicle in this DRASS model is an AFV, not an electric bus (EC). An EC emits almost no carbon, but it can be recharged at any of the available stations during the workday. In this sense, the collaborative design of DRASS routes and recharging station locations should consider the interactions between construction and operating expenses of charging piles and operation costs of these ECs. As a result, extending the DRASS model to simultaneously select the optimal locations of charging piles, design ECs routes, and assign ECs to recharging stations is worth further studying.

Data Availability

Some or all data, models, or code generated or used during the study are available from the corresponding author on request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was financially supported by the Humanities and Social Sciences Foundation of the Ministry of Education of China (20YJCZH176) and the central college basic scientific research operating expenses fund in Civil Aviation University of China (3122020079).

References

[1] T. Liu and A. Ceder, "Analysis of a new public-transport-service concept: customized bus in China," *Transport Policy*, vol. 39, pp. 63–76, 2015.

[2] J. Wang, T. Yamamoto, and K. Liu, "Key determinants and heterogeneous frailties in passenger loyalty toward customized buses: an empirical investigation of the subscription termination hazard of users," *Transportation Research Part C: Emerging Technologies*, vol. 115, Article ID 102636, 2020.

[3] J. Wang, T. Yamamoto, and K. Liu, "Role of customized bus services in the transportation system: insight from actual performance," *Journal of Advanced Transportation*, vol. 2019, Article ID 6171532, 14 pages, 2019.

[4] P.-Y. Yang, J.-F. Tang, Y. Yu, and J.-X. Pei, "Minimizing carbon emissions through vehicle routing and scheduling in the shuttle service of picking up and delivering customers to the airport," *Acta Automatica Sinica*, vol. 39, no. 4, pp. 424–432, 2013.

[5] Y. Xiao and A. Konak, "The heterogeneous green vehicle routing and scheduling problem with time-varying traffic congestion," *Transportation Research Part E: Logistics and Transportation Review*, vol. 88, pp. 146–166, 2016.

[6] E. Demir, T. Bektaş, and G. Laporte, "A review of recent research on green road freight transportation," *European Journal of Operational Research*, vol. 237, no. 3, pp. 775–793, 2014.

[7] B. Sun, M. Wei, and W. Wu, "An optimization model for demand-responsive feeder transit services based on ride-sharing car," *Information*, vol. 10, no. 12, pp. 370–386, 2019.

[8] L. Q. Wang, Z. Jun, and W. Wei, "Hyperpath-based vehicle routing and scheduling method in time-varying networks for airport shuttle service," *Natural Computing*, vol. 18, no. 4, pp. 769–784, 2017.

[9] B. Sun, M. Wei, and S. Zhu, "Optimal design of demand-responsive feeder transit services with passengers' multiple time windows and satisfaction," *Future Internet*, vol. 10, no. 3, pp. 30–46, 2018.

[10] X. Li, M. Wei, J. Hu, Y. Yuan, and H. Jiang, "An agent-based model for dispatching real-time demand-responsive feeder bus," *Mathematical Problems in Engineering*, vol. 2018, no. 1, 11 pages, Article ID 6925764, 2018.

[11] M. Wei, B. Sun, and S. L. Zhu, "A hybrid-type indicator set pairs analysis model for evaluating transit operational efficiency," *Journal of Nonlinear and Convex Analysis*, vol. 30, no. 5, pp. 895–904, 2019.

[12] Y. Xia and Z. Fu, "Improved tabu search algorithm for the open vehicle routing problem with soft time windows and satisfaction rate," *Cluster Computing*, vol. 22, no. S4, pp. 8725–8733, 2019.

[13] A. K. Beheshti and S. R. Hejazi, "A novel hybrid column generation-metaheuristic approach for the vehicle routing problem with general soft time window," *Information Sciences*, vol. 316, no. 6, pp. 598–615, 2015.

[14] J. Xu, F. Yan, and S. Li, "Vehicle routing optimization with soft time windows in a fuzzy random environment," *Transportation Research Part E Logistics & Transportation Review*, vol. 47, no. 6, pp. 1071–1091, 2011.

[15] F. Daniela, M. Elena, and P. Paola, "Ant colony system for a VRP with multiple time windows and multiple visits," *Journal of Interdisciplinary Mathematics*, vol. 10, no. 2, pp. 263–284, 2007.

[16] S. Belhaiza, P. Hansen, and G. Laporte, "A hybrid variable neighborhood tabu search heuristic for the vehicle routing problem with multiple time windows," *Computers & Operations Research*, vol. 52, pp. 269–281, 2014.

[17] W. Jun and L. I. Bo, "Multi-objective tabu search algorithm for vehicle routing problem with fuzzy due-time," *Computer Integrated Manufacturing Systems*, vol. 17, no. 4, pp. 858–866, 2011.

[18] M. Adelzadeh, V. Mahdavi Asl, and M. Koosha, "A mathematical model and a solving procedure for multi-depot vehicle routing problem with fuzzy time window and heterogeneous vehicle," *The International Journal of Advanced Manufacturing Technology*, vol. 75, no. 5–8, pp. 793–802, 2014.

[19] J. Brito, F. J. Martínez, J. A. Moreno, and J. L. Verdegay, "An ACO hybrid metaheuristic for close-open vehicle routing problems with time windows and fuzzy constraints," *Applied Soft Computing*, vol. 32, pp. 154–163, 2015.

[20] N. Norouz, M. Sadegh-Amalnick, and R. Tavakkoli-Moghaddam, "Modified particle swarm optimization in a time-dependent vehicle routing problem: minimizing fuel consumption," *Optimization Letters*, vol. 11, no. 1, pp. 1–14, 2016.

[21] D. Z. Zhang, X. Wang, S. Y. Li, N. Li, and Z. Zhang, "An Joint optimization of green vehicle scheduling and routing problem

with time-varying speeds," *PLoS One*, vol. 13, no. 2, pp. 1–20, 2018.

[22] Y. Deng, W. Zhu, H. Li, and Y. Zheng, "Multi-type ant system algorithm for the time dependent vehicle routing problem with time windows," *Journal of Systems Engineering and Electronics*, vol. 29, no. 3, pp. 625–638, 2018.

[23] F. Hooshmand and S. A. MirHassani, "Time dependent green VRP with alternative fuel powered vehicles," *Energy Systems*, vol. 10, no. 3, pp. 721–756, 2019.

[24] M. A. Figliozzi, "The impacts of congestion on time-definitive urban freight distribution networks CO_2 emission levels: results from a case study in Portland, Oregon," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 5, pp. 766–778, 2011.

[25] O. Jabali, T. Van Woensel, and A. G. De Kok, "Analysis of travel times and CO_2 emissions in time-dependent vehicle routing," *Production and Operations Management*, vol. 21, no. 6, pp. 1060–1074, 2012.

[26] K. Harwood, C. Mumford, and R. Eglese, "Investigating the use of metaheuristics for solving single vehicle routing problems with time-varying traversal costs," *Journal of the Operational Research Society*, vol. 64, no. 1, pp. 34–47, 2013.

[27] İ. Kara, B. Y. Kara, and M. K. Yetis, *Energy Minimizing Vehicle Routing Problem*, pp. 62–71, Springer Berlin Heidelberg, Heidelberg, Germany, 2007.

[28] Y.-J. Kwon, Y.-J. Choi, and D.-H. Lee, "Heterogeneous fixed fleet vehicle routing considering carbon emission," *Transportation Research Part D: Transport and Environment*, vol. 23, pp. 81–89, 2013.

[29] T. Bektaş and G. Laporte, "The pollution-routing problem," *Transportation Research Part B: Methodological*, vol. 45, no. 8, pp. 1232–1250, 2011.

[30] A. Lee and M. Savelsbergh, "An extended demand responsive connector," *EURO Journal on Transportation and Logistics*, vol. 6, no. 1, pp. 25–50, 2017.

[31] J. X. Shen, S. Q. Yang, X. M. Gao, and F. Qiu, "Vehicle routing and scheduling of demand-responsive connector with on-demand stations," *Advances in Mechanical Engineering*, vol. 9, no. 6, pp. 1–10, 2017.

[32] X. Lu, J. Yu, X. Yang, S. Pan, and N. Zou, "Flexible feeder transit route design to enhance service accessibility in urban area," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 507–521, 2016.

[33] P. Carotenuto, C. Cis, S. Rismondo, and G. Storchi, "Hybrid genetic algorithm to approach the DaRP in a demand responsive passenger service," *IFAC Proceedings Volumes*, vol. 39, no. 3, pp. 315–320, 2006.

[34] K. Liu, T. Yamamoto, and T. Morikawa, "Impact of road gradient on energy consumption of electric vehicles," *Transportation Research Part D: Transport and Environment*, vol. 54, pp. 74–81, 2017.

[35] T. Q. Tang, J. Zhang, and K. Liu, "A speed guidance model accounting for the driver's bounded rationality at a signalized intersection," *Physica A: Statistical Mechanics and its Applications*, vol. 473, pp. 45–52, 2017.

[36] A. Ceder, "Stepwise multi-criteria and multi-strategy design of public transit shuttles," *Journal of Multi-Criteria Decision Analysis*, vol. 16, no. 1-2, pp. 21–38, 2009.

[37] B. Sun and M. Wei, "Optimal integrated location and dispatching decisions for feeder bus route design problem," *Transport*, preprinting, 2020.

[38] X. Luo, L. Dong, Y. Dou et al., "Factor decomposition analysis and causal mechanism investigation on urban transport CO_2 emissions: comparative study on Shanghai and Tokyo," *Energy Policy*, vol. 107, pp. 658–668, 2017.

[39] J. Jemai, M. Zekri, and K. Mellouli, "An NSGA-II algorithm for the green vehicle routing problem," *Evolutionary Computation in Combinatorial Optimization*, pp. 37–48, 2012.

[40] E. Demir, T. Bektaş, and G. Laporte, "The bi-objective pollution-routing problem," *European Journal of Operational Research*, vol. 232, no. 3, pp. 464–478, 2014.

[41] A. Baniamerian, M. Bashiri, and F. Zabihi, "Two phase genetic algorithm for vehicle routing and scheduling problem with cross-docking and time windows considering customer satisfaction," *Journal of Industrial Engineering International*, vol. 14, no. 1, pp. 15–30, 2018.

[42] S. Baldi, N. Maric, R. Dornberger, and T. Hanne, "Pathfinding optimization when solving the paparazzi problem comparing A* and Dijkstra's algorithm," in *Proceedings of the 2018 6th International Symposium on Computational and Business Intelligence (ISCBI)*, pp. 12–16, Basel, Switzerland, August 2018.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 3753601, 24 pages
https://doi.org/10.1155/2020/3753601

Research Article

Application of Modified NSGA-II to the Transit Network Design Problem

Jie Yang [1,2] **and Yangsheng Jiang** [1,2]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu, Sichuan 611756, China*

Correspondence should be addressed to Yangsheng Jiang; jiangyangsheng@swjtu.cn

Received 5 January 2020; Revised 22 June 2020; Accepted 13 July 2020; Published 1 August 2020

Academic Editor: Yu Jiang

The transit network design problem involves determining a certain number of routes to operate in an urban area to balance the costs of the passengers and the operator. In this paper, we simultaneously determine the route structure of each route and the number of routes in the final solution. A novel initial route set generation algorithm and a route set size alternating heuristic are embedded into a nondominated sorting genetic algorithm-II- (NSGA-II-) based solution framework to produce the approximate Pareto front. The initial route set generation algorithm aims to generate high-quality initial solutions for succeeding optimization procedures. To explore the solution space and to have solutions with a different number of routes, a route set size alternating heuristic is developed to change the number of routes in a solution by adding or deleting one route. Experiments were performed on Mandl's network and four larger Mumford's networks. Compared with a fixed route set size approach, the proposed NSGA-II-based solution method can produce an approximate Pareto front with much higher solution quality as well as improved computation efficiency.

1. Introduction

Followed by frequency setting, timetable development, bus scheduling, and driver scheduling, the transit network design problem (TNDP) [1] aims to determine a set of routes to operate in an urban area to balance the costs of the passengers and the operator. The performance of the succeeding four stages highly depends on the quality of the results of the transit network design. Thus, the TNDP has been continuously studied during the last five decades. Known as a NP-hard problem, the TNDP is a difficult combinatorial optimization problem, and its optimal solution is difficult to find even by using supercomputers.

Despite the attention given TNDP, the determination of a suitable number of routes (or a range) for a specific network has received little attention. The route set size is an important parameter in the transit network design because of the following reasons. First, the route set size implicitly reflects the total route length, which directly affects the fuel costs, the mileage of the public transport, and other operation costs. Second, there is a trade-off between the route set size and the average travel time, which is an important indicator of the transit network performance. A larger route set size usually indicates more direct services for the passengers. Optimizing the route set size can lead to solutions with a wide range of trade-off levels between the operator cost and the passenger cost.

Pattnaik et al. [2] proposed a variable string length coding method in which the route set size is determined along with the route set. Similarly, Tom and Mohan [3] incorporated the variation for the route set size in the coding to address the transit network routing and frequency setting problem. However, the mentioned research studies imply two limitations: (i) the optimal route set is only the subset of a large set of candidate routes generated by a candidate route generation algorithm. This method is not practical for large

networks as the candidate route set scales exponentially with the network size. (ii) The resultant model can only give one answer per run. To address these issues, in this work, we consider the number of routes as an endogenous variable and vary it during the search for the approximate Pareto front. In other words, a route set size alternating heuristic is used together with the NSGA-II for solving simultaneously the TNDP and the route set size determination problem.

With respect to objective functions, most previous studies [2–9] have adopted the weighted sum approach to achieve different objectives (i.e., the weighted sum of the operator cost and the passenger cost). This method can achieve a relatively good balance between the passenger cost and the operator cost. However, the determination of a suitable set of weights usually requires many experiments, which is time consuming, and the suitability of the weights obtained may vary from network to network. Besides, some researchers [10–18] approached the TNDP by considering the operator cost or the passenger cost as the only objective to be optimized. As a result, given the conflicting nature of the objectives, the optimal solution obtained might not be practical from the viewpoint of the other stakeholder. Therefore, in this work, we try to produce an approximate Pareto front from which the desired solution can be selected by the decision maker.

The main contributions of this research are threefold: (a) the development of an initial route set generation algorithm to produce high-quality initial solutions for the TNDP; (b) the proposal of a route set size alternating heuristic to alternate the number of routes in a solution during the optimization procedure; and (c) the illustration of the applicability of the proposed solution method in five networks with different scales.

The remainder of this paper is organized as follows. In Section 2, we review existing works on route set initialization and optimization methods. The transit network design problem is described in Section 3. Section 4 elaborates the proposed initial route set generation algorithm and route set size alternating heuristic. Section 5 presents numerical results, and we conclude in Section 6.

2. Literature Review

The typical solution method for TNDP comprises two parts: an initialization procedure and an optimization algorithm. In this section, we systematically review related works from these two perspectives.

2.1. Initialization Procedure. Most initialization procedures can be classified into two categories: initial route set generation and candidate route set generation. One route set consists of a predefined number of stop sequences which can be used to define a transit network. As our work is focusing on the transit routing problem, interesting readers for other transit network planning problems can refer to Guihaire and Hao [19], Kepaptsoglou and Karlaftis [20], Farahani et al. [21], Abedin et al. [22], and Liu and Ceder [23].

Lampkin and Saalmans [24] first designed a heuristic algorithm to generate route set. They produced a skeleton route of four nodes and inserted acceptable nodes one by one into the skeleton route according to a predefined criterion until a complete route is obtained. After each insertion, demand satisfied by the formed route will be eliminated in order to guide the process. This framework of constructing route was adopted to develop initial route generation algorithm by Baaj and Mahmassani [5]. Mandl [25] approached urban transit routing problem with two steps: first all the shortest paths connecting every origin-destination pair were found to generate feasible initial route set, and then used heuristics to improve the quality of initial route set. With the consideration of transit demand, Baaj and Mahmassani [5] designed a route generation algorithm (RGA) to generate initial route set. Specifically, high demand node pairs are selected to define the initial skeletons with the help of a designer. Once the skeletons were specified, they were expanded into routes with a node selection or insertion strategy. The algorithm can also generate different sets of routes with different trade-offs between passengers and operator costs. Inspired by RGA [5], Mauttone and Urquhart [26] proposed another route generation algorithm called pair insertion algorithm. They used a novel strategy of inserting pairs of nodes, instead of the original expansion of routes by inserting single vertices. Experimental results reported by them evidenced that pair insertion algorithm can produce better solutions from the operator viewpoint with similar quality from the passengers' viewpoint.

To avoid defining a skeleton first, Chakroborty and Wivedi [12] determined each route by first selecting the starting node and then adding adjacent nodes to the former node until some termination criteria are satisfied. The probability of selecting a node i is determined by its activity level which is calculated based on the number of trips originating from node i and starting from node i. Similarly, Mumford [27] constructed the first route by adding adjacent nodes to a randomly selected start node and initiated other routes in the same way with a start node used in previously generated routes. Kechagiopoulos and Beligiannis [13] and Owais and Osman [28] proposed a similar procedure for initial route set generation. Instead of selecting nodes randomly, Nikolic and Teodorovic [16] developed a deterministic procedure for initial route set generation. They generated the initial solution by always selecting the shortest path with the largest number of passengers that enjoy the direct service till the number of routes generated equals a predefined number. Most recently, Kilic and Gok [14] formulated an initial route set procedure based on the edge usage statistics. The statistics was obtained by computing all the shortest paths in the graph and then calculating the edge usage score based on the total traffic passing over a single edge. Each edge score was weighted by the inverse length of the edge and normalized. The normalized value is set as the discrete probability for the edge. Each route is then expanded by selecting adjacent edges with predefined probabilities. Their method improves upon published results for Mumford's larger networks.

In an initial route set generation procedure, multiple initial route sets are required for population-based optimization methods. Candidate route set generation methods generate a set of candidate routes. Fan and Machemehl [7] and Fan and Machemehl [29] modified Yen's k-shortest path algorithm [30] to generate all feasible routes whose distances satisfy the minimum and maximum length constraints for all nonzero demand origin-destination pairs. Similarly, Arbex and Cunha [31] generated a route database: for each origin-destination pair with nonzero travel demand, the associated OD path that corresponds to the ideal shortest direct travel time is added to the database as a bus route, along with those routes whose travel time does not exceed the fastest time by a maximum route detour factor specified by the planner. In addition, Cipriani et al. [6] developed a heuristic algorithm to generate three different and complementary sets of rational and realistic routes. A set of candidate routes is formed by using Newton gravity theory and a special shortest path procedure [10]. In Afandizadeh et al. [4], paths with lengths within a desirable limit (i.e., travel time up to 1.2 shortest path travel time) are selected as initial paths. The best subset of the large candidate route set is selected by an optimization procedure.

2.2. Optimization Algorithm. After the generation of an initial route set, heuristics- or metaheuristics-based optimization algorithms are utilized to improve solution quality, e.g., genetic algorithm (GA), simulated annealing, bee colony algorithm, artificial bee colony, particle swarm optimization, nondominated sorting genetic algorithm-II, tabu search, memetic algorithm, differential evolution, and selection hyperheuristics. For candidate route set approaches, the most commonly used algorithm is GA.

A GA with seven proposed genetic operators is designed to facilitate the search within a reasonable amount of time [8]. The model designs the bus routes in two phases: the route improvement algorithm using genetic operators and coordination of headways to improve the efficiency of the network. The results showed that the proposed model is more efficient than the binary-coded genetic algorithm benchmark. Fan and Machemehl [7] developed a GA to select an optimal set of routes from the candidate route set solution space. Numerical results showed that genetic algorithms outperform local search methods with multiple starting points and provide no worse solution quality than either simulated annealing or tabu search. Cipriani et al. [6] proposed a parallel GA for finding a suboptimal set of routes with the associated frequencies. Numerical experiments carried out on network of the city of Rome showed that transit demand can be served effectively by a bus network composed of a lower number of lines. Chew et al. [32] developed a GA to address the biobjective transit routing problem. The computational results for Mandl's network showed that the proposed algorithm performs better than the previous best published results in most cases. Afandizadeh et al. [4] applied GA to the city of Mashhad and the results indicated highly effective improvement in terms of travel time and fleet size. Nayeem et al. [15] developed a GA with elitism to optimize the size of satisfied passengers, the overall travel time of served passengers, and the total number of transfers. Owais and Osman [28] proposed a GA to tackle the multiobjective transit network design problem by optimizing average travel time and the required bus fleet.

Besides, Fan and Machemehl [29] used a simulated annealing algorithm for the first time to solve the optimal bus transit route network design problem at the distribution node level. Zhao [33] proposed a stochastic local search method based on simulated annealing and fast descent search. Numerical experiments showed that the proposed method is capable of tackling large-scale transit network design optimization problems and producing results in a reasonable amount of time. Later, Zhao and Zeng [18] developed a metaheuristic search scheme that combines simulated annealing, tabu, and greedy search methods to optimize transit route network structure, vehicle headways, and timetables. Kilic and Gok [14] applied hill climbing algorithm and tabu search algorithm, respectively, on the transit network design problem. The experiment results for Mumford's four larger networks showed improvements over results in Mumford [27].

In the last decade, Nikolic and Teodorovic [16] designed a bee colony optimization algorithm for transit network design. This algorithm uses a similarity among the way in which bees in nature look for food and the way in which optimization algorithms search for an optimum of a combinatorial optimization problem. The numerical experiments performed on Mandl's network and a larger instance showed that BCO algorithm is competitive with other approaches in the literature, and it can generate high-quality solutions. Szeto and Jiang [34] developed a hybrid artificial bee colony algorithm to solve the transit route and frequency design problem. The memetic algorithm (MA) is one of the recent growing evolutionary computation algorithms. Zhao et al. [35] developed a MA algorithm to design both optimal route configuration and service frequency for the urban transit network. Compared with traditional algorithms, the results obtained by MA demonstrates that the proposed algorithm could improve the computational performance. Kechagiopoulos and Beligiannis [13] developed an optimization algorithm based on particle swarm optimization.

Most recently, Buba and Lee [36] applied differential evolution algorithm and a repair mechanism for the multiobjective transit routing problem. Numerical experiments are performed on Mandl's network and four larger networks of Mumford [27]. Buba and Lee [37] proposed a hybrid differential evolution with particle swarm optimization to design the transit network structure and associated frequencies. The computational results showed improvements over results in most of the previous studies. Ahmed et al. [38] evaluated the performance of a set of selection hyperheuristics on the route design problem of bus networks, with the goal of minimizing the passengers' travel time and the operator's costs.

3. Problem Definition

The transit network design problem (TNDP) can be defined by an undirected graph $G = (V, E)$ and a travel demand matrix D. V is the set of vertices $\{v_1, v_2, \ldots, v_n\}$ representing predefined pick-up/drop-off points. E is the set of edges $\{e_1, e_2, \ldots, e_m\}$ standing for direct transport edges between vertices. Each edge $e = (i, j) \in E$ is associated with a travel time $l(i, j)$ denoting the necessary time for a vehicle to travel from vertex i to vertex j. ($l(i, j) = \infty$, if $e(i, j) \notin$). Travel demand matrix $D = \{d_{ij}, i, j \in V\}$ is given to present the travel demand between every pair of vertices; d_{ij} denotes the number of trips per time unit from vertex i to vertex j. A route r can be expressed as a sequence of adjacent vertices $(v_1, v_2, \ldots, v_q)$. A solution set consists of a set of routes $R = \{r_1, r_2, \ldots, r_K\}$. Let $G_R = (V_R, E_R)$ be the subgraph corresponding to route set R, where $V_R = \cup_{r \in R} V_r = \cup_{r \in R} \cup_{i=1}^{|r|} v_i$ is the set of vertices of route set R ($|r|$ is the number of vertices in node sequence r). $E_{r_k} = \{e(r_{ki}, r_{k,i+1}): i = 1, \ldots, |r_k| - 1\}$ denotes the set of edges of route r_k and $E_R = \cup_{r \in R} E_r$. To compare our outcomes with existing results, this study mainly focuses on the generation of route sets and leaves the frequency setting problem for future studies.

From the viewpoint of passengers, the travel time and number of transfers are the major concerns. The average travel time ATT (including in-vehicle travel time and a transfer penalty) is selected as the indicator of the service level to depict the passenger cost. Since the average travel time includes the transfer cost, the number of transfers is implicitly included in the ATT. The objective function from the passengers' viewpoint is defined as follows:

$$\min ATT(R, p) = \frac{\sum_{(i,j) \in V^2} d_{ij} t_{ij}(R, p)}{\sum_{(i,j) \in V^2} d_{ij}}, \tag{1}$$

where d_{ij} denotes the number of trips per time unit from vertex i to vertex j. p is the transfer cost representing the inconvenience of moving from one vehicle to another, and $t_{ij}(R, p)$ represents the shortest travel time from vertex i to vertex j by using route set R. To be consistent with former research studies, the transfer cost p is set as 5 minutes [4, 6, 16, 27, 32]. To compare our results with existing published results, we assume that passengers always choose the shortest path to reach their destinations.

From the operator's perspective, the sum of the lengths of all routes in a route set is directly proportional to the fleet size, which is an important component in the operator cost [26]. Thus, the total route length (TRL) of a route set is taken as a proxy. The objective function from the operator's viewpoint is given as follows:

$$\min TRL(R) = \sum_{r \in R} L(r), \tag{2}$$

where $L_r = \sum_{i=2}^{r} l(r_{i-1}, r_i)$ represents the length of route r.

Mathematically, this optimization problem can be stated as

$$\min ATT(R, p) = \frac{\sum_{(i,j) \in V^2} d_{ij} t_{ij}(R, p)}{\sum_{(i,j) \in V^2} d_{ij}}, \tag{3}$$

$$\min TRL(R) = \sum_{r \in R} L(r), \tag{4}$$

$$\text{s.t. } N_{\min} \le |r| \le N_{\max}, \quad \forall r \in R, \tag{5}$$

$$\cup_{r \in R} V_r = V, \tag{6}$$

$$G_R = \left(\cup_{r \in R} V_r, \cup_{r \in R} E_r \right) \text{is connected}, \tag{7}$$

$$\left| V_r \right| = q_r, \quad \forall r \in R, \tag{8}$$

$$\sum_{i=2}^{r} l(r_{i-1}, r_i) \ne \infty, \quad \forall r \in R. \tag{9}$$

Constraint (5) limits the number of vertices of each route. Constraint (6) ensures that all vertices in V can be found in at least one route. Constraint (7) ensures that every vertex is reachable from other vertices in the subgraph defined by R [36]. Constraint (8) ensures that each node can only be visited by a route at most once. Recall that route r_k is described as a node sequence $(v_{k1}, v_{k2}, \ldots, v_{kq})$, where q denotes the number of nodes in node sequence r_k. Constraint (9) ensures that all consecutive vertices are connected by an existing edge ($l(i, j) \ne \infty$). The uniqueness of route r in R is ensured through the definition of route set R.

4. Solution Method

4.1. Overall Solution Framework. In the developed solution method, a novel initial route set generation algorithm is constructed to produce high-quality initial route sets, while a route set size alternating heuristic is developed to switch the route set size of a solution by adding or deleting an entire route. These two algorithms are embedded into the framework of NSGA-II to search for the approximate Pareto front in terms of the operator cost (i.e., the total route length, TRL) and the passenger cost (i.e., the average travel time, ATT).

Multiple high-quality initial solutions are firstly produced by the initial route set generation algorithm. Figure 1 presents the flowchart of the proposed initial route set algorithm. Two objective values are then calculated for each initial solution to evaluate the fitness of initial solutions. And all initial solutions are regarded as parents and used to generate the offspring through genetic operators (i.e., selection, crossover, and mutation operators). In this work, the application of crossover operator generates only one offspring. Therefore, the number of solutions in the offspring remains the same as the parents. After the mutation, the route set size alternating heuristic is applied to the offspring to change the number of routes in an offspring by adding/deleting if its solution quality has not been improved for a

predefined number of generations. In order to select the individuals of the next generation, the NSGA-II is subsequently applied to the parents and the offspring. For extensive surveys on NSGA-II, readers are referred to Deb et al. [39]. This process is repeated until the stopping criteria are satisfied.

The following three datasets are needed to perform route combination algorithm (RCA): (i) initial candidate route set (containing all the shortest paths) denoted by P; (ii) transit demand served by each candidate route and denoted by P_v; and (iii) the origin-destination pairs served by each candidate route and denoted by P_d. Note that due to the symmetry of demand and travel time, only triangular elements are used.

4.2. Initial Route Set Generation Algorithm. The proposed initial route set generation algorithm is based on the observation that maximizing the travel demand that can be satisfied directly (without transfers) is a key component in the generation of an initial route set. This algorithm consists of two procedures: absorption and combination. As the algorithm is based on the combination of routes, it is called RCA. These procedures generate an initial route set based on the following parameters: SP, D, and G. SP contains all the shortest paths connecting every origin-destination pair. G is the undirected graph. An example in which a graph and a matrix D is used to illustrate the initial route set generation algorithm is shown in Figure 2. The numbers adjacent to the links represent the travel times in minutes.

Table 1 gives the dataset structure that is needed to generate an initial route set. The second column gives all the shortest paths between the vertices in the example graph. OD pairs served by the shortest paths are shown in the third column. And the last column contains the travel demands served directly by each shortest path. For example, for the second route in Table 1, the shortest path (1, 2, 3) starts from vertex 1, goes through vertex 2, and terminates at vertex 3. This path serves the travel demand of one OD pair (1-3).

4.2.1. Absorption Procedure. If a route is completely included in another route, we consider that this route can be absorbed by another route. The basic principle of absorption is to eliminate routes which are completely included in another route. By doing this, the number of routes in R and the total route length can be efectively decreased. The absorption procedure is composed of four steps:

Step 1. Sort the routes in R in the order of the number of vertices of each route. The routes that serve no demand are eliminated. Let M be the total number of routes in R and set $m = 1$. For example, the routes in Table 1 are sorted and given in Figure 3(a). Particularly, the third route (1, 4) in Table 1 is eliminated since it serves no demand.

Step 2. Find all the routes (except route m) that can absorb route m. Let Q be the set of routes that can absorb route m. If $m = M$, stop; otherwise, go to Step 3.

Step 3. If Q is empty, set $m = m + 1$, and return to Step 2; otherwise, based on the travel demands served by routes in Q, a probability p_q is appointed to each route q and reflects the probability of each route to be selected as the route to absorb route m. The probability of selecting route q is calculated by

$$p_q = \frac{D(q)}{\sum_{i \in Q} D(i)}. \tag{10}$$

Step 4. Use the roulette wheel selection method to select a route from Q based on the values of probabilities p_q. As soon as a route is selected to absorb route m, the ODs and demands served by route m are assigned to route q. Then, we need to delete route m from R and return to Step 2.

An example for the absorption procedure is shown in Figure 3. In Figure 3(a), the first route (1, 2) is completely included in route 6 (1, 2, 3), namely, the first route is absorbed by route 6 and the travel demand of the first route can be served by route 6. Therefore, we need to erase the first route from R.

In Figure 3(b), the second route (2, 3) can be absorbed by three routes (routes 6, 8, and 9). According to equation (10), the probabilities of selecting the three routes 6, 8, and 9 are 0.5, 0.375, and 0.125, respectively. Assume that route 6 is selected. Then, the ODs (2-3) and demands (4) served by route 2 are assigned to route 6. As presented in Figure 3(c), route 6 serves three ODs (1-2, 1-3, 2-3), and the total demand is 12. In the following steps, routes 3, 4, and 5 are absorbed by routes 8, 9, and 7, separately.

From Figure 3, it can be seen that the number of routes in R is decreased from 9 to 4, but the travel demand served by each route is increased. After the absorption procedure, $3 * N_{rou}$ routes with high demands are selected as the input datasets for the following procedure, where N_{rou} is the number of routes in a solution set (predefined according to Table 2).

4.2.2. Combination Procedure. The combination procedure aims to construct new routes with pairs of routes in R. The basic principle of the combination procedure is to combine pairs of routes with high demands. In this way, we can simultaneously decrease the number of routes and increase the travel demands served by one route. Before constructing new routes, pairs of routes are selected. The selection of one pair of routes relies on the relationship between two routes. In the light of the positions of vertices shared by two routes, the relationships between two routes can be classified into six types. Figure 4 shows the six types and their combined results.

If two routes fall into types of 1, 2, 3, or 5, then these two routes can be combined. The combination procedure is composed of five steps:

Step 1. Sort the routes in R in descending order based on the transit demand of each route. Let M be the total

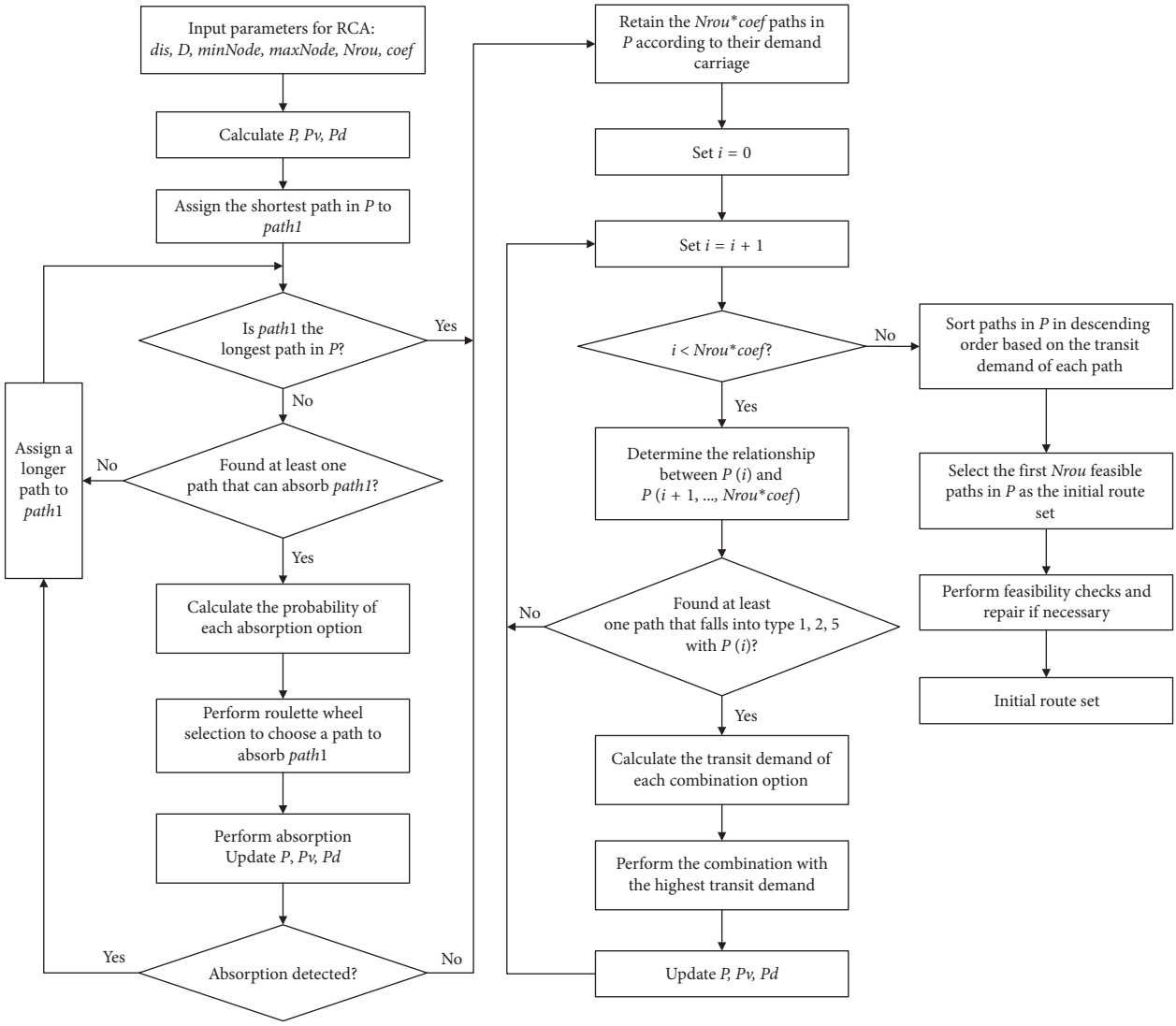


Figure 1: Initial route set generation procedure.

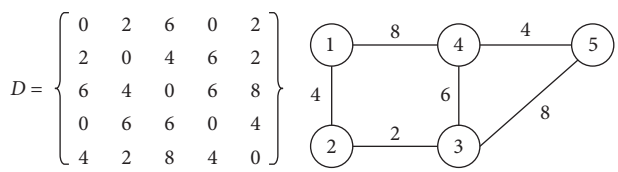


Figure 2: Illustrative example.

Table 1: Dataset structure of the example.

#	Route	OD	Demand
1	1, 2	1-2	2
2	1, 2, 3	1-3	6
3	1, 4	1-4	0
4	1, 4, 5	1-5	4
5	2, 3	2-3	4
6	2, 3, 4	2-4	6
7	2, 3, 5	2-5	2
8	3, 4	3-4	6
9	3, 5	3-5	8
10	4, 5	4-5	4

#	R	OD	D
1	1, 2	1-2	2
2	2, 3	2-3	4
3	3, 4	3-4	6
4	3, 5	3-5	8
5	4, 5	4-5	4
6	1, 2, 3	1-3	6
7	1, 4, 5	1-5	4
8	2, 3, 4	2-4	6
9	2,3,5	2-5	2

(a)

#	R	OD	D
~~1~~	~~1, 2~~	~~1-2~~	~~2~~
2	2, 3	2-3	4
3	3, 4	3-4	6
4	3, 5	3-5	8
5	4, 5	4-5	4
6	1, 2, 3	1-2, 1-3	8
7	1, 4, 5	1-5	4
8	2, 3, 4	2-4	6
9	2, 3, 5	2-5	2

(b)

#	R	OD	D
~~1~~	~~1, 2~~	~~1-2~~	~~2~~
~~2~~	~~2, 3~~	~~2-3~~	~~4~~
3	3, 4	3-4	6
4	3, 5	3-5	8
5	4, 5	4-5	4
6	1, 2, 3	1-2, 1-3, 2-3	12
7	1, 4, 5	1-5	4
8	2, 3, 4	2-4	6
9	2, 3, 5	2-5	2

(c)

#	R	OD	D
~~1~~	~~1, 2~~	~~1-2~~	~~2~~
~~2~~	~~2, 3~~	~~2-3~~	~~4~~
~~3~~	~~3, 4~~	~~3-4~~	~~6~~
4	3, 5	3-5	8
5	4, 5	4-5	4
6	1, 2, 3	1-2, 1-3, 2-3	12
7	1, 4, 5	1-5	4
8	2, 3, 4	2-4, 3-4	12
9	2, 3, 5	2-5	2

(d)

#	R	OD	D
~~1~~	~~1, 2~~	~~1-2~~	~~2~~
~~2~~	~~2, 3~~	~~2-3~~	~~4~~
~~3~~	~~3, 4~~	~~3-4~~	~~6~~
~~4~~	~~3, 5~~	~~3-5~~	~~8~~
5	4, 5	4-5	4
6	1, 2, 3	1-2, 1-3, 2-3	12
7	1, 4, 5	1-5	4
8	2, 3, 4	2-4, 3-4	12
9	2, 3, 5	2-5, 3-5	10

(e)

#	R	OD	D
~~1~~	~~1, 2~~	~~1-2~~	~~2~~
~~2~~	~~2, 3~~	~~2-3~~	4
~~3~~	~~3, 4~~	~~3-4~~	~~6~~
~~4~~	~~3, 5~~	~~3-5~~	~~8~~
~~5~~	~~4, 5~~	~~4-5~~	~~4~~
6	1, 2, 3	1-2, 1-3, 2-3	12
7	1, 4, 5	1-5, 4-5	8
8	2, 3, 4	2-4, 3-4	12
9	2, 3, 5	2-5, 3-5	10

(f)

Figure 3: Illustration of the absorption procedure. (a) Route 1⟶6. (b) Route 2⟶6/8/9. (c) Route 3⟶8. (d) Route 4⟶9. (e) Route 5⟶7. (f) None.

Table 2: Properties of five networks.

Network	Location	Vertices	Edges	Number of routes	Vertices' route	LB_{ATT} (min)
Mandl	Switzerland	15	21	4-8	2-8	10.0058
Mumford0	—	30	90	12	2-15	13.0121
Mumford1	Yubei	70	210	15	10-30	19.2695
Mumford2	Brighton	110	385	56	10-22	22.1689
Mumford3	Cardiff	127	425	60	12-25	24.7453

number of routes in R and set $m = 1$. For example, the routes in Figure 3 are sorted and given in Figure 5(a).

Step 2. Find all the routes (except route m) that can combine with route m. Let Z be the set of routes that can combine with route m. If $m = M$, go to Step 5; otherwise, go to Step 3.

Step 3. If Z is empty, set $m = m + 1$, and return to Step 2; otherwise, according to the travel demands served by routes in Z, a probability p_z is distributed to each route z and depicts the probability of each route to be selected as the route to combine with route m. The probability of selecting route z is computed by

$$p_z = \frac{D(z) + D(m)}{\sum_{i \in Z} (i) + D(m)}. \tag{11}$$

Step 4. Use the roulette wheel selection method to choose a route from Z based on the values of probabilities p_z. As soon as a route z is selected, route m and route z are combined into a new route according to Figure 4. The new route serves the travel demands and ODs of both routes. Then, we should delete routes m and z from R and return to Step 2.

Step 5. Sort routes again in R in descending order by the transit demand of each route. The initial route set is formed of the first N_{rou} (the number of routes in a route set) feasible routes in R. Check the feasibility of the initial route set by using feasibility checks described in Section 4.3.

An example for the combination procedure using the output of the absorption example is shown in Figure 5. First, the routes are sorted in descending order based on the travel

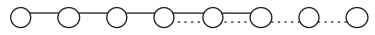
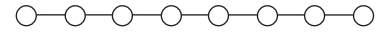
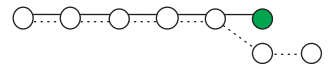
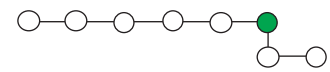
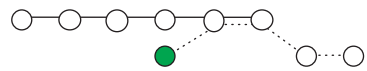
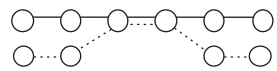
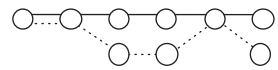
#	Definitions	Examples	Combined results
1	The shared vertices are on the first part of route 1 and on the last part of route 2.		
2	The shared vertices are on the first parts of both routes.		
3	The shared vertices are on the first part of route 1 and on the middle part of route 2.		
4	The shared vertices are on the middle parts of both routes.		None
5	There is no vertex shared by both routes.		
6	The shared vertices are not continuous on both routes.		None

Figure 4: Six kinds of route relationships and their corresponding combined results.

#	R	OD	D
1	1, 2, 3	1-2, 1-3, 2-3	12
2	2, 3, 4	2-4, 3-4	12
3	2, 3, 5	2-5, 3-5	10
4	1, 4, 5	1-5, 4-5	8

(a)

#	R	OD	D
1	1, 2, 3, 4	1-2, 1-3, 2-3, 2-4, 3-4	24
2	2, 3, 5	2-5, 3-5	10
3	1, 4, 5	1-5, 4-5	8

(b)

#	R	OD	D
1	1, 2, 3, 4, 5	1-2, 1-3, 2-3, 2-4, 2-5, 3-4, 3-5	34
2	1, 4, 5	1-5, 4-5	8

(c)

Figure 5: Illustration of combination procedure. (a) Route 1 + 6. (b) New route 1 + 2. (c) None.

demand served directly by each route. The relationship between the first route and other routes are determined, that is, routes 1 and 2 fall into type 1, routes 1 and 3 belong to type 1, and routes 1 and 4 are in type 2. With the aid of equation (11), the probabilities of selecting these three routes are 0.36, 0.33, and 0.30. We assume that route 2 is chosen to combine with route 1, and the new route is constructed according to Figure 4. It should be noted that the demand covered by the new route is doubled. The total route length of R is also decreased from 36 minutes to 34 minutes. The number of routes also decreases from 4 to 3, as shown in Figure 5(b). Similarly, in Figure 5(b), routes 1 and 2 are combined.

From Figure 5, it is not uncommon that not only the number of routes in R is decreased from 4 to 2, but also the travel demand served by each route is increased. Hence, the result of the combination procedure can be used as an initial route set for optimization algorithms.

4.3. Feasibility Checks. The TNDP is a multiconstrained problem, and the feasibility of the solutions should be checked carefully. Inspired by [27, 31], our proposed feasibility checks are detailed as follows:

(1) Route feasibility check: check if every route is feasible, that is, no repeated vertices appear in a route, and the number of vertices in a route must be within a predefined range.

(2) Route uniqueness check: if one route is the same as a segment of another route or another entire route, we regard this route as nonunique. As a result, this route is directly eliminated.

(3) Area coverage check: check if all the vertices appear in the solution set at least once. This process constructs an appeared-vertex list by searching all the routes and checks if every vertex appears in the list.

(4) Network connectivity check: the purpose of this procedure is to determine whether a route is connected to the transit network.

After feasibility checks, a repair mechanism is applied to repair infeasible solutions. It should be noted that the generated initial solutions are always feasible for the first two feasibility checks. Only infeasible initial solutions which fail area coverage check are repaired. Other infeasible solutions (network connectivity check) are simply discarded during

initial solution generation and the initial route set generation algorithm is called again. The repair mechanism for initial solutions is described as follows.

The repair mechanism randomly selects a route from the solution and then applies insert operation. The insert operation constructs a list of candidate nodes by searching all positions and checking the existence of the edges (in the origin graph) between the used and unused nodes of the modified route, and then it randomly selects one of the candidate nodes and inserts it into the route. The above procedure is repeated until all nodes can be found in the route set. If there are nodes that cannot be inserted into the initial route set, this solution is discarded. Note that this procedure is also applied to the output solutions of route set size alternating heuristic.

4.4. Route Operators. Due to the complex structure of the route set, one crossover and three mutation operators are designed to facilitate the generation of offspring.

The route crossover operator is developed to exchange two routes between two individuals. Each individual of the current generation is selected in turn as the first parent. The second parent is randomly determined. Then, two routes from each parent are randomly determined and exchanged.

Three mutation operators, namely, the sequence, insert, and delete mutation operators, are proposed in this study to mutate the route structure of an individual. Specifically, the sequence mutation operator is developed to replace a random sequence of a route with a k-th shortest path. For the inserted mutation operator, all the positions in the route will be checked to determine whether there exists a vertex that can be connected to two adjacent vertices in the route with only two edges. Then, a suitable vertex will be randomly selected and inserted into the route. Comparatively, the delete mutation operator constructs a candidate list of vertices and randomly deletes a vertex.

Before a crossover/mutation operation is made, the result of crossover/mutation will be checked for route feasibility and route uniqueness. If these checks return true, the crossover/mutation is made. The solutions are feasible for the first two checks all the time. After all solutions go through crossover and mutations, the last two feasibility checks are performed for child solutions.

If the area coverage constraint is not satisfied, a route in the child solution is selected from the solution and then insert operation is applied. The insert operation constructs a list of candidate nodes by searching all positions and checking the existence of the edges (in the origin graph) between the used and unused nodes of the modified route, and then it randomly selects one of the candidate nodes and inserts it into the route. The above procedure is repeated until all unused nodes are inserted. If there are nodes that cannot be inserted into the child solution, the child solution is replaced by its parent.

If the subgraph defined by a child solution is not connected, the child solution is substituted by its parent.

4.5. Route Set Size Alternating Heuristic. As mentioned before, the route set size is regarded as an endogenous variable. Therefore, for every individual that survives into the next generation, a descent route set size alternating heuristic is employed to change the route set size. To ensure efficiency and secure potential individuals, only individuals who have not been improved for a predefined number of generations (set as 5 in this article) will go through this heuristic. The proposed route set size alternating heuristic comprises two procedures: route addition and route deletion. For each selected route set, either route addition or route deletion is randomly selected with the same probability. Solutions going through this heuristic are checked for area coverage constraint and repaired when necessary using the repair mechanism described in Section 4.3. If the solution is not feasible after repair, the changes made by route set size alternating heuristic are undone. The process of this heuristic is detailed as follows.

4.5.1. Route Addition. For each route set R, a deviation ratio is calculated for every nonzero transit demand to estimate the quality of their trips. Then, the OD pairs with a low service quality are selected. Finally, the greedy algorithm proposed by Nikolic and Teodorovic [16] is used to determine the most suitable route to add into the route set R.

The quality indicator for passengers between vertices i and j is defined as follows:

$$\Delta_{ij}(R) = \frac{t_{ij}(R)}{t_{ij}(G)}. \tag{12}$$

Using routes of R, this represents a deviation ratio of the minimum travel time from the theoretical best value (which only depends on the road network G). $t_{ij}(R)$ is computed by using an all-or-nothing assignment approach, and $t_{ij}(G)$ is the in-vehicle travel time of the shortest path in G between vertices i and j. According to this formulation, lower values of $\Delta_{ij}(R)$ are desired. Then, the OD pairs with a high value of $\Delta_{ij}(R)$ are considered as not well served. The set of these OD pairs can be derived as follows:

$$S = \left\{(i,\ j) \mid \Delta_{ij}(R) \geq \text{threshold}\right\}. \tag{13}$$

The OD pair that has the highest value of $ds(i, j)$ is used to generate the most suitable route. This route is obtained by connecting this OD pair with the shortest path. By inserting this route into R, we can increase the percentage of the satisfied travel demand with zero transfer.

4.5.2. Route Deletion. The route deletion procedure is proposed to delete one route from the route set R. This procedure is based on the analysis of the allocation of each travel demand. The route deletion procedure works in the following way. For each route set R that is selected to perform the route deletion procedure, the route whose removal from the route set has the least influence on the quality of the current route set is determined. Then, this route should be deleted from the route set R. Based on the analysis of the allocation of each transit demand, every route

r_k is assigned a weight w_k, which reflects the importance of each route. And this weight is calculated as follows:

$$w_k = \sum_{(i,j)\in V^2} \frac{t_{ij}(r_k) d_{ij}}{T_k}, \quad (14)$$

where T_k is the round-trip time of route r_k. d_{ij} is the transit demand per unit time from vertex i to vertex j. $t_{ij}(r_k)$ represents the in-vehicle travel time on route r_k for passengers traveling between vertices i and j. The value of $t_{ij}(r_k)$ is obtained as follows:

(1) If there is a direct service between vertices i and j using route r_k, the $t_{ij}(r_k)$ value is the in-vehicle travel time.

(2) If the passengers between vertices i and j cannot complete their trips with only route r_k, the $t_{ij}(r_k)$ value is the in-vehicle travel time from the boarding/alighting stop to the transfer stop on route r_k.

(3) If the passengers between vertices i and j can complete their trips without route r_k, the $t_{ij}(r_k)$ value is 0.

This definition has significant implications: (i) more passengers and longer in-vehicle travel time spent on route r_k mean more contribution to the weight w_k; (ii) the weight of route r_k rapidly diminishes when the length route r_k increases. After the weight calculation procedure, the least important route is determined and removed from the route set R.

5. Experimental Results

To illustrate the effectiveness of the proposed initial route set generation algorithm and demonstrate the effectiveness of the route set size alternating heuristic on solution quality, in this section, various experiments have been conducted and these solutions are systematically compared with existing published results. The proposed algorithms are tested on Mandl's benchmark network [25] and four larger instances in Mumford [27]. Table 2 gives a brief description of these five networks. The solutions are evaluated by the percentages of satisfied demand with 0, 1, and 2 transfers (d_0, d_1, d_2), the percentages of unsatisfied demand (d_{un}), and the average travel time (ATT), as well as the total route length (TRL). All the experiments are performed on a standard PC that has an Intel Core i5-3470 chip with 3.20 GHz and 8 GB RAM.

The number of individuals in a generation is set to 20. The probabilities of all four route operators are set to $1/|R|$, $1/|R|$, 0.005, and 0.005. NSGA-II terminates when the approximate Pareto fronts are not improved for 100 generations.

5.1. Effect of the Proposed Initial Route Set Generation Algorithm. To illustrate the effectiveness of the proposed initial route set generation algorithm, we compare our initial solutions with reported initial solutions for Mandl's network [25] and Mumford's large networks [27].

5.1.1. Mandl's Network. In this section, we compare our initial solutions with reported initial solutions in Nikolic and Teodorovic [14, 16] for Mandl's network. Initial route sets that contain 4, 6, 7, and 8 routes for Mandl's network were obtained. Other initialization procedures mentioned in the literature review are not compared, since the initial solutions produced by their methods are not reported in their works. The solutions are plotted on the two-dimensional space defined by the objectives (ATT and TRL). Note that the 4-route solution in Nikolic and Teodorovic [16] is not plotted as it is not feasible.

In Figure 6, it can be seen that the initial solutions obtained by the proposed algorithm have small values of ATT, while other solutions have small values of TRL. The results from greedy algorithm have a better trade-off in terms of both objective function values. However, the greedy algorithm can only produce one specific solution for a network and $|R|$ since it contains no randomization. Greedy algorithm may not be suitable for population-based optimization algorithms which require multiple different initial solutions.

Table 3 compares our initial solutions for Mandl's network with other approaches in terms of six parameters (i.e., d_0, d_1, d_2, d_{un}, ATT, and TRL). It can be seen that the initial solutions obtained by RCA have better values of ATT than the ones in Nikolic and Teodorovic [16] and Kilic and Gok [14]. 4-route solution in Nikolic and Teodorovic [16] has a smaller ATT. However, 6.96% demand are unsatisfied in that solution. On the other hand, the initial solutions of Nikolic and Teodorovic [16] and Kilic and Gok [14] have lower values of TRL. For the initial route sets, the increase of the number of routes $|R|$ has a direct impact on the values of ATT and TRL. For example, when $|R|$ increases from 4 to 8, the value of ATT of the solutions acquired by RCA gradually decreases from 11.01 to 10.12, while the TRL value increases from 116 to 259. It should be noted that the values of d_2 and d_{un} in all the initial solutions acquired by RCA are zero. This indicates that all the passengers can reach their destination with at most one transfer.

5.1.2. Mumford's Networks. In this section, we further compare our initial solutions for Mumford's large networks with existing initial solutions [14]. Other initialization procedures are not compared, since the initial solutions of their methods are not reported in their works. We also plot the final solutions in Mumford [27] and Kilic and Gok [14] for four larger networks.

In Figure 7, the initial solutions generated by the proposed algorithm outperform other initial solutions in terms of both objectives. The initial solutions generated by the proposed algorithm have similar TRL values, and the ATT values of our results are generally smaller than the initial solutions in Kilic and Gok [14]. Note that for network Mumford2 and network Mumford3, the initial solutions produced by the proposed algorithm are even better than the final solutions in Mumford [27] and Kilic and Gok [14], which implies that the proposed initial route set generation algorithm can produce high-quality initial solutions for large

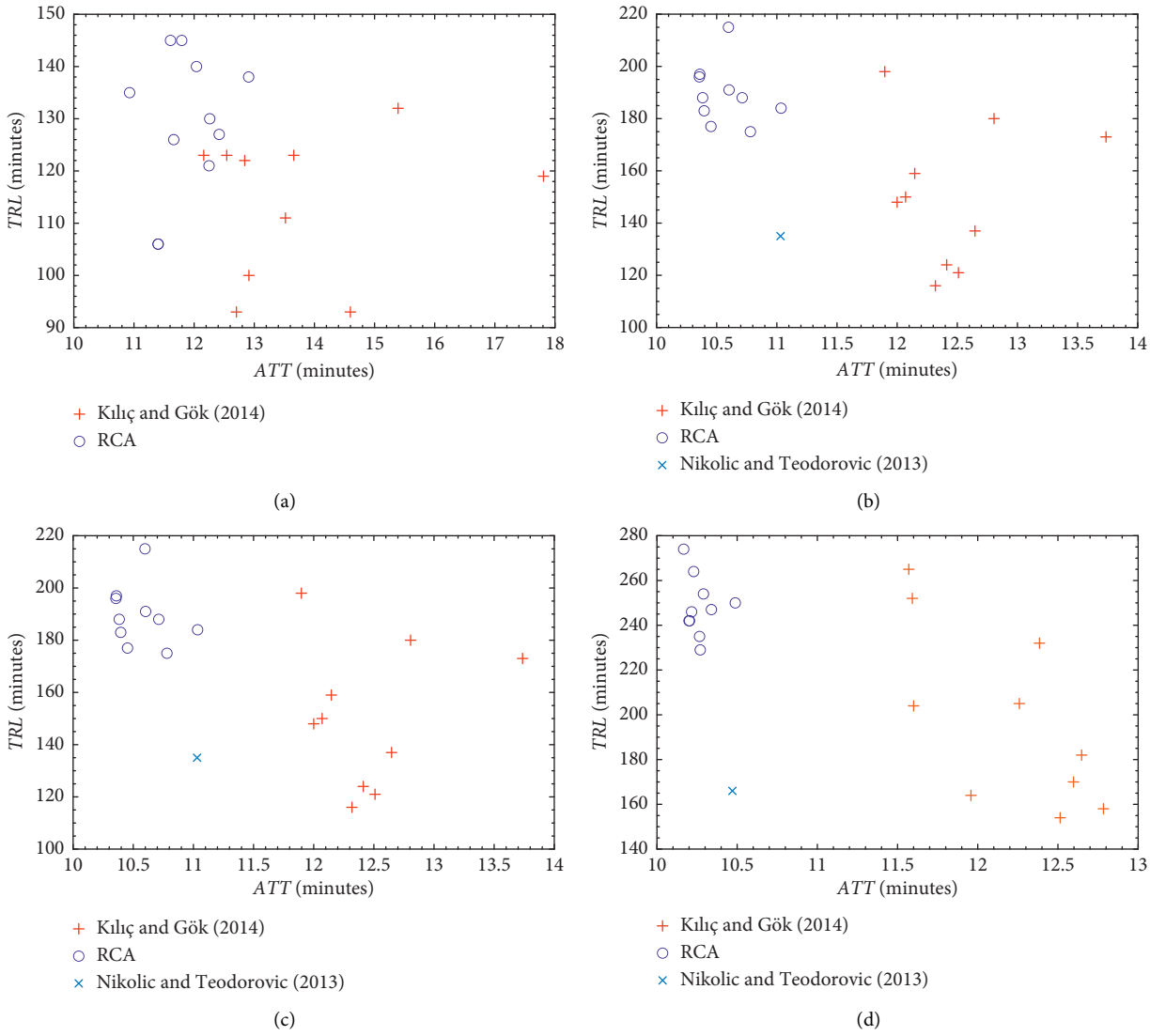


Figure 6: Comparisons of the initial solutions for Mandl's network. (a) $|R| = 4$. (b) $|R| = 6$. (c) $|R| = 7$. (d) $|R| = 8$.

networks. For network Mumford3, some initial solutions produced by Kilic and Gok [14] are also dominating the final solution produced by Mumford [27]. The initial solutions used in Mumford [27] are randomly generated to encourage coverage and connectivity of the route network. The importance of the initial route set generation algorithm may be overlooked.

We further compare the initial solutions in Kilic and Gok [14] and ours for Mumford's networks in terms of six parameters (i.e., d_0, d_1, d_2, d_{un}, ATT, and TRL). The results are integrated and presented in Table 4. The number of routes in a solution is set as the recommended value in Table 2.

From the results in Table 4, it can be seen that the solutions obtained by RCA are better than the ones obtained by Kilic and Gok [14] in terms of both objectives (i.e., ATT and TRL). Compared to the solutions in Kilic and Gok [14], the solutions obtained by RCA for the four networks can reduce the ATT by 8.23%, 8.10%, 6.46%, and 5.88%, respectively, and reduce the TRL by 6.92%, 24.06%, 6.07%, and 9.12%, separately. These results demonstrate that the RCA can significantly improve the quality of the initial solutions for large networks when compared with the initial procedure proposed by Kilic and Gok [14]. Therefore, the RCA is used in the following experiments to generate high-quality initial solutions for the optimization procedure.

We also compare the computation time reported in Kilic and Gok [40] and ours to further display the superiority of the proposed algorithm. The corresponding outcomes are depicted in Table 5. Notwithstanding the difference of data structure and subroutines, it can be seen that RCA is very efficient, as it takes only 3 seconds to generate an initial solution for the largest network (i.e., Mumford3). On the contrary, the computation time for network Mumford3 in Kilic and Gok [14] is 8 hours. The computation resource needed for RCA is negligible.

5.2. Effect of the Route Set Size Alternating Heuristic. This experiment compares the approximate Pareto fronts obtained from the NSGA-II with and without a route set size

Table 3: Initial solutions for Mandl's network compared to other approaches.

$\|R\|$	Parameters	Nikolic and Teodorovic [16]	Kilic and Gok [14]	RCA
4	d_0 (%)	80.47	77.71	91.71
	d_1 (%)	12.33	20.3	8.29
	d_2 (%)	0.51	1.99	0
	d_{un} (%)	**6.69**	0	0
	ATT	10.22	12.57	11.01
	TRL	99	114	116
6	d_0 (%)	87.73	83.75	95.57
	d_1 (%)	11.75	15.22	4.43
	d_2 (%)	0.51	1.03	0
	d_{un} (%)	0.00	0	0
	ATT	11.03	11.65	10.27
	TRL	135	135	172
7	d_0 (%)	90.62	75.47	97.5
	d_1 (%)	8.86	21.51	2.5
	d_2 (%)	0.51	3.02	0
	d_{un} (%)	0.00	0	0
	ATT	10.57	11.89	10.16
	TRL	152	151	206
8	d_0 (%)	91.91	85.36	98.46
	d_1 (%)	7.58	14.19	1.54
	d_2 (%)	0.51	0.45	0
	d_{un} (%)	0.00	0	0
	ATT	10.47	10.9	10.12
	TRL	166	195	259

alternating heuristic. To make the comparison as fair as possible, the NSGA-II with variable $|R|$ starts with the same initial solutions for the NSGA-II with fixed $|R|$ in Table 2 (for Mandl's network, $|R| = 6$).

5.2.1. Mandl's Network. Figures 8(a)–8(e) compare the results of the NSGA-II with variable $|R|$ and fixed $|R|$. An immediate observation is that the approximate Pareto fronts are very close. When $|R|$ increases, the best ATT value decreases as expected, while the corresponding TRL value increases. For the NSGA-II with variable $|R|$, the theoretical best value of ATT (10.0058) is reached when $|R| = 19$, and all passengers can travel with the shortest path. This is the first time a transit network with theoretical ATT is presented for Mandl's network.

In addition, Figure 8(f) compares the approximate Pareto front obtained from the NSGA-II with and without a route set size alternating heuristic. The approximate Pareto front with fixed $|R|$ is obtained by comparing all the approximate Pareto fronts with $|R| = 4$ to $|R| = 8$. It can be seen that the NSGA-II with variable $|R|$ performs as well as the NSGA-II with fixed $|R|$. This shows that the NSGA-II with variable $|R|$ leads to the same solution quality of the solutions obtained with fixed $|R|$. Moreover, the NSGA-II with variable $|R|$ can produce an approximate Pareto front with a wider range of trade-offs between the two objectives.

Table 6 summarizes the computation time of each NSGA-II. Note that the computation time is measured by the number of generations generated by each NSGA-II. As shown in Table 6, the computation time of NSGA-II with variable $|R|$ is relatively close to that of the NSGA-II with fixed $|R|$. However, to obtain the approximate Pareto fronts with different values of $|R|$, the NSGA-II with fixed $|R|$ must be performed many times with different values of $|R|$. As a result, for fixed $|R|$ approaches, the total computation time grows as the number of possible values of $|R|$ increases. This suggests that the route set size alternating heuristic can save a large amount of computation time without deteriorating the quality of the approximate Pareto front obtained.

5.2.2. Mumford's Networks. To assess the merits of our route set size alternation heuristic, we run the optimization algorithm for Mumford's four networks with the same initial solutions. To make a fair comparison, each NSGA-II was terminated when 3000 generations were produced. The fixed value of $|R|$ is set to the recommended value in Mumford [27]. The results are plotted on the two-dimensional space defined by both objectives in Figure 9.

In Figure 9, the first observation is that two curves contact at some points. The joint points represent the best solutions from the passenger's perspective with predefined $|R|$. The operator's best solutions with fixed $|R|$ are dominated by the results with variable $|R|$ except the solution for Mumford0. Specifically, for network Mumford0 (i.e., Figure 9(a)), the solutions with fixed $|R|$ coincide with the lower part of the other curve; for network Mumford1 (i.e., Figure 9(b)), the approximate Pareto front of fixed $|R|$ is overlapped with the one with variable $|R|$ on the upper part. As shown in Figure 9(c), the overlapped solutions are few. One interesting observation in Figures 9(c) and 9(d) is that the overlapped solutions locate on the upper part of the curve, rather than on the middle area. This is result of a larger $|R|$, which also reinforces the idea that it is difficult to select a suitable $|R|$ for the best solution to TNDP.

Generally, the NSGA-II with fixed $|R|$ can produce high-quality solutions, but most of these solutions are dominated by solutions with a small $|R|$. As a result, the NSGA-II with fixed $|R|$ needs to be executed many times with different values of $|R|$ to produce the same approximate Pareto front generated by the NSGA-II with variable $|R|$. However, when the network grows in size, the evaluation of one solution will take a great deal of time, and there are more values of $|R|$ to be searched. As a result, a decision maker is forced to compromise on solution quality and computation resource. Note that all the results are produced by a single run of the NSGA-II.

Table 7 further presents the computation time for Mumford's networks. The computation times with variable $|R|$ are generally larger than the ones with fixed $|R|$ because NSGA-II with variable $|R|$ has to search solution space with larger $|R|$ values than the recommended value. To produce the same Pareto front, NSGA-II with fixed $|R|$ has to be run many times with different input values for $|R|$. The total computation time scales with the range of $|R|$. In practice, the NSGA-II with variable $|R|$ is easier to implement.

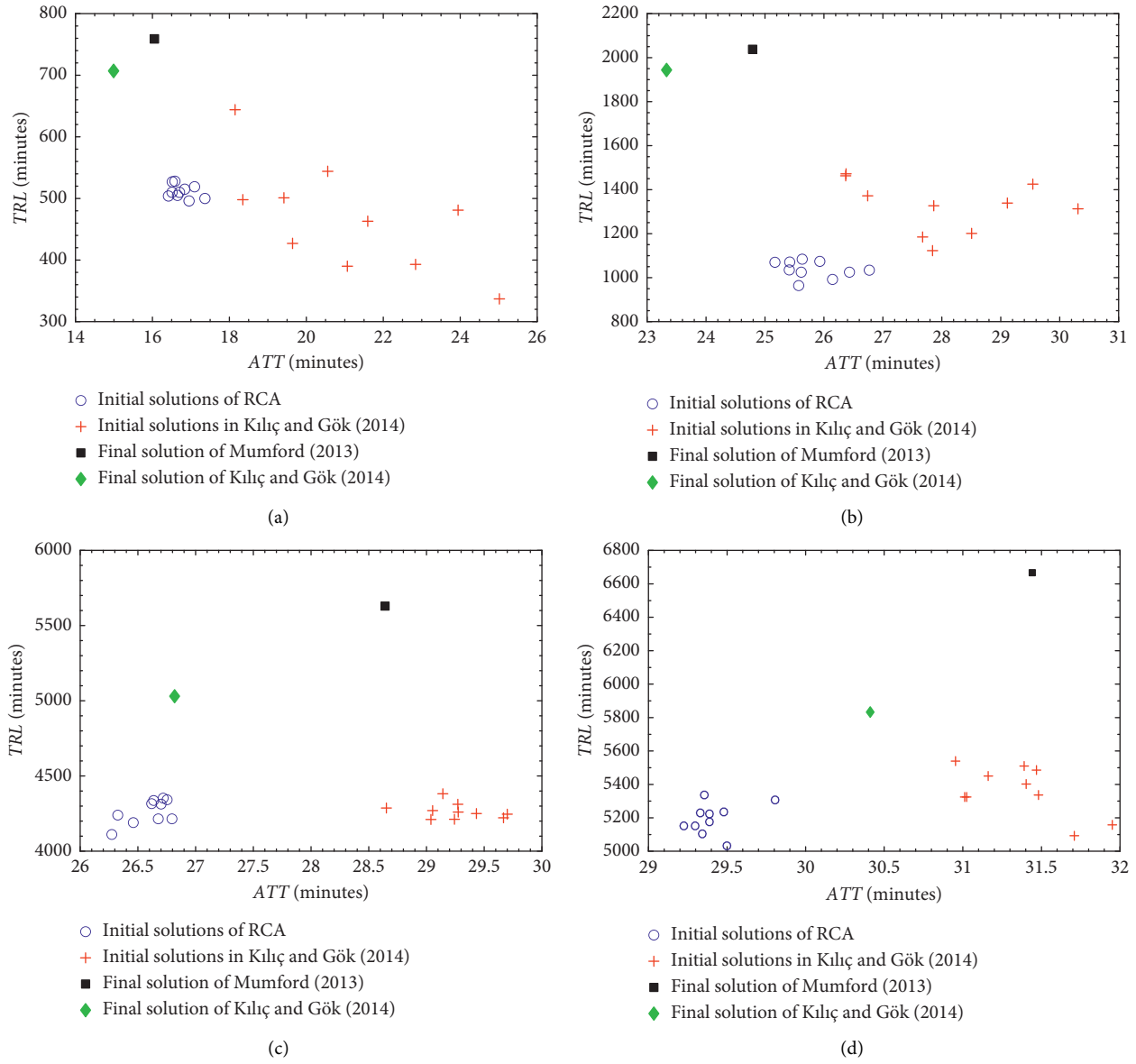


Figure 7: Comparisons of the initial solutions for Mumford's networks. (a) Mumford0 ($|R| = 12$). (b) Mumford1 ($|R| = 15$). (c) Mumford2 ($|R| = 56$). (d) Mumford3 ($|R| = 60$).

Table 4: Initial solutions for Mumford's networks compared to Kilic and Gok [14].

Parameters	Mumford0		Mumford1		Mumford2		Mumford3	
	Kilic and Gok [14]	RCA	Kilic and Gok [14]	RCA	Kilic and Gok [14]	RCA	Kilic and Gok [14]	RCA
d_0 (%)	45.8	57.72	30.68	35.1	27.33	39.59	25.18	38.28
d_1 (%)	49.12	37.97	47.48	53.83	56.18	57.11	51.23	56.14
d_2 (%)	5.08	4.31	19.43	10.23	15.76	3.29	20.51	5.58
d_{un} (%)	0	0	2.41	0.84	0.73	0	3.08	0.01
ATT	18.23	16.73	27.65	25.41	28.27	26.44	31.07	29.24
TRL	535	498	1363	1035	4482	4210	5668	5151

5.3. Comparison with Existing Published Results. To further demonstrate the feasibility and effectiveness of the proposed algorithms, we systematically compare final solutions with existing published results for Mandl's network and Mumford's large networks.

5.3.1. Mandl's Network. In Figure 10, we compare the final solutions for Mandl's network with the previously published results in Fan and Mumford [41], Mumford [27], Chew et al. [32], Nikolic and Teodorovic [16], Kilic and Gok [40], Kilic and Gok [14], Kilic [42], John et al. [43], Kechagiopoulos and

Table 5: Comparison of computation time of initialization procedures.

Network	Mandl-8	Mumford0	Mumford1	Mumford2	Mumford3
Kilic and Gok [14] (minutes)	0.01	0.69	30.82	253.48	484.83
RCA (seconds)	0.065	0.03	0.48	2.2	3.07

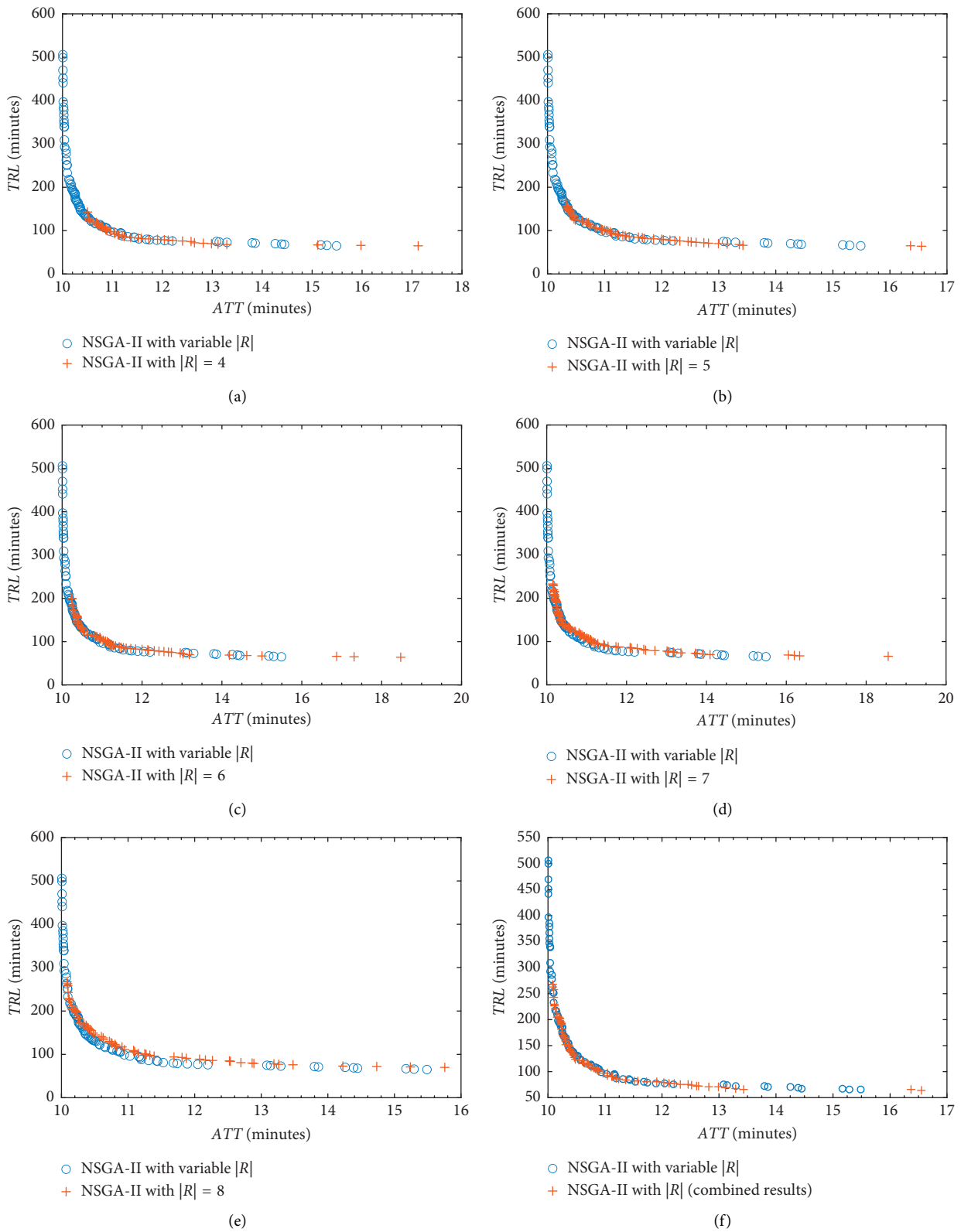


Figure 8: Approximate Pareto fronts generated by the NSGA-II with variable and fixed $|R|$ for Mandl's network. (a) Fixed $|R| = 4$. (b) Fixed $|R| = 5$. (c) Fixed $|R| = 6$. (d) Fixed $|R| = 7$. (e) Fixed $|R| = 8$. (f) Fixed $|R| = 4 - 8$.

Table 6: Computation time of the optimization procedure.

Computation time	Average	Minimum	Maximum	Std. dev.
NSGA-II with variable $\|R\|$	3244.9	1626	4500	834.67
NSGA-II with $\|R\| = 4$	1777.2	764	2910	745.49
NSGA-II with $\|R\| = 5$	3317.8	1617	5845	1407.82
NSGA-II with $\|R\| = 6$	3762.2	2141	5818	1386.76
NSGA-II with $\|R\| = 7$	2901	1690	4255	941.41
NSGA-II with $\|R\| = 8$	3113.5	1996	5696	1246.36

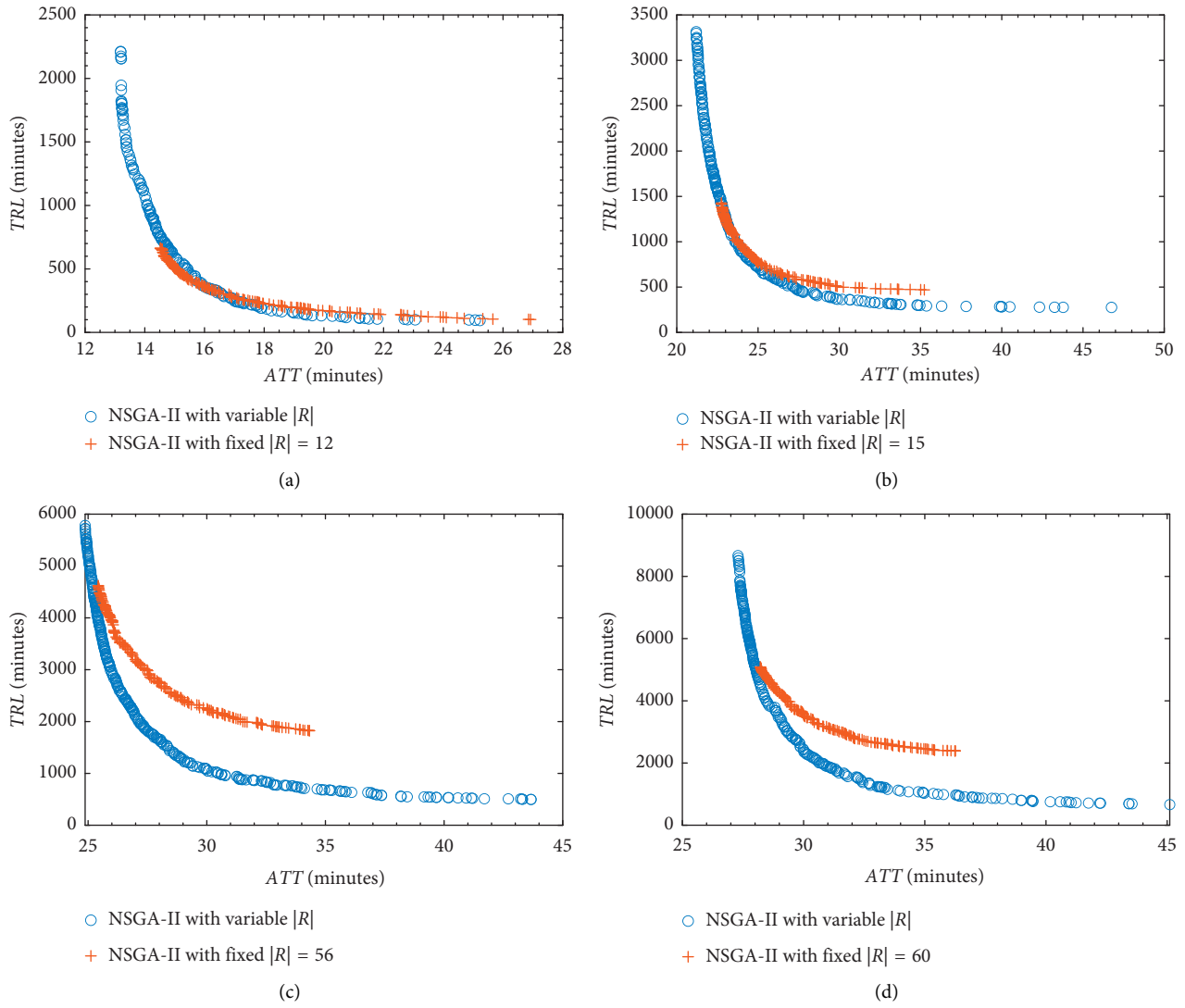


Figure 9: Approximate Pareto fronts generated by the NSGA-II with variable and fixed $|R|$ for Mumford's networks. (a) Mumford0. (b) Mumford1. (c) Mumford2. (d) Mumford3.

Table 7: Computation time of optimization procedure for Mumford's networks.

Network	Mumford0	Mumford1	Mumford2	Mumford3
NSGA-II with fixed $\|R\|$ (h)	0.262	1.8418	18.6342	82.0615
NSGA-II with variable $\|R\|$ (h)	1.782	2.2411	67.3288	179.038

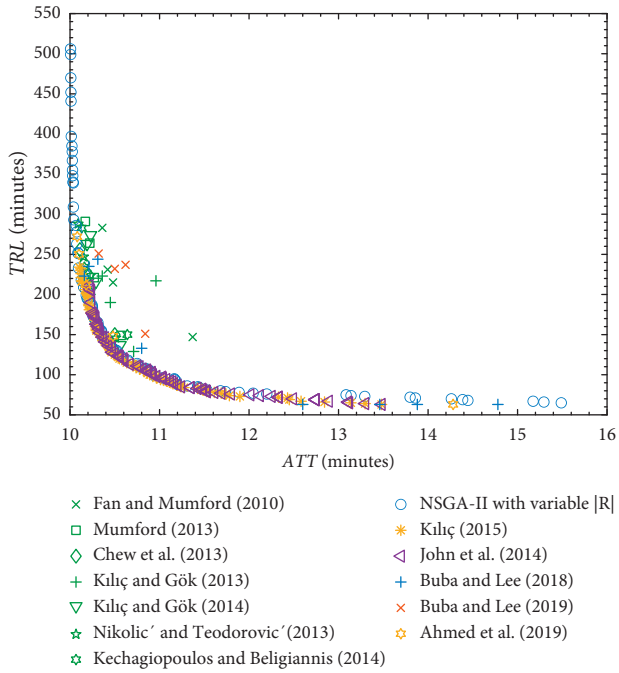


Figure 10: Comparison with previously published results for Mandl's network.

Beligiannis [13], Ahmed et al. [38], Buba and Lee [36], and Buba and Lee [37]. The approximate Pareto fronts of Kilic [42] and John et al. [43] plotted in Figure 10 are obtained via private communication.

From Figure 10, it can be seen that our results determined by varying the value of $|R|$ are competitive with other results. In general, our results are quite similar to the combined approximate Pareto front in Kilic [42] and the results in John et al. [43]. The ATT of our best 8-route solution is 10.09, which is almost the same as the ATT values in Nikolic and Teodorovic [16] (i.e., 10.09) and Ahmed et al. [38] (i.e., 10.08). However, their TRL values are, respectively, 288 and 272 minutes, which are much larger than ours (i.e., 233 minutes). Tables 8 and 9 separately compare the detailed performance indicators of the solutions in Figure 10 from the perspectives of passenger and operator.

In Table 8, we compare our results with former research studies in terms of d_0, d_1, d_2, d_{un}, ATT, and TRL. The presented solutions of our work are selected from the obtained approximate Pareto front generated by NSGA-II with variable $|R|$. The TRL values of Nikolic and Teodorovic [16] and Kechagiopoulos and Beligiannis [13] are obtained through our evaluation functions since they did not report these values.

From Table 8, the proposed solution method outperforms most of the former results. The ATT values of our method are very close to the smallest ATT values in four cases. For 4-route solutions, the best ATT (obtained by Ahmed et al. [38]) is 10.48 which is slightly smaller than ours (i.e., 10.56). However, our solution does have a much smaller TRL value (i.e., 124) when compared with the result (i.e., 148) in Ahmed et al. [38]. For 6-route solutions, the best ATT (obtained by Buba and Lee [36]) is 10.16, which is slightly smaller than ours (i.e., 10.19). The difference is less than 1%. On the contrary, the TRL of our solution is 195 (i.e., 12% less than 223). From the passenger's perspective, our solution method is competitive with other approaches in the literature.

Among 12 existing research studies mentioned above, only Mumford [27], Chew et al. [32], John et al. [43], Kilic and Gok [14], Buba and Lee [36], and Ahmed et al. [38] reported best solutions from the operator's perspective. Hence, to enhance the compatibility, we compare our results with existing literature [14, 27, 32, 36, 38, 43] in Table 9. Recall that the $|R|$ value is not predefined. As a result, there is only one best solution from the operator's perspective (in the last column in Table 9). So, we supplement the best solutions from NSGA-II with fixed $|R|$ in Figure 8.

The results of the proposed NSGA-II with fixed $|R|$ are less impressive than other literature. Our operator's costs are close to the lower bound of TRL. This is because the objective of our search algorithm is to search for the approximate Pareto front, instead of optimizing the passenger and/or operator cost. Generally, the best TRL values of our results are slightly larger than the lower bound of TRL (i.e., 12%). The best result for operator obtained from NSGA-II with variable $|R|$ contains only 3 routes.

5.3.2. Mumford's Networks. Figure 11 compares the obtained approximate Pareto front with recently published results for Mumford's larger networks in Mumford [27], Kilic and Gok [14], Kilic [42], John et al. [43], Buba and Lee [36], Nayeem et al. [15], and Ahmed et al. [38]. The approximate fronts obtained by John et al. [43] and Kilic [42] are obtained via private communication. Note that the results of Nayeem et al. [15] are obtained by evaluating the reported final solution in their work (the only feasible route set is for network Mumford1).

From Figure 11, it can be seen that for network Mumford0, the obtained front has a large duplication with former results. When network size grows, the gap between the obtained front and former results gets larger. This implies that our method has an advantage over large networks. The approximate Pareto fronts acquired by our method cover a larger area than other results because the $|R|$ value is not predefined. Tables 10 and 11 summarize the detailed performance indicators of the best solutions from the perspectives of both the passenger and the operator in Figure 11.

In Table 10, we compare our results from the passenger's perspective with other methods for Mumford's four larger networks. Note that the results of Nayeem et al. [15] are obtained by evaluating the reported final solution in their work (the only feasible route set is for network Mumford1).

From Table 10, it can be seen that our proposed method with variable $|R|$ outperforms the methods in Mumford [27], Nayeem et al. [15], Kilic and Gok [14], John et al. [43], Kilic [42], Buba and Lee [36], and Ahmed et al. [38] in terms of the ATT. The proposed solution method also found the best d_0 in all instances except network Mumford2. The percentages of demand satisfied with at most one transfer ($d_0 + d_1$) are

Table 8: Passenger perspective results compared to other approaches for Mandl's networks.

$\|R\|$	Indicators	Fan and Mumford [41]	Mumford [27]	Chew et al. [32]	Kılıç and Gök [40]	Kılıç and Gök [14]	Nikolić and Teodorović [16]	Kechagiopoulos and Beligiannis [13]	John et al. [43]	Kılıç [42]	Ahmed and Lee [36]	Ahmed and Lee [37]	Ahmed et al. [38]	NSGA-II with variable $\|R\|$
4	d_0 (%)	93.26	90.43	91.84	91.84	91.33	92.1	91.84	—	91.2	91.46	92.23	91.84	91.07
	d_1 (%)	6.74	9.57	8.16	8.16	8.16	7.19	7.64	—	8.8	8.54	7.71	8.15	8.22
	d_2 (%)	0	0	0	0	0.51	0.71	0.51	—	0	0	0.06	0	0.71
	d_{un} (%)	0	0	0	0	0	0	0	—	0	0	0	0	0
	ATT	11.37	10.57	10.5	10.71	10.56	10.51	10.64	—	10.5	10.8	10.84	10.48	10.56
	TRL	147	149	150	129	137	146a	150b	—	140	133	151	148	124
6	d_0 (%)	91.52	95.38	96.79	92.81	95.5	95.63	96.21	—	96.4	97.24	94.93	97.17	96.92
	d_1 (%)	8.48	4.56	3.21	6.87	4.5	4.37	3.47	—	3.6	2.76	5.07	2.82	3.08
	d_2 (%)	0	0.06	0	0.32	0	0	0.32	—	0	0	0	0	0
	d_{un} (%)	0	0	0	0	0	0	0	—	0	0	0	0	0
	ATT	10.48	10.27	10.21	10.45	10.29	10.23	10.23	10.22	10.25	10.16	10.5	10.18	10.19
	TRL	215	221	224	190	216	224a	203b	211	180	223	232	212	195
7	d_0 (%)	93.32	96.47	98.01	94.35	97.04	98.52	97.17	—	96.66	97.37	91.84	98.84	97.43
	d_1 (%)	6.36	3.34	1.99	5.14	2.83	1.48	2.83	—	3.34	2.63	8.16	1.15	2.57
	d_2 (%)	0.32	0.19	0	0.51	0.13	0	0	—	0	0	0	0	0
	d_{un} (%)	0	0	0	0	0	0	0	—	0	0	0	0	0
	ATT	10.42	10.22	10.16	10.36	10.23	10.15	10.16	—	10.2	10.21	10.62	10.1	10.15
	TRL	231	264	239	223	274	247a	234b	—	229	235	237	250	209
8	d_0 (%)	94.54	97.56	99.04	95.12	98.46	98.97	97.75	—	98.52	98.2	94.54	99.16	98.59
	d_1 (%)	5.46	2.31	0.96	4.56	1.54	1.03	2.25	—	1.48	1.8	5.46	0.83	1.41
	d_2 (%)	0	0.13	0	0.32	0	0	0	—	0	0	0	0	0
	d_{un} (%)	0	0	0	0	0	0	0	—	0	0	0	0	0
	ATT	10.36	10.17	10.11	10.96	10.19	10.09	10.13	—	10.1	10.31	10.32	10.08	10.09
	TRL	283	291	256	217	263	288a	283b	—	234	244	251	272	233

Table 9: Operator perspective results compared to other approaches for Mandl's network.

$\|R\|$	Indicators	Mumford [27]	Chew et al. [32]	John et al. [43]	Kılıç [42]	Ahmed and Lee [36]	Ahmed et al. [38]	NSGA-II with fixed $\|R\|$	NSGA-II with variable $\|R\|$
4	d_0 (%)	61.08	61.08	—	48.81	61.08	—	56.65	47.40
	d_1 (%)	36.61	36.61	—	45.54	36.61	—	41.55	46.05
	d_2 (%)	2.31	2.31	—	5.33	2.31	—	1.80	6.55
	d_{un} (%)	0	0	—	0.32	0	—	0	0.00
	ATT	13.88	13.88	—	14.67	13.88	—	13.29	15.49
	TRL	63	63	—	63	63	—	68	65.00
6	d_0 (%)	70.91	70.91	—	70.91	70.46	62.23	63.84	$\|R\| = 3$
	d_1 (%)	25.5	25.5	—	25.5	24.34	27.16	27.62	
	d_2 (%)	2.95	2.95	—	2.95	5.2	9.57	8.09	
	d_{un} (%)	0.64	0.64	—	0.64	0	1.03	0.45	
	ATT	13.48	13.48	13.48	13.48	12.6	14.28	14.19	
	TRL	63	63	63	63	63	63	69	
7	d_0 (%)	65.13	70.65	—	70.65	68.96	—	57.55	
	d_1 (%)	22.93	21.13	—	21.13	26.29	—	33.27	
	d_2 (%)	10.34	7.13	—	7.13	3.92	—	8.73	
	d_{un} (%)	1.61	1.09	—	1.09	0.83	—	0.45	
	ATT	14.25	13.76	—	13.76	13.46	—	13.92	
	TRL	63	63	—	63	63	—	71	
8	d_0 (%)	57.93	61.91	—	43.93	60.76	—	36.48	
	d_1 (%)	31.92	29.67	—	21.52	25.63	—	44.32	
	d_2 (%)	9.7	6.87	—	14.96	10.34	—	11.11	
	d_{un} (%)	0.45	1.54	—	19.59	3.26	—	8.09	
	ATT	14.45	14.22	—	17.95	14.78	—	15.76	
	TRL	63	63	—	63	63	—	70	

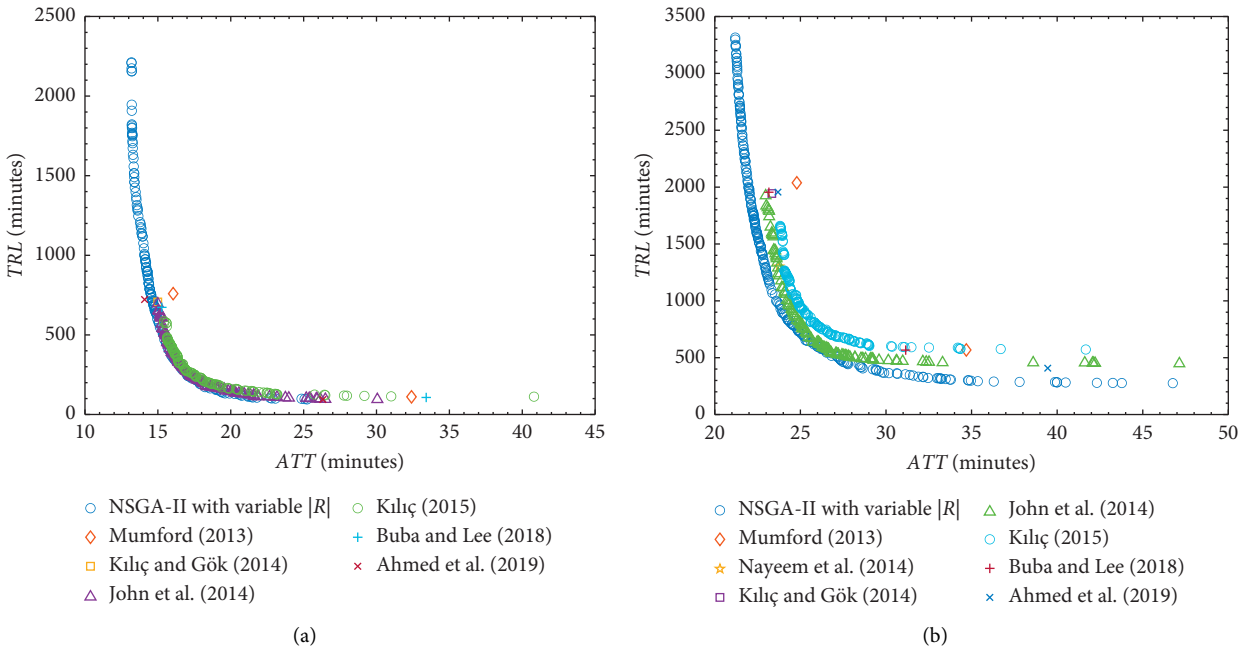


Figure 11: Continued.

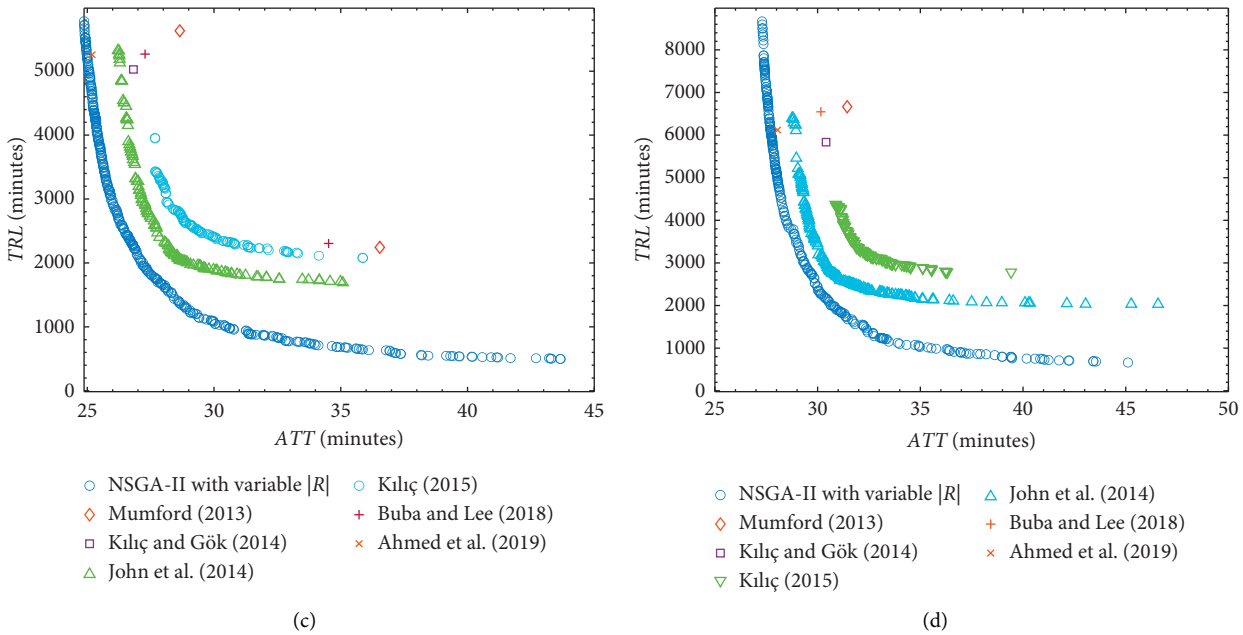


Figure 11: Comparison with previously published results for Mumford's networks. (a) Mumford0. (b) Mumford1. (c) Mumford2. (d) Mumford3.

Table 10: Passenger perspective results compared to other approaches for Mumford's networks.

Network	Parameters	Mumford [27]	Nayeem et al. [15]	Kılıç and Gök [14]	John et al. [43]	Kılıç [42]	Ahmed and Lee [36]	Ahmed et al. [38]	NSGA-II with variable $\|R\|$
Mumford0	d_0 (%)	63.2	—	69.73	—	66.88	65.41	88.74	98.77
	d_1 (%)	35.82	—	30.03	—	32.76	34.24	11.25	1.23
	d_2 (%)	0.98	—	0.24	—	0.37	0.35	0	0
	d_{un} (%)	0	—	0	—	0	0	0	0
	ATT	16.05	—	14.99	14.99	15.43	15.27	14.09	13.19
	TRL	759	—	707	696	587	673	722	2212
Mumford1	d_0 (%)	36.6	38.77	45.1	—	40.8	38.77	65.75	71.04
	d_1 (%)	52.42	55.45	49.08	—	52.46	54.23	34.18	28.96
	d_2 (%)	10.71	5.79	5.76	—	6.74	5.12	0.07	0
	d_{un} (%)	0.26	0	0.06	—	0	1.88	0	0
	ATT	24.79	23.75	23.33	22.97	23.82	23.16	21.69	21.19
	TRL	2038	1301	1944	1925	1659	1955	1956	3315
Mumford2	d_0 (%)	30.92	—	33.88	—	29.14	31.47	56.68	53.74
	d_1 (%)	51.29	—	57.18	—	60.21	58.23	43.26	46.24
	d_2 (%)	16.36	—	8.77	—	10.48	9.6	0.05	0.02
	d_{un} (%)	1.44	—	0.17	—	0.17	0.7	0	0
	ATT	28.65	—	26.82	26.21	27.67	27.28	25.19	24.88
	TRL	5632	—	5027	5326	3957	5268	5257	5783
Mumford3	d_0 (%)	27.46	—	27.56	—	26.17	28.12	50.41	56.32
	d_1 (%)	50.97	—	53.25	—	56.78	54.35	48.81	43.48
	d_2 (%)	18.76	—	17.51	—	16.37	16.84	0.77	0.2
	d_{un} (%)	2.81	—	1.68	—	0.68	0.69	0	0
	ATT	31.44	—	30.41	28.76	30.87	30.16	28.05	27.27
	TRL	6665	—	5834	6400	4373	6547	6119	8668

over 99% for all instances. Although the *TRL* values of our approach are generally larger than other approaches, a solution that dominates the results of existing methods in terms of both objectives can be found in Figure 11.

Besides, Table 11 compares our results from the operator's perspective for Mumford's four larger networks with recently published results in Mumford [27], John et al. [43], Kilic [42], Buba and Lee [36], and Ahmed et al. [38]. Our approach found the best results in networks Mumford1, Mumford2, and Mumford3 in terms of *TRL*. This is not unexpected, since the number of routes in our approach can be alternated by the route set size alternating heuristic as

Table 11: Operator perspective results compared to other approaches for Mumford's networks.

Network	Parameters	Mumford [27]	John et al. [43]	Kılıç [42]	Ahmed and Lee [36]	Ahmed et al. [38]	NSGA-II with variable $\|R\|$
Mumford0	D_0 (%)	18.42	—	25.51	16.99	14.61	39.96
	D_1 (%)	23.4	—	16.55	30.72	31.59	46.8
	d_2 (%)	20.78	—	21.38	28.92	36.41	13.23
	d_{un} (%)	37.4	—	36.55	23.38	17.37	0
	ATT	32.4	30.06	40.8	33.41	26.32	25.23
	TRL	111	95	112	107	94	95
Mumford1	d_0 (%)	16.35	—	19.7	22.39	18.02	29.08
	d_1 (%)	29.06	—	32.29	40.57	29.88	42.01
	d_2 (%)	29.93	—	30.01	34.33	31.9	23.57
	d_{un} (%)	24.66	—	18	2.71	20.19	5.34
	ATT	34.69	47.14	41.67	31.15	39.45	46.75
	TRL	568	448	572	567	408	276
Mumford2	d_0 (%)	13.76	—	15.62	15.32	13.63	16.06
	d_1 (%)	27.69	—	38.07	29.31	23.58	41.25
	d_2 (%)	29.53	—	33.7	31.08	23.94	32.39
	d_{un} (%)	29.09	—	12.62	24.29	38.82	10.29
	ATT	36.54	35.08	35.87	34.52	46.86	43.67
	TRL	2244	1691	2078	2305	1330	497
Mumford3	d_0 (%)	16.71	—	14.24	25.42	16.28	17.44
	d_1 (%)	33.69	—	33.44	41.26	24.87	44.22
	d_2 (%)	29.18	—	34.1	24.91	26.34	31.19
	d_{un} (%)	20.42	—	18.22	8.41	32.44	7.16
	ATT	36.92	46.58	39.42	34.81	46.05	45.11
	TRL	2830	2030	2782	2732	1746	663

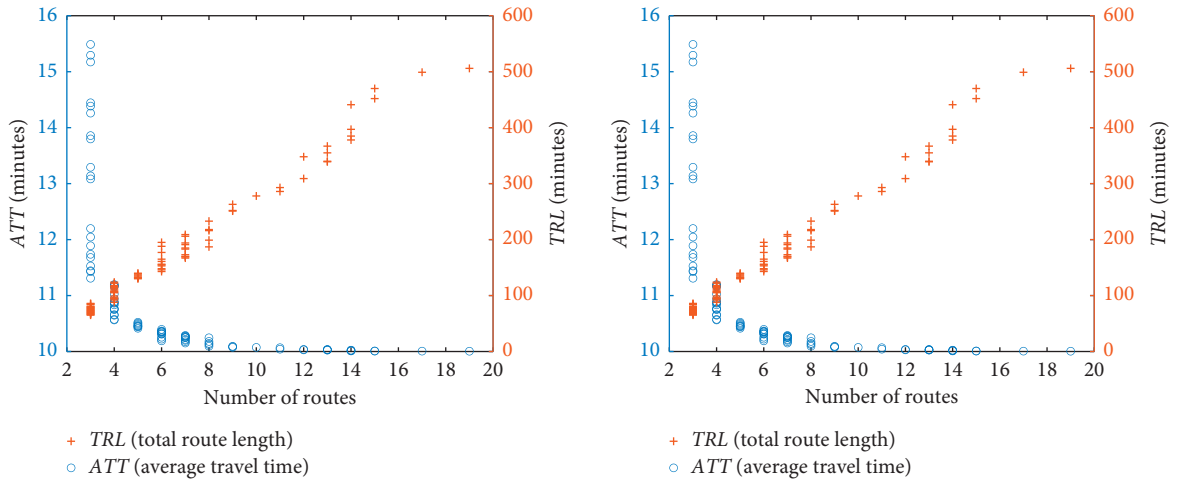


Figure 12: Continued.

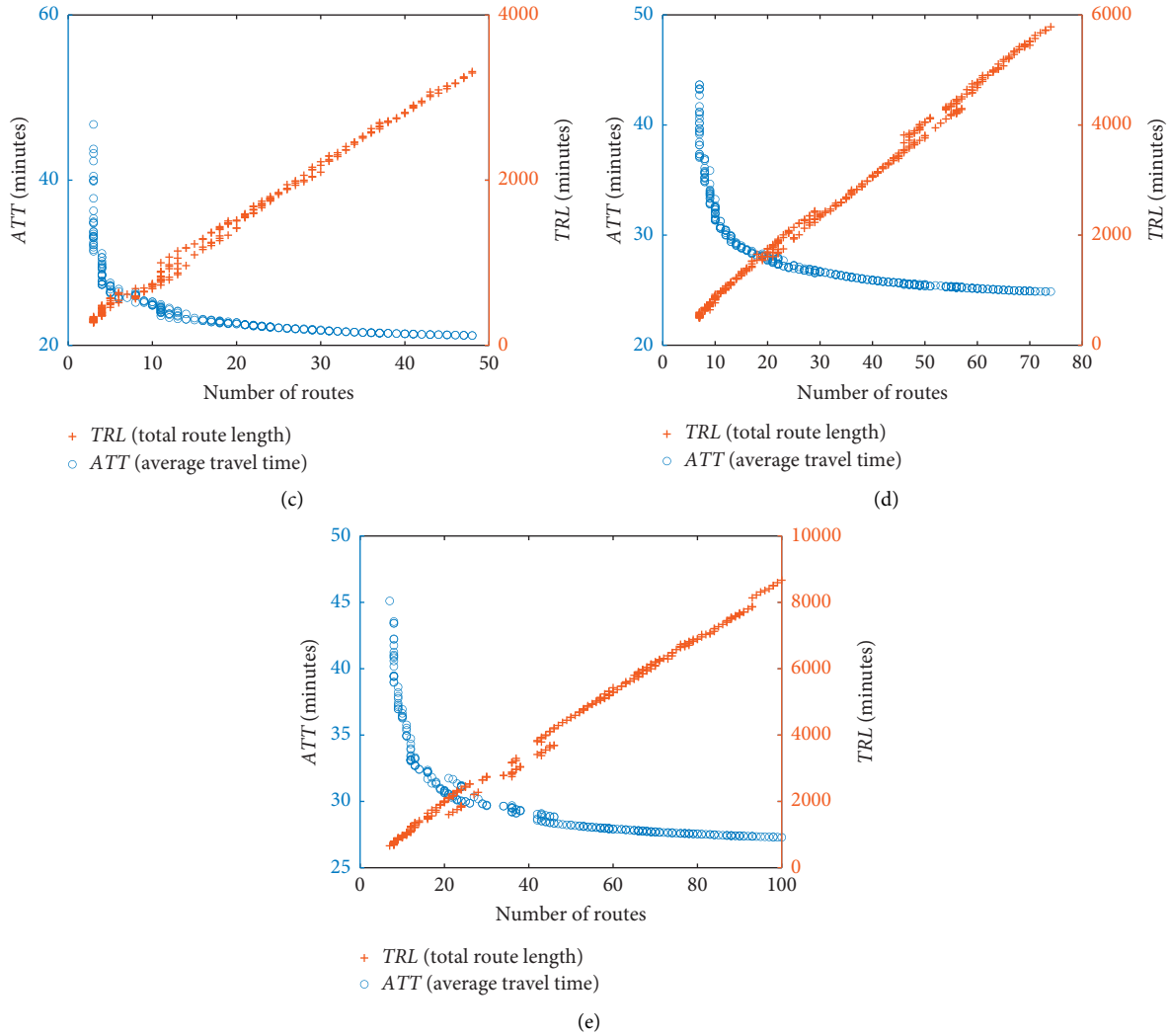


Figure 12: Effect of $|R|$ on both objective functions in five networks. (a) Mandl's network. (b) Mumford0. (c) Mumford1. (d) Mumford2. (e) Mumford3.

long as the solution is feasible. The best solution for network Mumford0 is obtained by Ahmed et al. [38]. The *TRL* of their result is one minute less than ours (i.e., 95 minutes). On the contrary, the *ATT* of our result is smaller than theirs, and all demands can be satisfied with our route set. Generally, the proposed solution method is competitive with existing approaches in terms of both objectives for larger instances.

5.4. Recommended Value of $|R|$. In the TNDP, it is difficult to state a suitable number of routes for the best solution [44]. By embedding a route set generation heuristic in a bi-objective optimization framework, we try to find the approximate Pareto front without presetting $|R|$. The aim of this section is to demonstrate the efect of $|R|$ on both objective values in the approximate Pareto fronts.

Figure 12 shows the $|R|$ values of the approximate Pareto front solutions for five networks. As expected, both objectives are generally conflicting. With $|R|$ increasing, the *ATT* decreases and the *TRL* increases at the same time. When $|R|$ is reaching the theoretical largest value, the *ATT* will converge to the theoretical lowest value (shown in Table 2). When $|R|$ is larger than the theoretical largest value, the increase of $|R|$ will only lead to larger *TRL* but no decrease in the *ATT*. The figures reveal that when $|R|$ reaches a certain value, a further increase in the $|R|$ value can barely lead to decreasing of *ATT*. For example, for network Mumford3, the *ATT* decreases from 27.54 minutes to 27.29 minutes when $|R|$ increases from 80 to 100, while the *TRL* increases from 6894 to 8688. This demonstrates that a large $|R|$ is not always beneficial to both the passenger and the operator.

From Figure 12, it can be seen that the number of solutions with specific $|R|$ in the approximate Pareto fronts is fairly small for four larger networks. This indicates that the best solutions with predefined $|R|$ are only a small part of the approximate Pareto front. The proposed approach with variable $|R|$ is more practical and efficient for large networks as the search process for large networks is computationally time consuming.

Table 12: Recommended $|R|$ values.

Network	Mandl	Mumford0	Mumford1	Mumford2	Mumford3
Recommended $|R|$ range	[3, 12]	[8, 30]	[4, 25]	[8, 42]	[10, 47]

Based on the observation of Figure 12, we present a recommended range of the $|R|$ value for each instance in Table 12. It should be noted that the recommended values are only meant for future researchers to limit the number of routes to avoid wasting computation resources on solutions with extreme trade-off level.

6. Conclusions

An integrated solution method is proposed to simultaneously solve the transit network design problem and the route set size determination problem. The solution method includes a novel initial route set generation algorithm that aims to produce high-quality initial solutions for the optimization algorithm and to explore the solution space with a different number of routes, and it also contains a route set size alternating heuristic to alternate the number of routes in a solution. These two algorithms are embedded into NSGA-II-based solution framework to find the approximate Pareto front in terms of both objectives (i.e., the average travel time and total route length).

The proposed initial route set generation algorithm is based on the intuitive observation that maximizing the travel demand that can be satisfied directly (without transfers) is a key component in the generation of an initial route set. The initial solutions produced indicate that the proposed initial route set generation algorithm is capable of producing high-quality initial solutions for large networks. Note that the initial solutions obtained for networks Mumford2 and Mumford3 are dominating final solutions generated by some former research studies in terms of both objectives. The comparison of computation time also shows that the proposed algorithm is very efficient, as it only takes three seconds to generate an initial solution for the largest network (i.e., Mumford3).

Based on the analysis of the allocation of each travel demand, to improve the local optimal solution and explore a solution space with a different $|R|$ value, a route set size alternating heuristic is developed to switch the number of routes in a solution. The experimental results show that the route set size alternating heuristic can save considerable computation time without deteriorating the quality of the approximate Pareto front obtained. When compared with the fixed route set size approach, the approximate Pareto fronts obtained by varying the number of routes in a solution can lead to a wider range of trade-offs, which enables a decision maker to focus on the trade-offs within this constrained set of approximate Pareto front solutions, rather than needing to concentrate on a fixed number of routes before the generation of a transit network. For future work, the proposed solution method should be extended to include bus headway determination and fleet size constraint, and we will extend our research by extracting dynamic transit demand from available sources.

Data Availability

The transit network instances used to support the findings of this study have been deposited in http://users.cs.cf.ac.uk/C.L.Mumford/Research%20Topics/UTRP/Outline.html.

Disclosure

The authors are responsible for any error in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 71901183), the Open Research Fund for National Engineering Laboratory of Integrated Transportation Big Data Application Technology (nos. CTBDAT201901 and CTBDAT201907), and the Fundamental Research Funds for the Central Universities.

Supplementary Materials

This section includes the approximate Pareto fronts obtained by the proposed solution method. For each instance, there are two kinds of files: Excel and Txt. Excel file contains the performance indicators of the solutions in the obtained approximate Pareto front. Txt files contain the corresponding route structure. (*Supplementary Materials*)

References

[1] A. Ceder and N. H. M. Wilson, "Bus network design," *Transportation Research Part B Methodological*, vol. 20, no. 4, pp. 331–344, 1986.

[2] S. B. Pattnaik, S. Mohan, and V. M. Tom, "Urban bus transit route network design using genetic algorithm," *Journal of Transportation Engineering*, vol. 124, no. 4, pp. 368–375, 1998.

[3] V. M. Tom and S. Mohan, "Transit route network design using frequency coded genetic algorithm," *Journal of Transportation Engineering*, vol. 129, no. 2, pp. 186–195, 2003.

[4] S. Afandizadeh, H. Khaksar, and N. Kalantari, "Bus fleet optimization using genetic algorithm a case study of mashhad," *International Journal of Civil Engineering*, vol. 11, no. 1, pp. 43–52, 2013.

[5] M. H. Baaj and H. S. Mahmassani, "An ai-based approach for transit route system planning and design," *Journal of Advanced Transportation*, vol. 25, no. 2, pp. 187–209, 1991.

[6] E. Cipriani, S. Gori, and M. Petrelli, "Transit network design: a procedure and an application to a large urban area,"

Transportation Research Part C: Emerging Technologies, vol. 20, no. 1, pp. 3–14, 2012.

[7] W. Fan and R. B. Machemehl, "Optimal transit route network design problem with variable transit demand: genetic algorithm approach," *Journal of Transportation Engineering*, vol. 132, no. 1, pp. 40–51, 2006a.

[8] S. Ngamchai and D. J. Lovell, "Optimal time transfer in bus transit route network design using a genetic algorithm," *Journal of Transportation Engineering*, vol. 129, no. 5, pp. 510–521, 2003.

[9] W. Y. Szeto and Y. Wu, "A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong," *European Journal of Operational Research*, vol. 209, no. 2, pp. 141–155, 2011.

[10] S. A. Bagloee and A. Ceder, "Transit-network design methodology for actual-size road networks," *Transportation Research Part B-Methodological*, vol. 45, no. 10, pp. 1787–1804, 2011.

[11] P. Chakroborty, "Genetic algorithms for optimal urban transit network design," *Computer-aided Civil and Infrastructure Engineering*, vol. 18, no. 3, pp. 184–200, 2003.

[12] P. Chakroborty and T. Wivedi, "Optimal route network design for transit systems using genetic algorithms," *Engineering Optimization*, vol. 34, no. 1, pp. 83–100, 2002.

[13] P. N. Kechagiopoulos and G. N. Beligiannis, "Solving the urban transit routing problem using a particle swarm optimization based algorithm," *Applied Soft Computing*, vol. 21, pp. 654–676, 2014.

[14] F. Kilic and M. Gok, "A demand based route generation algorithm for public transit network design," *Computers & Operations Research*, vol. 51, pp. 21–29, 2014.

[15] M. A. Nayeem, M. K. Rahman, and M. S. Rahman, "Transit network design by genetic algorithm with elitism," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 30–45, 2014.

[16] M. Nikolic and D. Teodorovic, "Transit network design by bee colony optimization," *Expert Systems with Applications*, vol. 40, no. 15, pp. 5945–5955, 2013.

[17] Q. K. Wan and H. K. Lo, "A mixed integer formulation for multiple-route transit network design," *Journal of Mathematical Modelling and Algorithms*, vol. 2, no. 4, pp. 299–308, 2003.

[18] F. Zhao and X. Zeng, "Optimization of transit route network, vehicle headways and timetables for large-scale transit networks," *European Journal of Operational Research*, vol. 186, no. 2, pp. 841–855, 2008.

[19] V. Guihaire and J.-K. Hao, "Transit network design and scheduling: a global review," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1251–1273, 2008.

[20] K. Kepaptsoglou and M. Karlaftis, "Transit route network design problem: Review," *Journal of Transportation Engineering*, vol. 135, no. 8, pp. 491–505, 2009.

[21] R. Z. Farahani, E. Miandoabchi, W. Y. Szeto, and H. Rashidi, "A review of urban transportation network design problems," *European Journal of Operational Research*, vol. 229, no. 2, pp. 281–302, 2013.

[22] Z. U. Abedin, F. Busch, D. Z. W. Wang, A. Rau, and B. Du, "Comparison of public transport network design methodologies using solution-quality evaluation," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 8, 2018.

[23] T. Liu and A. Ceder, "Analysis of a new public-transport-service concept: customized bus in China," *Transport Policy*, vol. 39, pp. 63–76, 2015.

[24] W. Lampkin and P. D. Saalmans, "The design of routes, service frequencies, and schedules for a municipal bus undertaking: a case study," *Journal of the Operational Research Society*, vol. 18, no. 4, pp. 375–397, 1967.

[25] C. E. Mandl, "Evaluation and optimization of urban public transportation networks," *European Journal of Operational Research*, vol. 5, no. 6, pp. 396–404, 1980.

[26] A. Mauttone and M. E. Urquhart, "A route set construction algorithm for the transit network design problem," *Computers & Operations Research*, vol. 36, no. 8, pp. 2440–2449, 2009.

[27] C. L. Mumford, "New heuristic and evolutionary operators for the multi-objective urban transit routing problem," in *Proceedings of the 2013 IEEE Congress on Evolutionary Computation*, pp. 939–946, Cancun, Mexico, June 2013.

[28] M. Owais and M. K. Osman, "Complete hierarchical multi-objective genetic algorithm for transit network design problem," *Expert Systems with Applications*, vol. 114, pp. 143–154, 2018.

[29] W. Fan and R. B. Machemehl, "Using a simulated annealing algorithm to solve the transit route network design problem," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 122–132, 2006.

[30] J. Y. Yen, "Finding the K Shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, 1971.

[31] R. O. Arbex and C. B. Cunha, "Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm," *Transportation Research Part B: Methodological*, vol. 81, pp. 355–376, 2015.

[32] J. S. C. Chew, L. S. Lee, and H. V. Seow, "Genetic algorithm for biobjective urban transit routing problem," *Journal of Applied Mathematics*, vol. 2013, Article ID 698645, 15 pages, 2013.

[33] F. Zhao, "Large-scale transit network optimization by minimizing user cost and transfers," *The Journal of Public Transportation*, vol. 9, no. 2, p. 6, 2006.

[34] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach," *Transportation Research Part B: Methodological*, vol. 67, pp. 235–263, 2014.

[35] H. Zhao, W. Xu, and R. Jiang, "The memetic algorithm for the optimization of urban transit network," *Expert Systems with Applications*, vol. 42, no. 7, pp. 3760–3773, 2015.

[36] A. T. Buba and L. S. Lee, "Differential evolution with improved sub-route reversal repair mechanism for multi-objective urban transit routing problem," *Numerical Algebra, Control and Optimization*, vol. 8, no. 3, p. 351, 2018.

[37] A. T. Buba and L. S. Lee, "Hybrid differential evolution-particle swarm optimization algorithm for multiobjective urban transit network design problem with homogeneous buses," *Mathematical Problems in Engineering*, vol. 2019, Article ID 5963240, 16 pages, 2019.

[38] L. Ahmed, C. Mumford, and A. Kheiri, "Solving urban transit route design problem using selection hyper-heuristics," *European Journal of Operational Research*, vol. 274, no. 2, pp. 545–559, 2019.

[39] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[40] F. Kiliç and M. Gök, "A public transit network route generation algorithm," *IFAC Proceedings Volumes*, vol. 46, no. 25, pp. 162–166, 2013.

[41] L. Fan and C. L. Mumford, "A metaheuristic approach to the urban transit routing problem," *Journal of Heuristics*, vol. 16, no. 3, pp. 353–372, 2010.

[42] F. Kilic, *Route generation algorithms for public transit network design*, Ph.D. thesis, Cukurova University, Adana, Turkey, 2015.

[43] M. P. John, C. L. Mumford, and R. Lewis, "An improved multi-objective algorithm for the urban transit routing problem," *Evolutionary Computation in Combinatorial Optimisation*, Springer, Berlin, Germany, pp. 49–60, 2014.

[44] H. Cancela, A. Mauttone, and M. E. Urquhart, "Mathematical programming formulations for transit network design," *Transportation Research Part B: Methodological*, vol. 77, pp. 17–37, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 4271871, 17 pages
https://doi.org/10.1155/2020/4271871

WILEY | Hindawi

Research Article

Estimating Wait Time and Passenger Load in a Saturated Metro Network: A Data-Driven Approach

Hezhou Qu,[1,2] Xiaoyue Xu,[1] and Steven Chien [3,4]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, Sichuan 611756, China*
[2]*National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu, Sichuan 611756, China*
[3]*College of Transportation Engineering, Chang'an University, Xi'an, Shaanxi 710064, China*
[4]*John A. Reif, Jr. Department of Civil and Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA*

Correspondence should be addressed to Steven Chien; chien@njit.edu

Received 6 February 2020; Revised 22 March 2020; Accepted 6 May 2020; Published 1 August 2020

Academic Editor: Yu Jiang

The service quality of public transit, such as comfort and convenience, is an important factor influencing ridership and fare revenue, which also reflects the passengers' perception to the transit performance. Passengers are frustrated while waiting to board a crowded train especially during the peak hours, while the fail-to-board (FtB) situation commonly exists. The service performance measures determined by deterministic passenger demand and service frequency cannot reflect the perceived service of passengers. With the automatic fare collection system data provided by Chengdu Metro, we develop a data-driven approach considering the joint probability of spatiotemporal passenger demand at stations based on posted train schedule to approximate passenger travel time (e.g., in-vehicle and out-of-vehicle times). It was found that the estimated wait time can reflect the actual situation as passengers FtB. The proposed modeling approach and analysis results would be useful and beneficial for transit providers to improve system performance and service planning.

1. Introduction

Urban rail rapid transit, also called metro, plays an important role to transport a vast number of passengers, especially in the peak period. In the past decade, metro networks have been expanding rapidly. Due to a dramatic increase of ridership, those networks are overcrowding over space and time, which has become a major problem affecting system efficiency as well as service quality.

The quality of service (QoS) for fixed-route transit consists of two categories: "Availability" and "Comfort and Convenience" [1]. Service availability reveals a minimum requirement for transit being a travel option, while comfort and convenience contribute to users' satisfaction to the service and their likelihood of using it. For a metro system, passenger load, service reliability, and travel time are such indices associated with comfort and convenience. Some passengers even weight more on crowdedness in the train than on travel time or distance [2–4]. In addition, the service measures were determined in previous studies using deterministic passenger demand and service frequency, which cannot reflect the perceived QoS of passengers.

From transit users' perspective, the passenger load affects the comfort during their travel, which varies over space and time in a network. The passenger load on a transit vehicle affects the comfort of the on-board vehicle portion of a transit trip in terms of both being able to find a seat and in overall crowding levels within the vehicle, represented by average space per standee and load factor. From transit operators' perspective on the other hand, the variation of passenger load may affect the system performance such that service frequency or number of cars per train need to justify for easing congestion and improving comfort [1]. It is desirable to know the spatiotemporal distributions of

passenger flows entering and exiting a metro system, so that appropriate service schedule as well as frequency can be determined [5–7]. A recent study [8] indicated that many transit users rely on real-time information (i.e., estimated travel time via smartphone) to make a travel choice. The accuracy of estimated travel time, consisting of in-vehicle and out-of-vehicle times, is critical for transit users to travel effectively.

The automatic fare collection system (AFCS) has been commonly applied (i.e., Metro Card in New York City, Oyster Card in London, etc.). Passenger flows entering and exiting metro stations can be analyzed with the AFCS transaction records, while wait time can be approximated in conjunction with the train schedule. However, transfer time is difficult to estimate because the itinerary of passengers (i.e., travel path) is unknown. In addition, overcrowding discouraged passengers boarding the train. This situation may result in fail-to-board (FtB), and extra wait/transfer time shall be expected. Previous research in estimating wait/transfer time under this context is still very limited.

The objective of this study is to develop a data-driven approach applied for approximating passenger itinerary, spatiotemporal passenger load (e.g., average space per standee and load factor), and travel time with the AFCS data considering station layout, train schedule, and the probability of FtB. Passenger travel time consists of in-vehicle and out-of-vehicle times (e.g., walking, wait, and transfer times).

The remainder of this paper is organized as follows. Section 2 describes previous research and practices related to this study. Sections 3 and 4 discuss the development of the proposed model and solution approach. Then, in Section 5, the numerical analysis based on the data provided by Chengdu Metro is conducted. Finally, the research findings and future studies are concluded.

2. Literature Review

Many studies have been conducted on analyzing transit system performance and service quality. Xin et al. [9] applied the methods suggested by Transit Capacity and Quality of Service Manual (TCQSM) to evaluate the QoS indicators (e.g., service frequency and coverage, hours of service, and transit-auto travel time) for Grand River Transit in Canada. Bunker [10] developed a method to profile route-based load factor as on-board passenger comfort and investigated the time series correlation between the load factor and passenger average travel time. Furth et al. [11] analyzed wait time and passenger load of a bus transit using the automatic vehicle location (AVL) and automatic passenger counter data to improve the performance and management of transit systems. For overcrowding situations, many passengers waiting on the platform find it difficult to board the immediate coming train due to limited train capacity and must wait for another available one. The TCQSM method [1] seems to be unable to accurately estimate wait time for a saturated metro network because the impact of FtB was not considered.

The AFCS data has been widely applied in public transit systems for OD demand estimation [12], timetable design [13], passenger flow assignment [14, 15], passenger behavior analysis [16], and transfer coordination [17–19]. Since passengers swipe the metro card only at the gates of origin and destination stations, detailed information, such as the location where a passenger makes a transfer, is unknown. Thus, wait/transfer time is very difficult to estimate, especially under overcrowding situations.

The spatiotemporal passenger load information is important for metro operation and service planning. Sun et al. [20] developed a regression model to estimate passenger demand of a transit line, using the minimum travel time between each pair of origin and destination stations. Zhang et al. [21] estimated walking time between the gate of a station and the platform and transfer time. With AFCS, AVL, and general transit feed specification data, Luo et al. [22] developed a method to generate passenger load profiles for transit vehicles, based on passenger check-in time and vehicle arrival time. However, none of the models discussed above considered FtB.

The FtB situation has become an increasingly important issue as the demand increases in metro systems. Zhu et al. [23] used the maximum likelihood method and a Bayesian inference method to estimate the probability of FtB with the AFCS data. Later, they [24] developed a probabilistic model for assigning passengers to trains and estimated the number of skipped trains because of overcrowding. With a bilevel regression model, Miller et al. [25] estimated the number of FtB passengers and calculated the shortage of cumulative capacity. However, the studies discussed above did not consider transfer passengers in a large metro network.

In addition to FtB, it is also important to consider transfer efficiency which would significantly affect metro system performance and level of service [26]. However, it is difficult to estimate transfer demand at stations because the AFCS data do not have passengers' itinerary information but have the records of them entering and exiting from stations. For a complicated network where passengers need to make multiple transfers to reach their destinations [27], estimating transfer demand becomes even more challenging. Passengers would select transfer locations to the best of their interests (i.e., less travel time or distance, greater reliability and comfort, etc.). Few studies investigated spatiotemporal travel-path choice under overcrowding situations. Kusakabe et al. [28] estimated the boarded trains by railway passengers whose travel-path choice was determined by the shortest travel path with the minimum wait, egress, and transfer times. Sun and Schonfeld [29] developed a schedule-based path-choice model for a rail transit network considering the probability of FtB. Zhao et al. [30] developed a probabilistic model to estimate route choices, assuming that the number of trains waited for by passengers at origin and transfer stations and the itineraries being chosen in a specific time obey the polynomial distribution. Both studies assumed that the numbers of FtB passengers at different stations are independent, which might contradict the reality because the train rode by a passenger at a transfer station is dependent on the train the passenger rode at the origin station.

While estimating wait time, previous studies considered idealized conditions, either ignoring FtB [31] or assuming random passenger arrivals [32]. Recently, more studies

applied AFCS data for estimating wait time. With a Bayesian inference model, Zhang and Yao [33] estimated walking, wait, transfer, and in-vehicle times as those follow truncated normal distributions. Considering the corelation of travel times between pairs of stations, Lee et al. [34] analyzed those travel time components with a decomposition method. However, transfer demand was not considered. Tavassoli et al. [35] estimated passenger wait time for nontransfer and transfer passengers, while assuming the travel path is given. However, the paths with more than one transfer were not considered. Considering passenger arrival distribution, Ingvardson et al. [36] assumed that wait time follows a joint uniform and beta distribution, but the effect of FtB to wait time was not assessed. In general, the studies discussed earlier did not consider the situations in which passengers may have multiple path choices with transfers to reach their destinations.

This study proposes a data-driven approach to approximate passenger itineraries and associated travel times (i.e., the sum of in-vehicle and out-of-vehicle times) for a saturated metro network with the AFCS data. The service indicators such as passenger load (i.e., average space per standee and load factor) and travel time components (e.g., walking, wait, and transfer times) can be estimated for monitoring system operation and improving service planning.

3. Methodology

In a metro system, passenger travel time consists of in-vehicle time and out-of-vehicle time. The in-vehicle time can be determined by train schedule if the passenger itinerary is known. On the other hand, the out-of-vehicle time consists of walking time and wait time. The walking time considered here includes access and egress walking time and transfer walking time (if needed), which can be determined by walking distance divided by walking speed. Hence, wait time is out-of-vehicle time less walking time. Since the passenger's itinerary is not available in the AFCS data, passenger's in-vehicle and out-of-vehicle times are difficult to estimate, especially under overcrowding situations. A data-driven approach is proposed and discussed here for approximating passenger itinerary with the use of AFCS data and train schedule. Thus, some performance measures concerning QoS (i.e., wait time, average space per standee, and load factor) can be analyzed.

3.1. General Network. A general metro network consists of a set of routes defined as R and a set of stations defined as N. A route consists of two directional lines (e.g., outbound and inbound), and a set of lines is denoted as L. Each station is given a unique ID. For example, the station ID on line 1 begins with 1 and ends at N_1, and then the station ID on line 2 begins with $N_1 + 1$ and ends at N_2, and so on. Hence, the station IDs of line l are $N_{l-1} + 1$ through N_l. For line l, the trains are indexed by m (e.g., $m = 1, 2, \ldots, M_l$), where M_l is the number of trains running on line l within the study time period. Note that a set of transfer stations denoted as S also carries IDs indexed by s.

3.2. Assumptions. The assumptions considered in this study are described as follows and the definitions of variables and model parameters are summarized in Table 1:

(1) Train dispatching and running follow the posted timetable

(2) Walking time is determined by walking distance divided by average walking speed

(3) Passengers will not stay at stations except for waiting for trains

(4) The number of passengers making more than two transfers is small and negligible

3.3. Passenger Itinerary. Figure 1 shows potential itineraries for a passenger, whose access time is consumed by walking from the entry gate to the platform. The wait time is consumed at the platform for the train, considering FtB. The egress time is determined by the passenger who walks from the platform at the destination to the exit gate. For transfer passengers, transfer time consists of walking time and wait time. The walking time is dependent on the distance between the platforms of connecting trains, while the wait time is incurred at the platform for boarding the pick-up train. In this study, passengers are classified into nontransfer and transfer groups.

A nontransfer passenger, after swiping the metro card at entry station, will walk to the platform (i.e., access time is required), wait for boarding the immediate coming train m when the train m is not full, or might wait longer for later trains. After boarding the train, the passenger will spend in-vehicle time and then arrive at the destination station. Finally, the passenger will alight from the train and walk to the gate (i.e., egress time is required), swipe the card, and then exit. Under assumption 3, the in-vehicle and out-of-vehicle times of the passenger as shown in Figure 1(a) can be determined.

For transfer trips after alighting from a delivery train as shown in Figure 1(b), passengers will walk to the platform and wait for a pick-up train. Note that extra wait time is expected if the capacity of the immediate coming train is insufficient, which could be estimated if the probability of passengers FtB is known. The blue solid line represents the case that a passenger boards the pick-up train based on the estimated egress time with assumption 3. The branches with orange dash lines represent the case where passengers could board the immediate coming train m or board the train $m + 1$ (or $m + 2$) due to FtB. Therefore, the passenger with one transfer as shown in Figure 1(b) has three potential itineraries (e.g., Itinerary 1, Itinerary 2, or Itinerary 3).

Similarly, passenger itineraries with more than one transfer can be analyzed. Note that the access, egress, and transfer walking times are dependent on the layout of each metro station and walking speed. Since passengers only

Table 1: Notations.

Variable	Unit	Description
A_{mil}/D_{mil}	—	Arrival/departure time of train m at station i of line l
B_{mlij}	pass	Number of passengers who board train m from station i to station j of line l
C_l	pass	Capacity per train of line l
c_{ls}	pass	Number of seats per train of line l
d, i, j	—	Indices of stations
g, h, k	—	Number of trains skipped by passengers
K_i/K_s	—	Weighted average number of skipped trains at station i/s
L_{dl}	km	The length of the link connecting stations d and $d+1$ of line L
l, l', l''	—	Indices of lines
m, m', m''	—	Indices of trains
$n_i/n_i'/n_s'$	—	Maximum number of trains skipped by passengers at station $i/i'/s$
P_{mlij}	—	Probability of boarding train m of line l from station i to station j
p	—	Index of itineraries
Q_{mdl}	pass	Number of passengers in train m on link d of line l
$r_{ll's}$	sec	Weighted average transfer wait time from line l to line l' at station s
$r_{pll's}$	sec	Wait time at station s for transfer passengers from line l to line l' using itinerary p
s	—	Indices of transfer stations
t_{ai}	sec	Access time from the entrance to the platform at station i
t_{ej}	sec	Egress time from the platform at station j to the exit
t_i/t_j	—	Times when passengers swipe the card at entrance/exit
t_{Vij}	sec	In-vehicle time from station i to station j
t_{wi}	sec	Weighted average wait time at station i
t_{wik}	sec	Wait time at station i skipped for k trains
W_{ik}	pass	Number of passengers who skipped k train at station i
$w_{ll's}$	sec	Transfer walking time from line l to line l' at station s
α_{ik}	—	Probability of the k trains skipped by passengers at station i
$\alpha_{ik}(p)$	—	Weight of skipping k trains at station i for boarding train $m+k$ on itinerary p
$\alpha_{pik}/\alpha_{p,s1,h}$	—	Probability of itinerary p for passengers who skipped k/g trains at station i/s_1
β_p/γ_p	—	Probability of itinerary p for transfer passengers with one transfer/two transfers
ψ_{ml}	m^2	Floor area for standees of train m on line l
φ_{mdl}	m^2/p	Average space per standee in train m on link d of line l
δ_{mdl}	—	Load factor of train m on link d of line l
δ_{ml}	—	Average load factor of train m on line l

swipe the card when entering and exiting the gates, the choice of transfer point(s) is unknown.

Assume that a passenger enters station i of line l at time t_i and would walk t_{ai} seconds to the platform. If this passenger could board the immediate coming train m, the constraint formulated as equation (1) must be sustained:

$$D_{m-1,i,l} \le t_i + t_{ai} < D_{mil}. \quad (1)$$

Note that equation (1) is to ensure that the arrival time of the passenger is between the interdeparture times of trains $m-1$ and m. However, under overcrowding situation, the passenger will wait for more than one train and the resulting wait time will be discussed later.

Based on assumption 3, after a passenger arrives at destination station j, he/she would walk t_{ej} seconds from platform to the exit at time t_j. Thus, the passenger who boards the train m must satisfy

$$A_{mjl} \le t_j - t_{ej} < A_{m+1,j,l}. \quad (2)$$

For a passenger who needs to transfer at station s, the relation among the delivery train arrival time, transfer walking time, and pick-up train departure time will be the concern to infer the possible itinerary. When the departure time of train m' of line l' at station s ($D_{m'sl'}$) is greater than the arrival time of train m of line l (A_{msl}) plus transfer walking time ($w_{ll's}$), the transfer is able to be made successfully. Thus,

$$D_{m'sl'} \ge A_{msl} + w_{ll's}. \quad (3)$$

Note that all possible itineraries for a passenger can be found using equations (1)–(3).

3.4. In-Vehicle Time. The in-vehicle time, denoted as t_{Vij}, can be obtained with train timetable and passenger's itinerary, using

$$t_{V_{ij}} = A_{mjl} - D_{mil}, \quad (4)$$

where A_{mjl} and D_{mil} represent the arrival and departure times of train m on line l at stations j and i, respectively.

3.5. Out-of-Vehicle Time. As discussed earlier, out-of-vehicle time consists of walking time and wait time. The average walking speed can be determined empirically. Hence, the access, egress, and transfer walking times are dependent on the layout of each station. Wait time (e.g., at origin and transfer stations) is a critical factor affecting the variation of

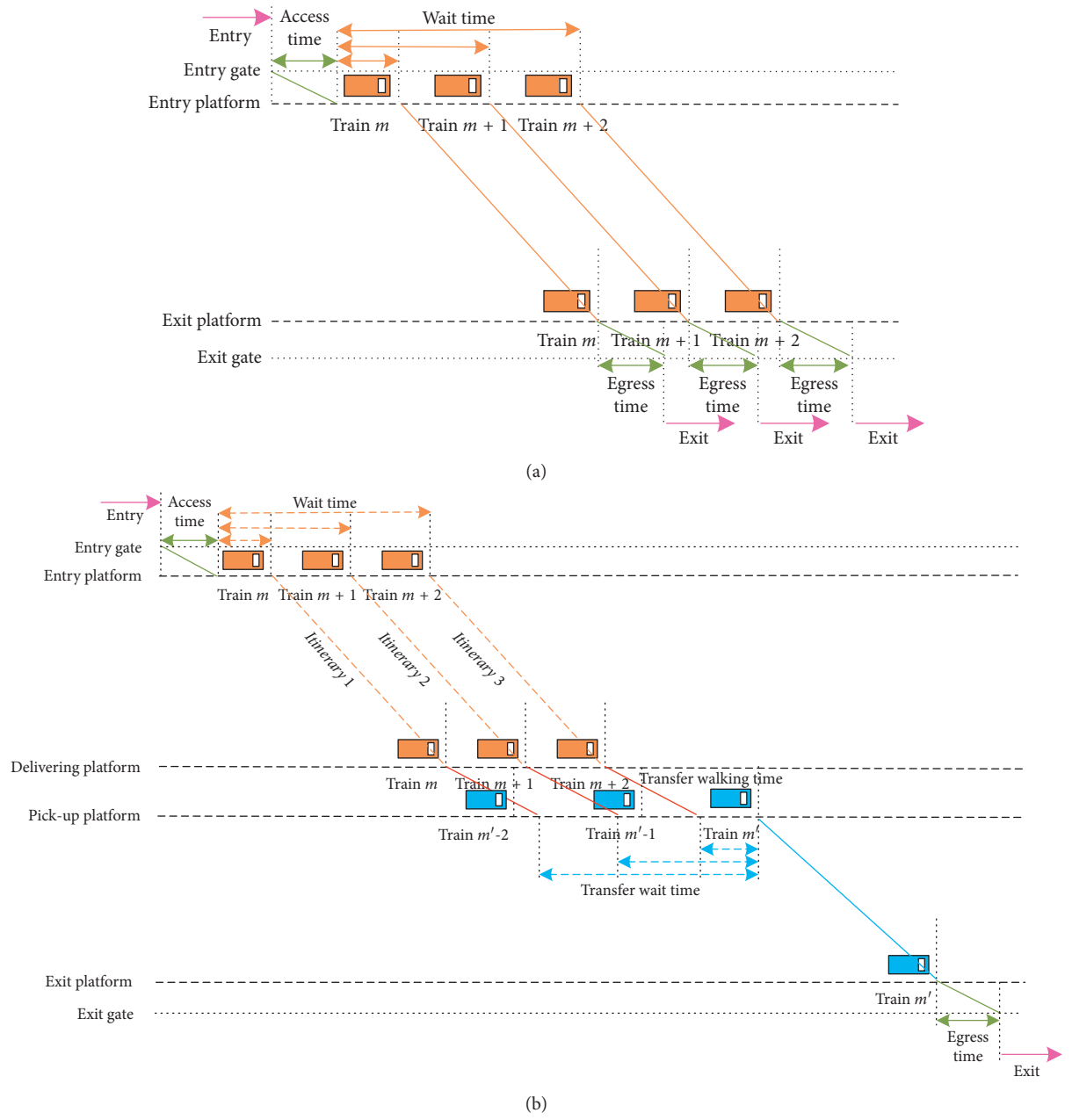


FIGURE 1: Space-time diagrams of passenger itineraries. (a) Passenger itineraries without transfer. (b) Passenger itineraries with one transfer.

travel time, which is dependent on the crowding conditions at stations.

3.5.1. Wait Time for Nontransfer Passengers. After a passenger arrives at the platform of station i, the immediate coming train m can be determined by equation (1), while train $m+k$ actually rode by the passenger can be determined by equation (2). Considering overcrowding situations, the wait time (t_{wik}) for train $m+k$ is expressed by

$$t_{wik} = D_{m+k,i,l} - t_{ai} - t_i. \tag{5}$$

Note that k is zero if the passenger is able to board the immediate coming train m. The number of passengers skipped for k trains at origin station i denoted as W_{ik} can be estimated using the AFCS data, and the probability of skipping k trains denoted as α_{ik} can be determined by

$$\alpha_{ik} = \frac{W_{ik}}{\sum_k W_{ik}}, \quad k = 0, 1, 2, \ldots, n_i, \tag{6}$$

where n_i is the maximum number of skipped trains by passengers at station i. The number of skipped trains at station i, denoted as K_i, is the weighted sum of skipped trains expressed by

$$K_i = \sum_{k=0}^{n_i} k\alpha_{ik}. \tag{7}$$

3.5.2. Wait Time for Transfer Passengers. According to the AFCS transaction records, train timetables, and the layout of the metro network, a set of potential itineraries of a passenger with one transfer is P. In addition to enter and exit stations, itinerary p ($p \in P$) also includes the station(s) to make transfer as well as the delivery and the pick-up train IDs. For instance, the pick-up train m' of line l' at transfer station s can be determined by the exit transaction record and equation (2). Hence, the departure time of train m' at s ($D_{m'sl'}$) can be determined, while the arrival time of delivery train $m+k$ at station i of line l ($A_{m+k,s,l}$) must satisfy the following condition:

$$A_{m+k,s,l} \le D_{m'sl'} - w_{ll's}, \quad k = 0, 1, 2, \ldots, n_i'. \tag{8}$$

Since the immediate coming train m of line l at station i can be inferred by equation (1) and the entry transaction records, the number of skipped trains (i.e., k) can be deduced using equation (8). For instance, the latest train that the transfer passenger must take to arrive at station s is train $m + n_i'$, and the maximum number of trains to skip is n_i'. Since station s can be inferred by the pick-up train schedule, the exit transaction record at destination station j, and equation (2), the key is to find the delivery train boarded at the origin station i. Hence, the probability of itinerary p is analogue to the probability of the train that passenger boarded at station i. Thus, the probability of itinerary p, denoted as β_p, that a passenger travels from station i to station j via s, who skips k trains at i, can be obtained by

$$\beta_p = \alpha_{pik} = \frac{\alpha_{ik}(p)}{\sum_p \alpha_{ik}(p)}, \tag{9}$$

where α_{pik} represents the probability that the transfer passenger can board train $m+k$ (skipped k trains) at origin station i on itinerary p. $\alpha_{ik}(p)$ is the weight of skipping k trains at station i for boarding train $m+k$ on itinerary p, which can be derived by equation (6). Considering the latest arrival time of a delivery train with equation (8), the maximum number of skipped trains (n_i') at station i can be determined.

Under overcrowding situations, the wait time of a transfer passenger at origin station i (denoted as t_{wi}) is the weighted average of wait time considering the probability of passenger who would take itinerary p (i.e., k trains will be skipped at station i) and the associated wait time (t_{wik}) at station i. Thus,

$$t_{wi} = \sum_p t_{wik}\beta_p, \quad p \in P. \tag{10}$$

The number of skipped trains at station i (K_i') is the weighted sum of skipped trains for all itinerary p multiplied by the associated probability β_p. Thus,

$$K_i' = \sum_p k\beta_p, \quad p \in P. \tag{11}$$

For itinerary p, the transfer wait time at station s is denoted as $r_{pll's}$ is $D_{m'sl'}$ less $A_{m+k,s,l}$ and then less the transfer walking time ($w_{ll's}$). Thus,

$$r_{pll's} = D_{m'sl'} - A_{m+k,s,l} - w_{ll's}. \tag{12}$$

Similarly, the transfer wait time at s denoted as $r_{ll's}$ is the weighted average of transfer wait times. Thus,

$$r_{ll's} = \sum_p r_{pll's}\beta_p, \quad p \in P. \tag{13}$$

Since the pick-up train is fixed but the delivery train is still uncertain, the number of skipped trains at station s is dependent on the arrival time of the delivery train. If a passenger boards train $m+k$ at station i and arrives at transfer station s and takes $w_{ll's}$ seconds to walk between platforms of connecting trains, the immediate pick-up train $m' - g$ whose departure time is $D_{m'-g,s,l'}$ can be determined by

$$D_{m'-g,s,l'} \ge A_{m+k,s,l} + w_{ll's}, \quad g = 0, 1, 2, \ldots, n_s', \tag{14}$$

where n_s' represents the maximum number of skipped trains at station s. Note that the pick-up train m' is known from equation (2) and the exit transaction record. The number of skipped trains (g) can be determined by equation (14). Similarly, the number of skipped trains at station s denoted as K_s' can be determined by equation (11).

For itinerary p with two transfers (e.g., the 1st and the 2nd transfer stations are denoted as s_1 and s_2, respectively), a joint probability method is proposed to estimate passenger itinerary. Note that the 2nd pick-up train at s_2 can be determined by the exit transaction record and equation (2); therefore, the purpose is to estimate the delivery train at i and 1st pick-up train at s_1. The probability of itinerary p (γ_p) is dependent on the probability that passenger boarding delivery train $m+k$ at origin station i (denoted as α_{pik}) and the 1st pick-up train $m'+h$ at s_1 (denoted as $\alpha_{p,s1,h}$) can be determined by

$$\gamma_p = \frac{\alpha_{pik}\alpha_{p,s_1,h}}{\sum_p \alpha_{pik}\alpha_{p,s_1,h}}, \tag{15}$$

where $\alpha_{p,s1,h}$ can be derived similarly as discussed for α_{pik}. Note that the departure/arrival time ($D_{m'+h,s1,l'}/A_{m'+h,s2,l'}$) of the 1st pick-up train $m'+h$ at s_1/s_2 must satisfy the following conditions, and the immediate coming and latest pick-up trains at s_1 can be identified:

$$\begin{aligned} &D_{m'+h,s1,l'} \ge A_{m+k,s1,l} + w_{l,l',s1}, \quad h = 0, 1, 2, \ldots, n_{s1}', \\ &A_{m'+h,s2,l'} \le D_{m'',s2,l''} - w_{l',l'',s2}, \quad h = 0, 1, 2, \ldots, n_{s1}' \end{aligned}, \tag{16}$$

where n_{s1}' represents the maximum number of skipped trains at station s_1. Similarly, passenger wait time and the number of skipped trains by passengers with two transfers can be determined using equations (10), (11), and (13).

For a complicated metro network, passengers may face multipath choice decision with transfers. It is necessary to identify potential train(s) for passengers to ride at origin as well as transfer stations considering spatiotemporal constraints, such as train arrival times at different stations of a line and different lines of the network as well as the possibility of FtB. According to the number of transfers associated with each path (also called itinerary), the probability

of each itinerary can be calculated by equations (9) and (15). Then, the average wait time and number of skipped trains can be estimated.

3.6. Quality of Service (QoS). To analyze the performance measures concerning QoS, we investigate *passenger wait time*, *average space per standee*, and *load factor* in different levels of details (e.g., link and train). As discussed earlier, wait time can be estimated by equations (5), (10), and (13). Other equations applied to assess average space per standee and load factor are discussed next.

3.6.1. Average Space per Standee. Average space per standee (ASPS) in square feet (or meters) per passenger can be used to describe the level of crowding on board of the vehicle and that for train m on link d (link connects station d to $d+1$) of line l, denoted as φ_{mdl}, can be calculated by

$$\varphi_{mdl} = \frac{\psi_{ml}}{Q_{m\,dl} - c_{ls}}, \tag{17}$$

where ψ_{ml} represents the floor area for standees. c_{ls} represents the number of seats per train of line l, and Q_{mdl} represents the number of passengers which can be derived by

$$Q_{mdl} = \sum_{i=N_{l-1}+1}^{d} \sum_{j=d+1}^{N_l} B_{mlij} P_{mlij}, \tag{18}$$

where B_{mlij} and P_{mlij} represent the number of passengers from station i to station j and the associated probability of them boarding train m of line l, respectively. Table 2 describes levels of crowding for transit vehicles [1], along with potential implications for passenger perspective.

3.6.2. Load Factor. Load factor at a point is defined as the ratio of passengers transported to spaces offered at maximum schedule load [37], which is a normalized measure used in this study. Therefore, one of the important indicators of system productivity is the load factor for link (δ_{mdl}) and train (δ_{ml}), which can be defined as the total passenger-miles traveled divided by the total capacity-miles provided. Thus,

$$\begin{aligned} \delta_{mdl} &= \frac{Q_{mdl}}{C_l}, \\ \delta_{ml} &= \frac{\sum_{d=1}^{N_l-1} Q_{m\,dl} L_{dl}}{\sum_{d=1}^{N_l-1} C_l L_{dl}}, \end{aligned} \tag{19}$$

where L_{dl} represents the length of the link connecting stations d and $d+1$ of line l. C_l represents the capacity per train of line l.

4. Solution Approach

Considering the variation of travel time, the in-vehicle and out-of-vehicle times should be determined based on time varying demand and supply attributes. Since train running times between stations and dwell times at stations are based on posted timetables, the in-vehicle time is relatively stable compared to out-of-vehicle time. A data-driven approach is developed to estimate the probability of passenger itinerary and the associated wait time as well as passenger load using the AFCS data and train schedule. The step procedure is discussed as follows and illustrated in Figure 2:

Step 1. Data inputs

(i) Extract entry and exit transaction records from the AFCS data
(ii) Input the train timetables, average access, and egress time at each station, as well as walking times at transfer stations

Step 2. Data preprocessing

(i) Eliminate the abnormal trip data missing either entry or exit record
(ii) Classify passengers into nontransfer and transfer trips according to the AFCS transaction records and the layout of metro network

Step 3. Extract the data of nontransfer trips

(i) Identify the boarding train using exit transaction records and estimate wait time at the origin station
(ii) Determine the number of skipped trains based on the schedule of immediate coming train
(iii) Calculate the probability of the trains skipped by passengers and the associated wait time

Step 4. Extract the data of trips with one transfer

(i) Identify the pick-up train by exit transaction records and infer the latest train that a passenger may board at the origin station by transfer connection constraint
(ii) Determine the earliest coming train at the origin station by entry transaction records
(iii) Generate the set of potential delivery trains and all potential itineraries
(iv) Determine the maximum number of skipped trains at the origin and transfer stations and extract the probability of skipped trains at the origin station
(v) Calculate the probability of each potential itinerary and the associated wait time and the number of skipped trains

Step 5. Extract the data of trips with two transfers

(i) Identify the sets of potential delivery and the 1st pick-up train similarly to Step 4 and generate all potential itineraries
(ii) Extract the probability of skipped trains at the origin and the 1st transfer station and calculate the probability of each potential itinerary
(iii) Calculate wait time and the number of skipped trains at the origin and transfer stations

Step 6. Results output

Table 2: Levels of crowding.

Level	Space per standee		Passenger perspective
I	>10.8 ft^2/p	>1.00 m^2/p	Passengers are able to spread out
II	5.4–10.8 ft^2/p	0.5–1.0 m^2/p	Comfortable standing load that retains space between passengers
III	4.3–5.3 ft^2/p	0.40–0.49 m^2/p	Standing load without body contact
IV	3.2–4.2 ft^2/p	0.30–0.39 m^2/p	Occasional body contact
V	2.2–3.1 ft^2/p	0.20–0.29 m^2/p	Approaching uncomfortable conditions
VI	<2.2 ft^2/p	<0.20 m^2/p	Crush loading conditions

Note. See exhibits 5–17 [1].

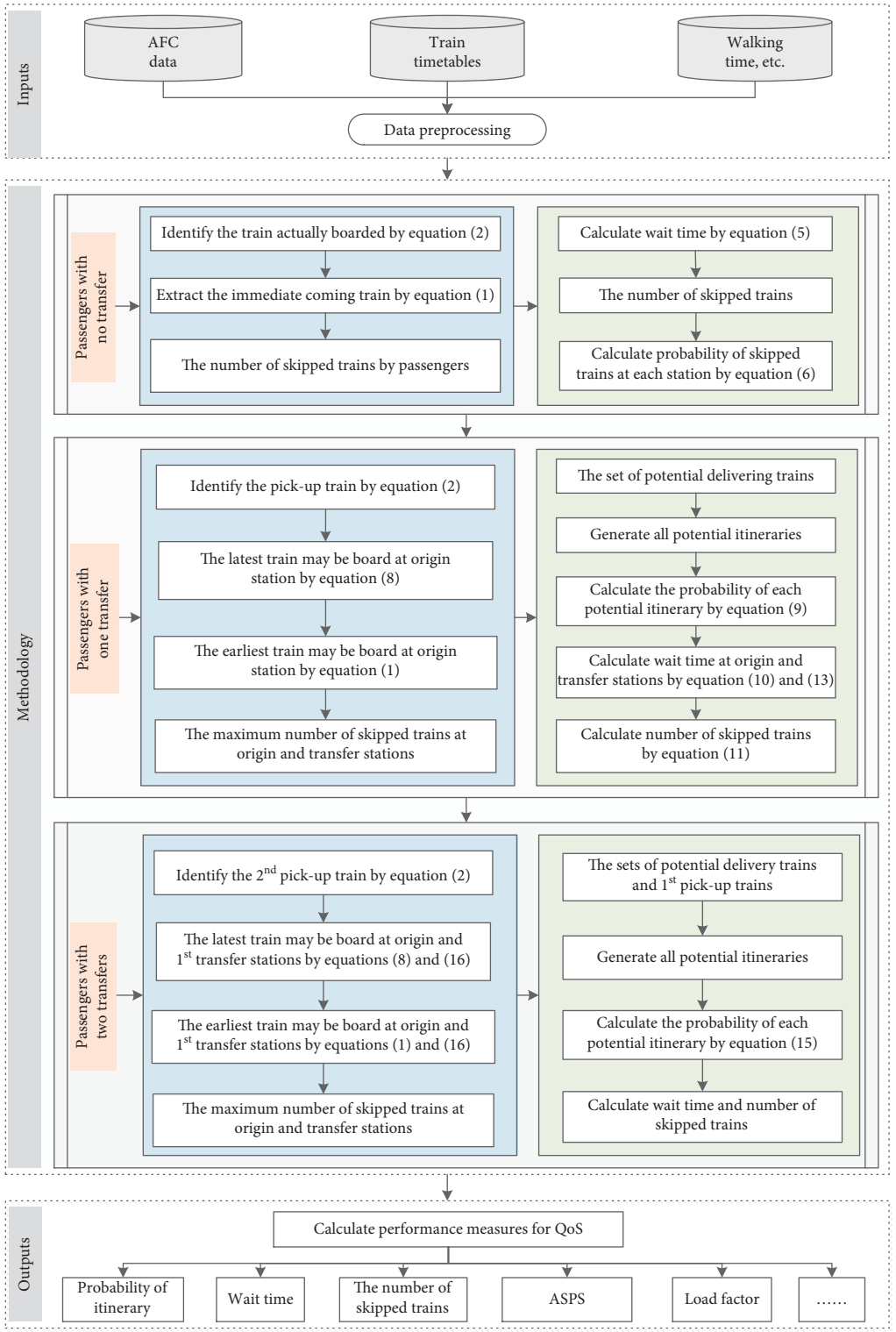


Figure 2: Framework of data-driven approach.

(i) Calculate the value of performance measures, including wait time, number of skipped trains, average space per standee, and load factor
(ii) Output the results

5. Case Study

The studied Chengdu Metro network consists of 6 routes intersecting at 14 transfer stations as shown in Figure 3. Route 7 is a ring that intersects routes 1, 2, 3, 4, and 10. Hence, the maximum number of transfers per trip is less than or equal to two. The ridership on a typical weekday is more than three million. The size of train units (i.e., cars per train) varies with different routes.

The input data are mainly from the AFCS and train timetables. At each station, metro staffs are placed at platforms to assist boarding and alighting passengers and to ensure the timeliness of train departures. Other input data, such as walking time, were estimated based on actual walking distance and speed via the data collected at each station.

5.1. Analysis of Passenger Flow

5.1.1. Ridership Distribution over Time. The ridership distribution over time on a typical weekday showed obviously bimodal peaks as shown in Figure 4, which are the AM peak (i.e., 8:00 am~9:00 am) and off-peak (i.e., 12:00 pm~1:00 pm). It is worth noting that the number of passengers accessing the metro network at the AM peak is nearly 15% of daily ridership, which leads to a very crowding situation.

5.1.2. Passenger Flows Entering and Exiting Stations. The passengers entering and exiting the metro system during AM peak are shown in Figure 5. The largest entering volume was found at XP station (Route 2, 10,300 pass/hr), and the largest exiting volume was at 3TFS station (Route 1, 17,000 pass/hr). Most stations with larger entering flows concentrate near the residential zones, while the major exiting flows fall in the stations near the employment areas.

5.2. Analysis of Travel Time

5.2.1. Travel Time Components. The assessment of travel time components for passengers is shown in Figure 6. In general, the average travel time of transfer passengers is obviously longer than that of nontransfer passengers because of longer travel distance and expected wait/transfer time. In-vehicle time accounts for 69.7% of the travel time for nontransfer passenger (Figure 6(a)). Wait time at origin station is the second longest (19.9% of travel time), while the entry and exit walking times are minors and around 5.0% of travel time, respectively. For passengers with one transfer, in-vehicle time accounts for 70.0% of the travel time, and the percentage of out-of-vehicle time is 30.0%, in which the transfer time ratio is over 15.0% (see Figure 6(b)). For passengers with two transfers, the average travel time is obviously the longest; however, the percentage (55.2%) of in-vehicle time is lower because of increased transfer time as shown in Figure 6(c).

5.2.2. Distribution of Wait Time. Figure 7 illustrates the probability distribution of wait time at different time period. The peaks represent the greatest probability of wait time, and the wait time distributions are similar between the AM peak and PM peak. About 60% of passengers spent less than 3 minutes to wait at station in AM and PM peaks. In general, more passengers experience less wait time at peak periods than that at off-peak due to the smaller headways.

The probability (P_{t_w}) density obeys the Gaussian distribution, which is $P_{t_w} = ae^{-((t_w - b)/c)^2}$, and t_w represents wait time. The fitting results such as the value of parameters (a, b, and c), the Sum Square Error (SSE), R-squared (R^2), Adjusted R-squared (Adjusted R^2), and Root Mean Square Error (RMSE) are shown in Table 3. The peak value $t_{w\max}$ after calculating the derivatives of the fitting functions represents the value of wait time with highest probability. At AM peak, $t_{w\max}$ is 1.98 minutes; at PM peak, $t_{w\max}$ is 2.05 minutes; at off-peak, $t_{w\max}$ is 2.88 minutes; at all day, $t_{w\max}$ is 2.80 minutes.

5.3. Average Wait Time and Number of Skipped Trains on Line 1.

The average wait times at stations as an example on line 1 at AM peak and off-peak are shown in Figure 8. Compared with Figure 9, it is found that the average wait time increases as the number of trains that passengers needed to wait for increases. In general, the average wait time at off-peak is longer than that at AM peak because the headway in off-peak is longer than that in the AM peak, except for a few stations (i.e., TFS, SCG, NJQ, and SRS). The average wait time at many stations (i.e., TFS, SCG, SRS, etc.) is longer than the wait time estimated by previous methods [1] in the peak because of overcrowding, while that is significantly reduced in the off-peak. Note that, as indicated in Figures 8 and 9, there are no data at WGS and SC stations, which are the end terminals for the feeder and main line services, respectively, as shown in Figure 3.

Figure 9(a) illustrates the probability of the number of skipped trains by passengers at stations of line 1 (WJN-SC) at AM peak. The number of FtB passengers varies with time and location, which reveals the crowding situation of the trains and the platform. During AM peak, passenger demand increases and sometimes exceeds the train capacity. Therefore, some FtB passengers must wait for later train(s). TFS, SCG, and SRS stations are deemed as bottleneck stations on line 1, which are all transfer stations. It is found that some passengers skip more than 5 trains and wait time is more than 13 minutes as the headway was 130 seconds. For the off-peak, Figure 9(b) illustrates the similar information to Figure 9(a). It reveals that the status of FtB was eased because of lighter demand.

5.4. Average Wait Time and Number of Skipped Trains at TFS Station on Line 1.

TFS station located in the center of Chengdu city is a transfer station intersected by Routes 1 and

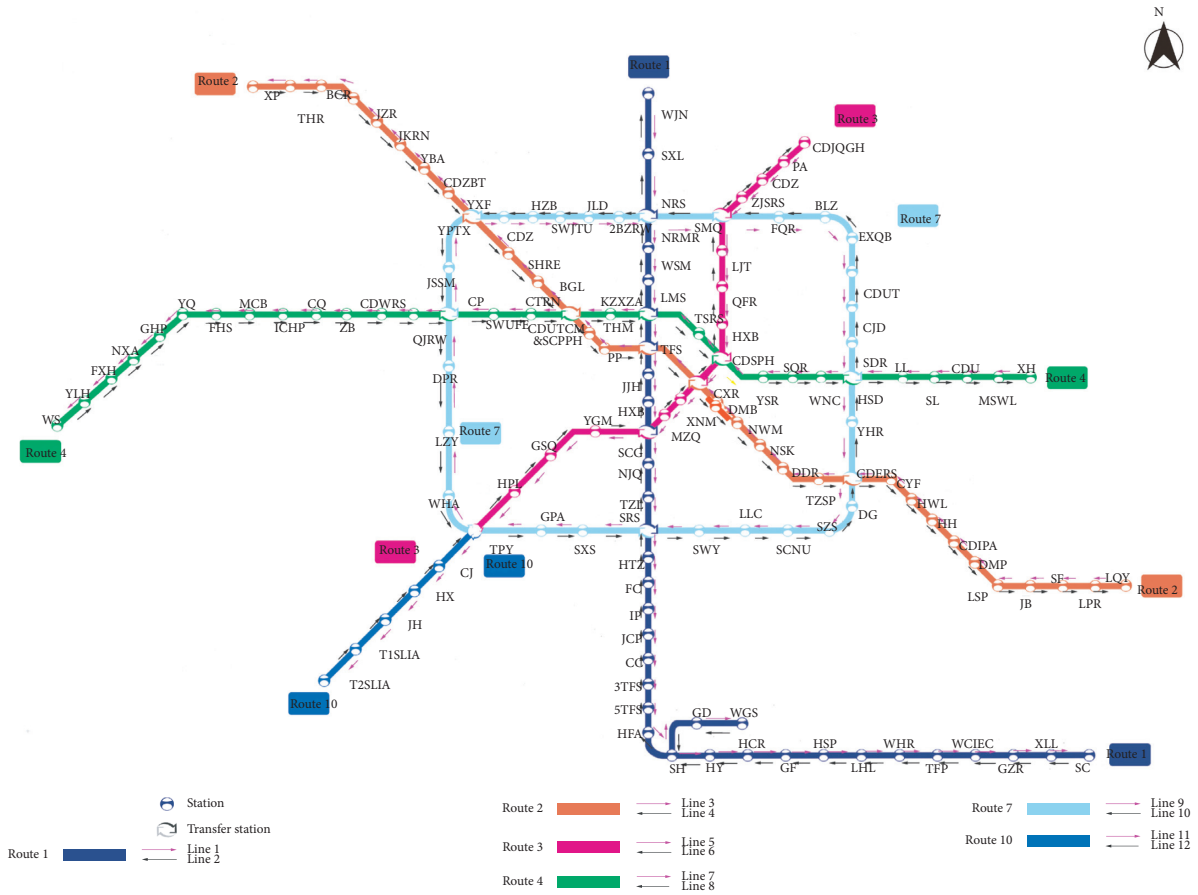


Figure 3: The Chengdu Metro network.

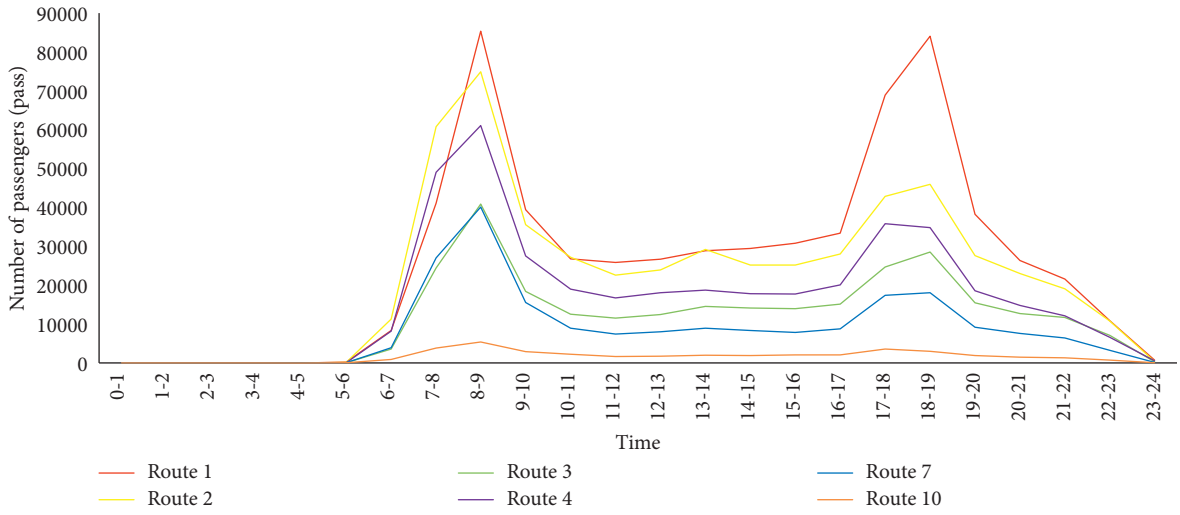


Figure 4: Ridership distribution over time on a typical weekday.

2. The daily volumes consist of nearly 41,000 nontransfer passengers and 180,000 transfer passengers. Taking line 1 as an example, the load factor of the link TFS-JJH affects the number of skipped trains at TFS as shown in Figure 10(a). It was found that as the load factor increases, the number of skipped trains increases.

As shown in Figure 10(b), the red and blue lines represent the average wait times estimated with the developed method

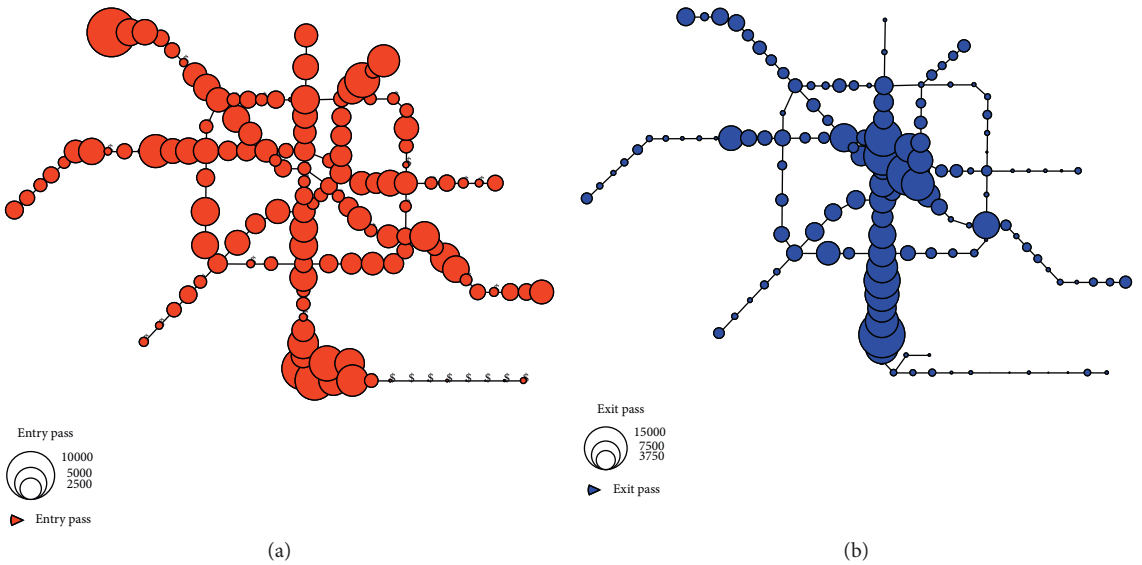


Figure 5: Passenger flows at stations during AM peak. (a) Entry pass. (b) Exit pass.

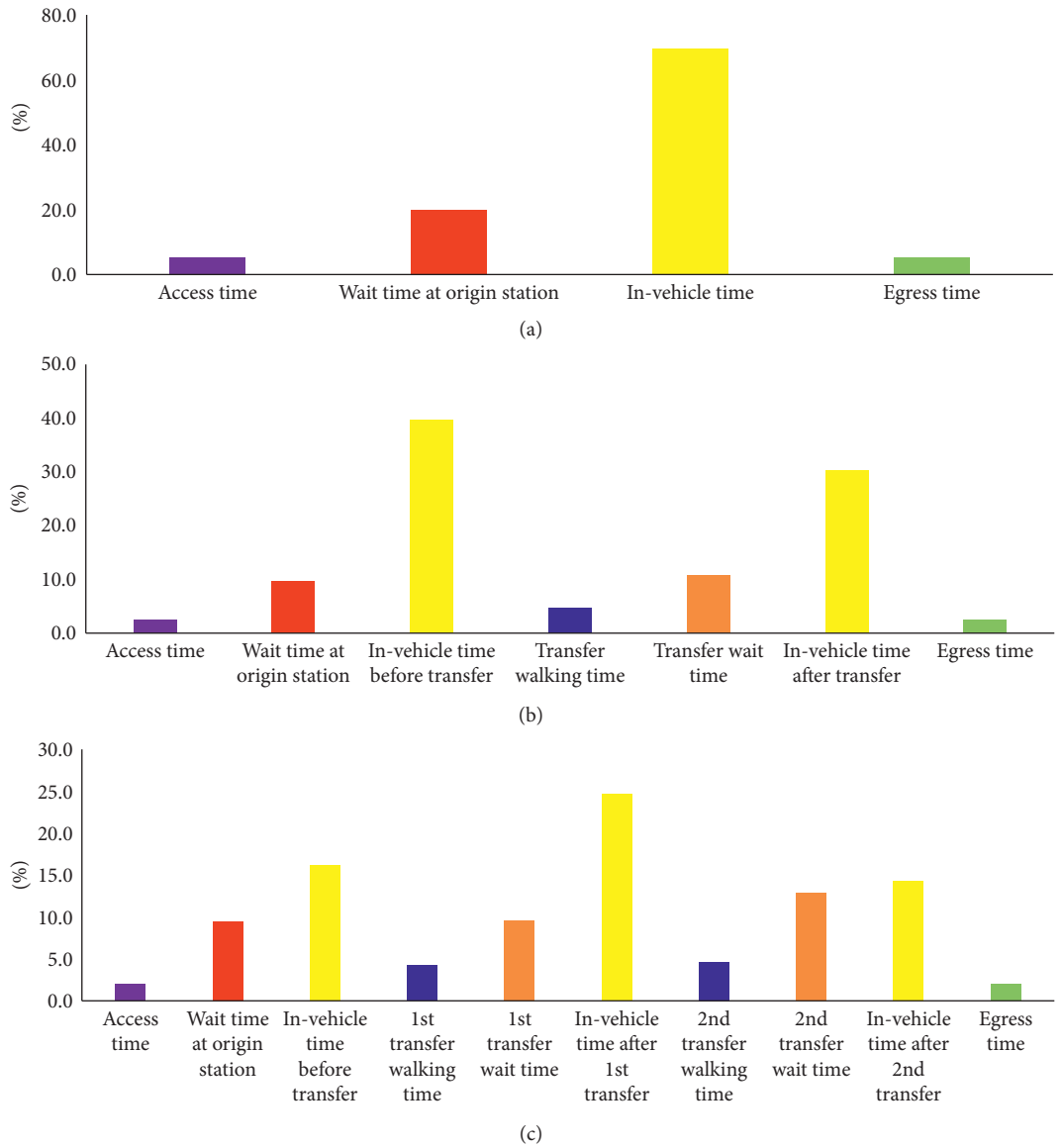


Figure 6: Travel time components for passengers. (a) Nontransfer passengers. (b) Passengers with one transfer. (c) Passengers with two transfers.

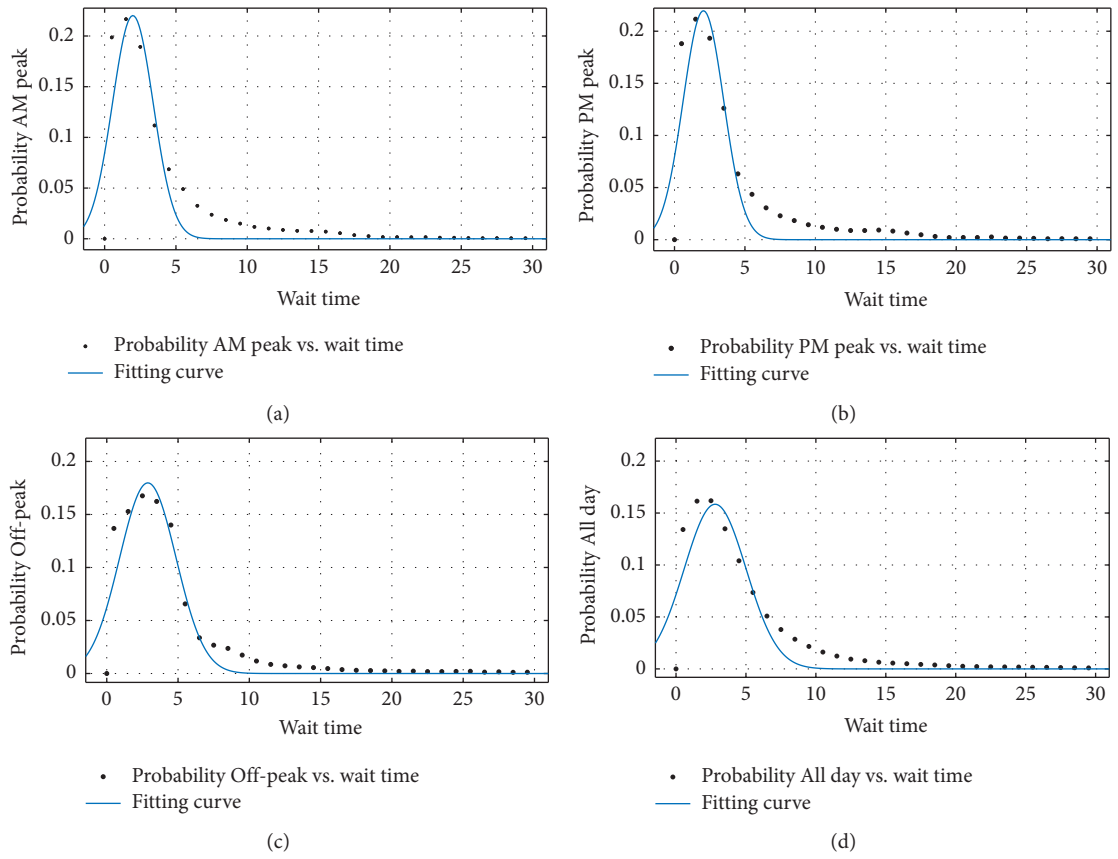


Figure 7: Probability distributions of wait time. (a) AM peak. (b) PM peak. (c) Off-peak. (d) All day.

Table 3: Fitting results.

	a	b	c	SSE	R^2	Adjusted R^2	RMSE
AM peak	0.2201	1.978	2.022	0.01712	0.8475	0.8366	0.02473
PM peak	0.2196	2.054	2.039	0.01456	0.8672	0.8577	0.02281
Off-peak	0.1798	2.882	2.801	0.00821	0.9101	0.9037	0.01712
All day	0.1585	2.797	3.136	0.01031	0.8686	0.8592	0.01919

and the method suggested by TCQSM, respectively, while the green line is headway. It was found that the wait time estimated by the proposed data-driven approach is much higher and closer to the wait time that passengers experienced.

5.5. Analysis of Passenger Load

5.5.1. Average Space per Standee. In the QoS analysis, TCQSM provides general guidance for passenger load standards that can be expressed as an average during a peak 15-, 30, or 60-min period [1]. As shown in Figure 11, the bidirectional (inbound and outbound) ASPS during AM peak is produced (8:00–9:00 AM). The segments colored in green through red represent the ASPS from high to low, which also reflects the average passenger load in metro network.

The distributions of passenger flows in and out of stations were heterogeneous over space and time. For Route 1, the crowding situations existed on the links of downtown or employment concentration areas in both directions. Most routes have similar crowding situations in the downtown area as Route 1, except loop Route 7.

The ASPS of any train can be obtained if the number of boarding passengers, number of seats, and the standing area per train are known. Therefore, the crowding situation of any coming trains is able to be displayed. Some passengers might be willing to skip the immediate coming train and wait a little longer for the next one (if less crowded).

The ASPS and associated standard deviation (SD) on each link of line 1 in the AM peak are illustrated in Figure 12. It indicates that the trains were not congested and there were no standees on the links of WJN-SXL, HSP-SC (main line), and SH-WGS (feeder line) because of high ASPS and low SDs. However, the trains running between stations NRS and 3TFS are constantly crowded with low ASPS and SD, where the ASPS on the links between stations LMS and JCP represents uncomfortable conditions (e.g., Levels V and VI according to Table 2).

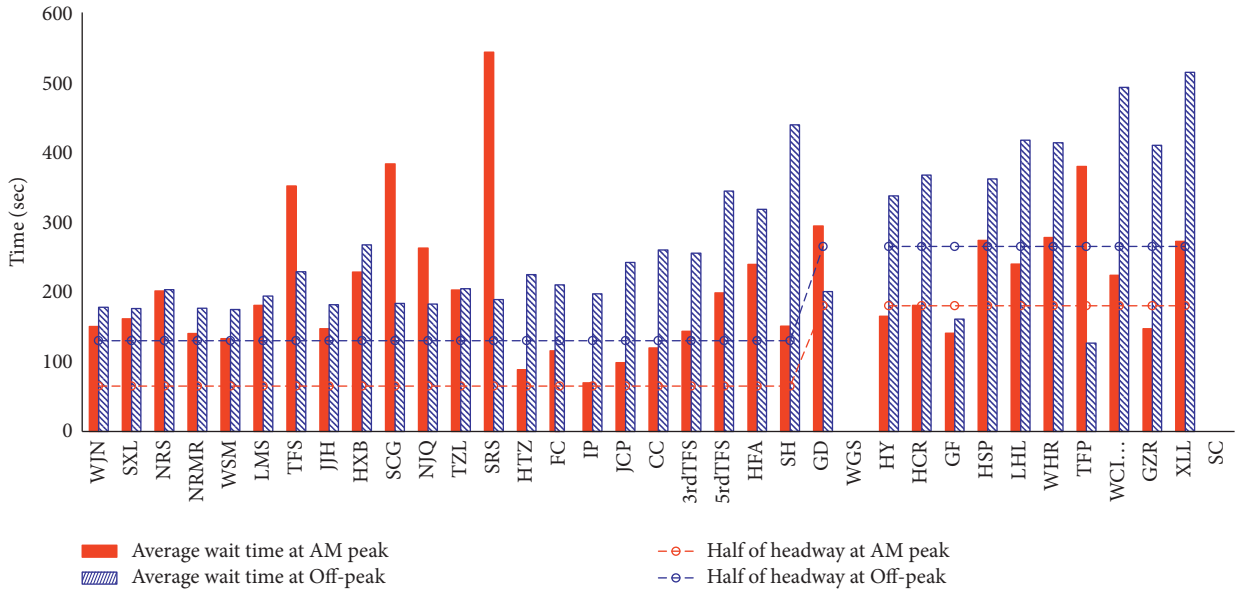


Figure 8: Average wait time at stations on line 1 (WJN-SC).

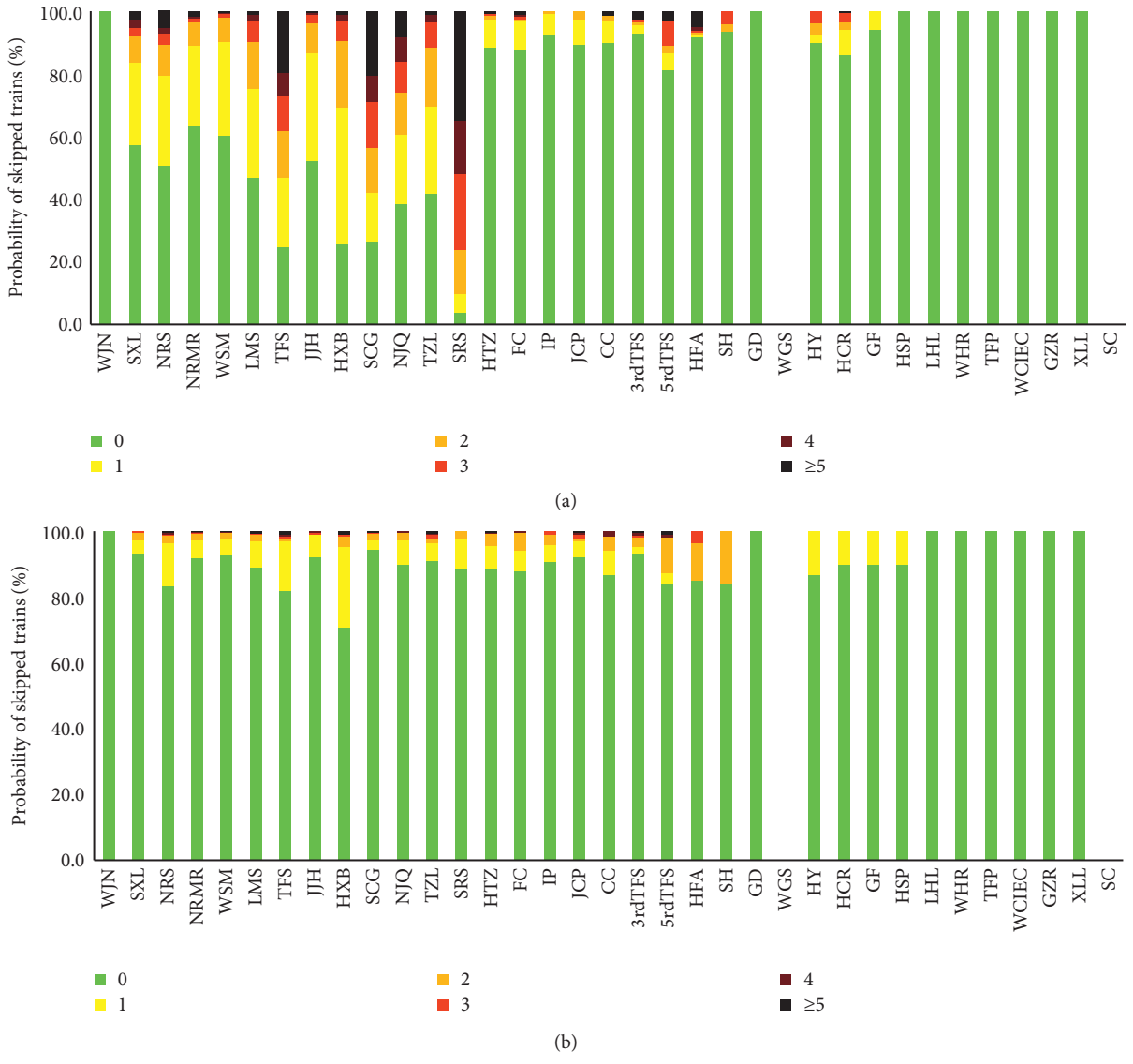


Figure 9: Probability of number of skipped trains by passengers on line 1 (WJN-SC). (a) AM peak. (b) Off-peak.

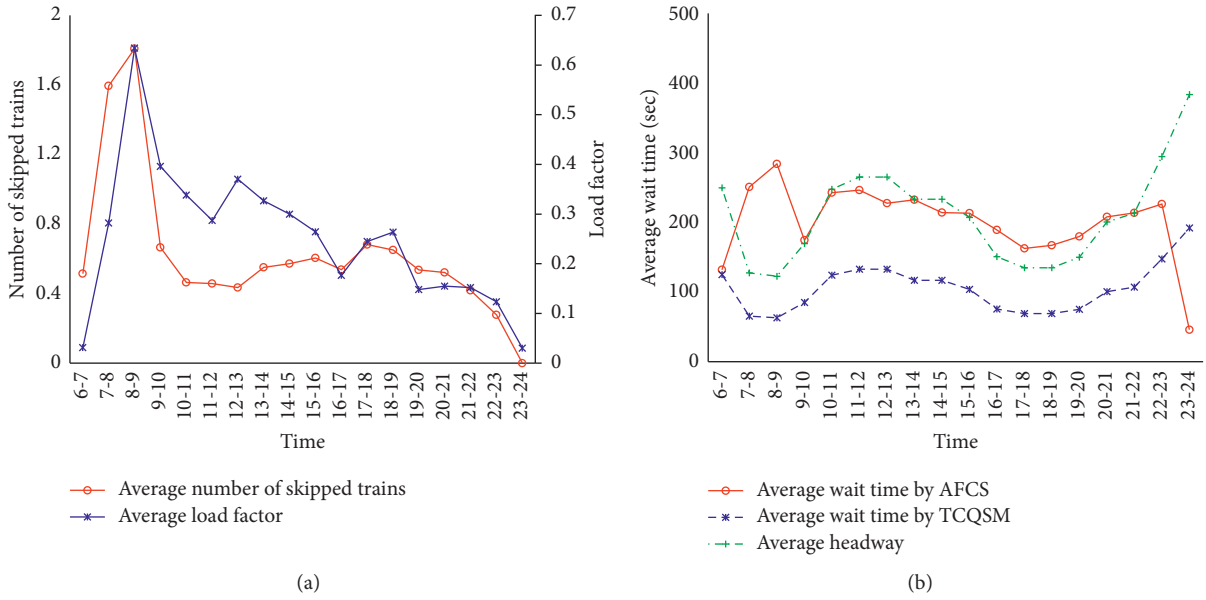


Figure 10: Number of skipped trains and average wait time at TFS station on line 1. (a) Number of skipped trains. (b) Average wait time.

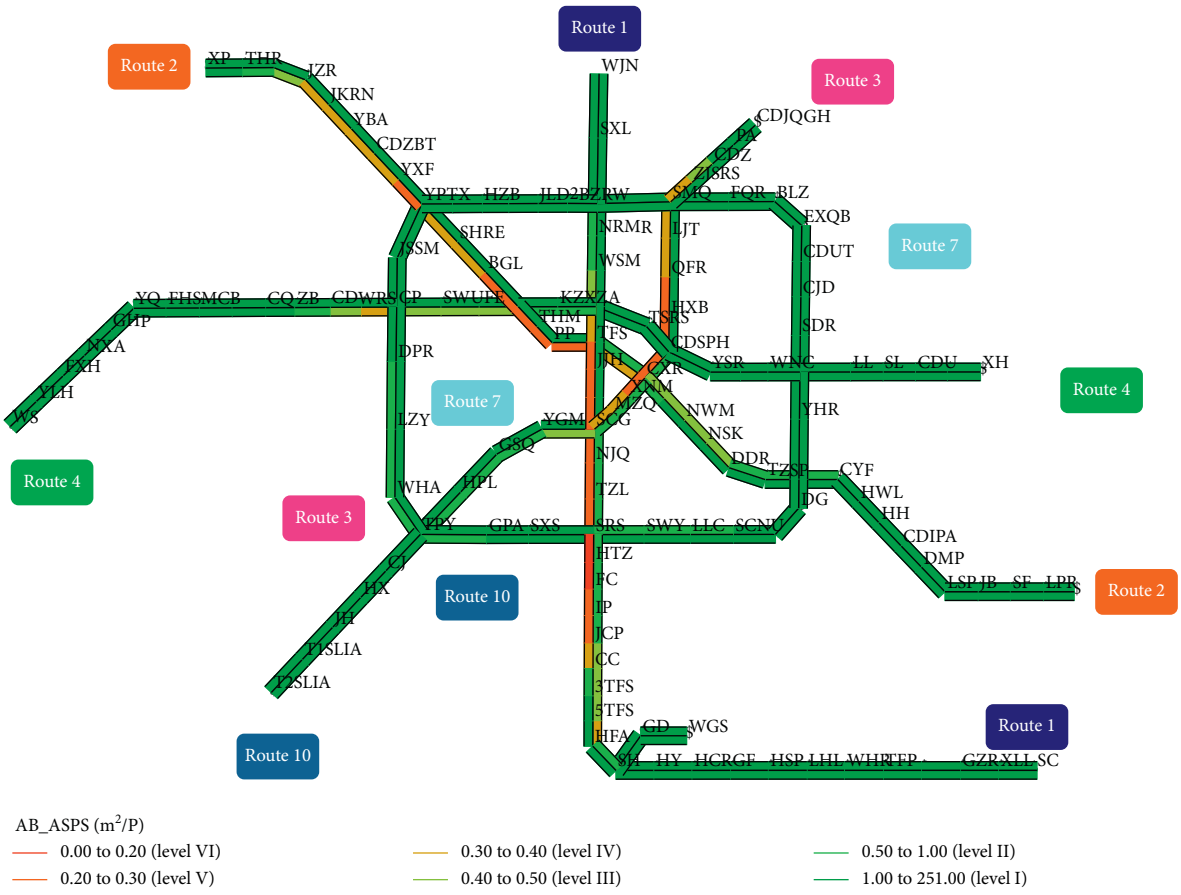


Figure 11: Average space per standee (8:00–9:00 AM).

5.5.2. Load Factor. The spatiotemporal load factor of each train on line 1 (WJN-SC) is shown in Figure 13. The empty areas in Figure 13 represent no service provided since line 1 offers the feeder line and main line train services. The color of a train profile represents the load factor on each link. The congestion level of each train is illustrated in a space-time diagram, in which the red segments with high load factors appeared during AM and PM peaks.

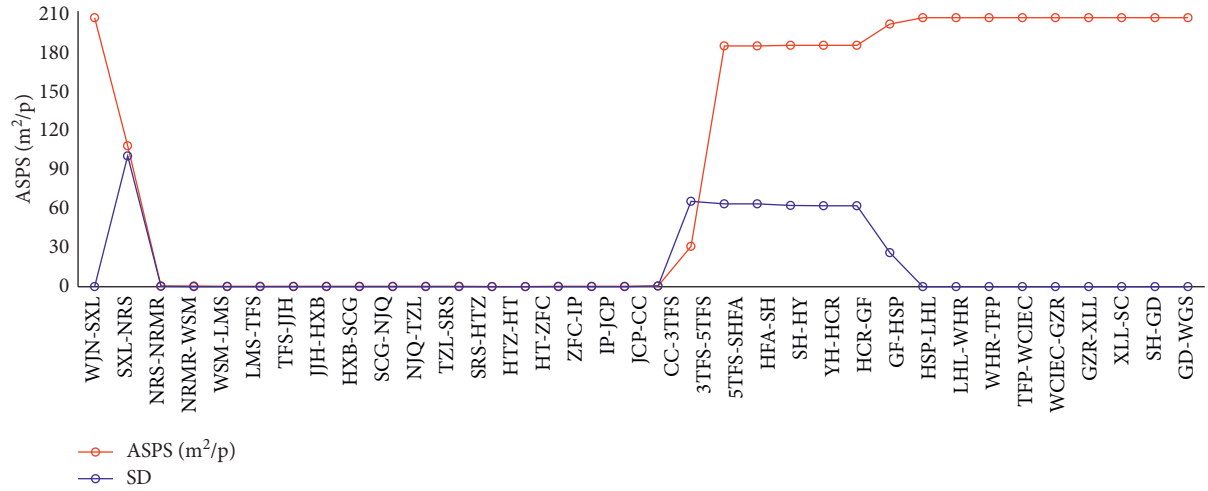


Figure 12: Average space per standee and standard deviation on links of line 1 (AM).

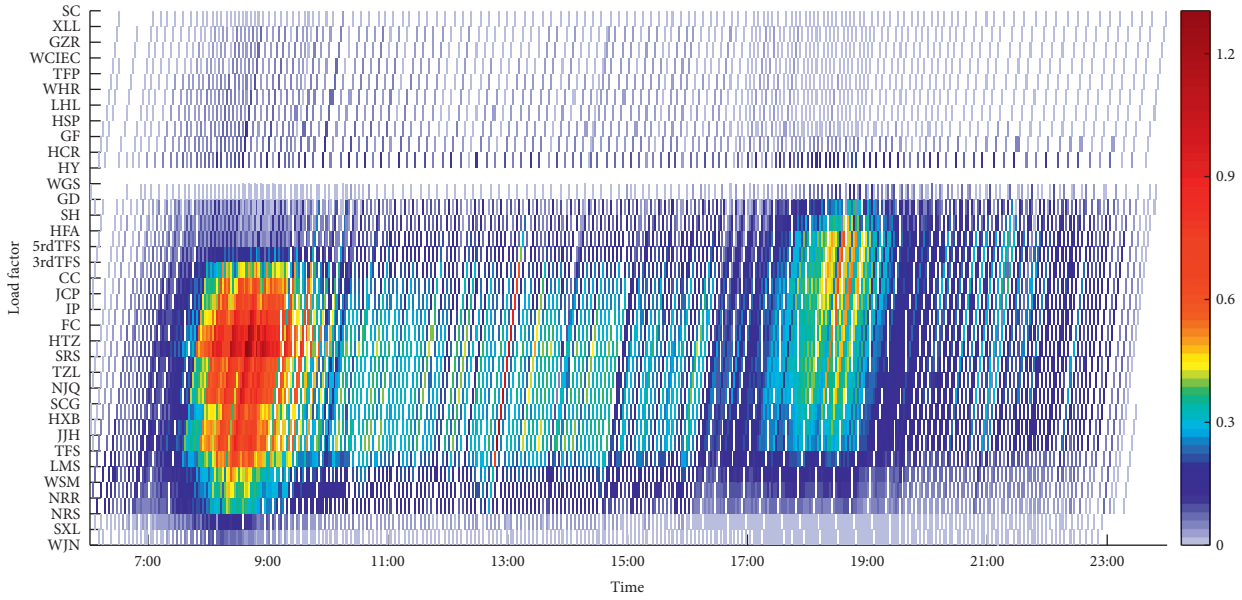


Figure 13: Spatiotemporal load factor for each train on line 1 (WJN-SC).

The spatiotemporal distribution of passenger flow is obviously uneven on line 1, which indicates that the load factor at AM peak is more serious than that at PM peak. For AM peak, more trains have high load factor with longer distance on line 1, while in PM peak, less trains have high load factor with shorter distance. The arrival times related to the work starting times in the AM are focused on, but the off-work time in the PM varies within a longer period. Some passengers work late, and others leave early.

The information about the congestion level of any coming train and station is useful for operator to draft the operation strategy such as the short-turn service, enhancing the train capacity, and proposing reasonable passenger flow limited measures at some crowding stations. In addition, passengers could obtain more accuracy travel time and make more appropriate travel plan.

6. Conclusions

With the AFCS data, the spatiotemporal passenger entry and exit distributions at metro stations can be determined. This study presents a data-driven approach to estimate the probabilities of passenger itineraries considering FtB for a congested metro network. This information is critical for estimating in-vehicle and out-of-vehicle travel times (e.g., walking, wait, and transfer times) as well as spatiotemporal passenger loads in different details (i.e., train, link, etc.). This information would be extremely useful in justifying and/or optimizing service planning (i.e., frequency, train size, service patterns, etc.) to elevate metro's QoS. A case study was conducted using the AFCS data provided by Chengdu Metro in China. The results of the analysis suggest that the probability of wait time obeys the Gaussian distribution. The average wait time, particularly at the AM peak, is much

higher than a half of headway applied in previous studies. The spatiotemporal wait time and the number of skipped trains at each station were able to be estimated effectively.

Several performance indicators related to passenger load were assessed, which include passenger flow, average space per standee, and load factor of any train running on the network. These indicators are critical for various extensions in advanced travel planning, travel time prediction, QoS assessment, developing strategic operating plans, demand forecasting, and dynamic passenger assignment in the context of Mobility as a Service (MaaS). As an immediate extension of this study, the path-choice survey will be conducted and the collected data with train AVL information may be applied to verify the results derived in this study.

Data Availability

The automatic fare collection system (AFCS) data and existing train timetables used to support the findings of this study are unavailable for the public due to confidentiality agreements with research collaborators. Supporting data can only be available to bona fide researchers subject to a nondisclosure agreement.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was financially supported by the Key Research and Development Plan of the Ministry of Science and Technology, China (Grant No. 2018YFB1601402) and the National Engineering Laboratory of Integrated Transportation Big Data Application Technology, China (Grant No. CTBDAT201910). The authors appreciate the data support from Chengdu Metro.

References

[1] Kittelson and Associates, Inc., Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, and Arup, *TCRP Report 165: Transit Capacity and Quality of Service Manual*, Transportation Research Board of the National Academies, Washington, DC, USA, 3rd edition, 2013.

[2] I. Ceapa, C. Smith, and L. Capra, "Avoiding the crowds: understanding tube station congestion patterns from trip data," in *Proceedings of the ACM SIGKDD international Workshop on Urban computing*, pp. 134–141, New York, NY, USA, 2012.

[3] S. Liu, Y. Liu, L. Ni, M. Li, and J. Fan, "Detecting crowdedness spot in city transportation," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, pp. 1527–1539, 2012.

[4] J. M. Bunker, "Assessment of transit quality of service with occupancy load factor and passenger travel time measures," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2535, no. 1, pp. 45–54, 2015.

[5] Y. Y. Ulusoy, S. I.-J. Chien, and C.-H. Wei, "Optimal all-stop, short-turn, and express transit services under heterogeneous demand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2197, no. 1, pp. 8–18, 2010.

[6] Y. Y. Ulusoy and S. I.-J. Chien, "Optimal bus service patterns and frequencies considering transfer demand elasticity with genetic algorithm," *Transportation Planning and Technology*, vol. 38, no. 4, pp. 409–424, 2015.

[7] H.-z. Qu, S. I.-J. Chien, X.-b. Liu, P.-t. Zhang, and A. Bladikas, "Optimizing bus services with variable directional and temporal demand using genetic algorithm," *Journal of Central South University*, vol. 23, no. 7, pp. 1786–1798, 2016.

[8] Z. Guo, "Mind the map! The impact of transit maps on path choice in public transit," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 7, pp. 625–639, 2011.

[9] Y. Xin, L. Fu, and F. F. Saccomanno, "Assessing transit level of service along travel corridors," *Transportation Research Record*, vol. 1927, pp. 259–267, 2005.

[10] J. M. Bunker, "How transit route passenger load and distance can together influence quality of service," in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2014.

[11] P. G. Furth, B. Hemily, T. H. Muller, and J. G. Strathman, *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management*, Transportation Research Board of the National Academies, Washington, DC, USA, 2006.

[12] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and destination estimation in New York city with automated fare system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1817, no. 1, pp. 183–187, 2002.

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

[14] W. Zhu, F. Zhou, J. Huang, and R. Xu, "Validating rail transit assignment models with cluster analysis and automatic fare collection data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2526, no. 1, pp. 10–18, 2015.

[15] L. Hong, W. Li, and W. Zhu, "Assigning passenger flows on a metro network based on automatic fare collection data and timetable," *Discrete Dynamics in Nature and Society*, vol. 2017, Article ID 4373871, 10 pages, 2017.

[16] A. Tirachini, L. Sun, A. Erath, and A. Chakirov, "Valuation of sitting and standing in metro trains using revealed preferences," *Transport Policy*, vol. 47, pp. 94–104, 2016.

[17] T. Liu, A. Ceder, J. Ma, and W. Guan, "Synchronizing public transport transfers by using intervehicle communication scheme: case study," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2417, no. 1, pp. 78–91, 2014.

[18] W. Li, R. Xu, Q. Luo, and S. Jones, "Coordination of last train transfers using automated fare collection (AFC) system data," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 2209–2225, 2016.

[19] X. Liu, M. Huang, H. Qu, and S. Chien, "Minimizing metro transfer waiting time with AFCS data using simulated annealing with parallel computing," *Journal of Advanced Transportation*, vol. 2018, Article ID 4218625, 17 pages, 2018.

[20] L. Sun, D.-H. Lee, A. Erath, and X. Huang, "Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system," in *Proceedings of the ACM SIGKDD international Workshop on Urban computing*, pp. 142–148, New York, NY, USA, 2012.

[21] F. Zhang, J. Zhao, C. Tian, C. Xu, X. Liu, and L. Rao, "Spatio-temporal segmentation of metro trips using smart card data," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1137–1149, 2015.

[22] D. Luo, L. Bonnetain, O. Cats, and H. van Lint, "Constructing spatiotemporal load profiles of transit vehicles with multiple data sources," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 175–186, 2018.

[23] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transportation Research Procedia*, vol. 23, pp. 362–379, 2017.

[24] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "A probabilistic passenger-to-train assignment model based on automated data," *Transportation Research Part B: Methodological*, vol. 104, pp. 522–542, 2017.

[25] E. Miller, G. E. Sánchez-Martínez, and N. Nassir, "Estimation of passengers left behind by trains in high-frequency transit service operating near capacity," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 8, pp. 497–504, 2018.

[26] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach," *Transportation Research Part B: Methodological*, vol. 67, pp. 235–263, 2014.

[27] H. Qu, X. Xu, and S. Chien, "Analysis of passenger load and wait time for a metro system with automatic fare collection data," in *Proceedings of the Transportation Research Board 98th Annual Meeting*, Washington, DC, USA, 2019.

[28] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.

[29] Y. Sun and P. M. Schonfeld, "Schedule-based rail transit path-choice estimation using automatic fare collection data," *Journal of Transportation Engineering*, vol. 142, no. 1, Article ID 04015037, 2016.

[30] J. Zhao, F. Zhang, L. Tu et al., "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 790–801, 2017.

[31] Y. Jiang and W. Y. Szeto, "Reliability-based stochastic transit assignment: formulations and capacity paradox," *Transportation Research Part B: Methodological*, vol. 93, pp. 181–206, 2016.

[32] T. Liu and A. Ceder, "Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: a bi-objective bi-level model using deficit function approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 935–955, 2018.

[33] Y. S. Zhang and E. J. Yao, "Splitting travel time based on afc data: estimating walking, waiting, transfer, and in-vehicle travel times in metro system," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 539756, 11 pages, 2015.

[34] H. Lee, D. Zhang, T. He, and S. H. Son, "Metrotime: travel time decomposition under stochastic time table for metro networks," in *Proceedings of the 2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, Hong Kong, China, pp. 1–8, May 2017.

[35] A. Tavassoli, M. Mesbah, and A. Shobeirinejad, "Modelling passenger waiting time using large-scale automatic fare collection data: an Australian case study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 58, pp. 500–510, 2018.

[36] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: a smart card data analysis," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 292–306, 2018.

[37] V. R. Vuchic, *Urban Transit: Operations, Planning, and Economics*, John Wiley & Sons, Hoboken, NJ, USA, 2017.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 8365194, 20 pages
https://doi.org/10.1155/2020/8365194

Research Article

Designing High-Freedom Responsive Feeder Transit System with Multitype Vehicles

Zhengwu Wang,[1,2] Jie Yu,[1,2] Wei Hao,[1,2] Tao Chen,[1] and Yi Wang[3]

[1]*Hunan Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems, Changsha University of Science and Technology, Changsha 410114, China*
[2]*School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410114, China*
[3]*International Business School, Hunan University of Technology and Business, Changsha 410205, China*

Correspondence should be addressed to Jie Yu; jiaotongyujie@163.com and Wei Hao; haowei@csust.edu.cn

Received 3 January 2020; Revised 31 May 2020; Accepted 8 July 2020; Published 25 July 2020

Academic Editor: Yu Jiang

The last mile travelling problem is the most challenging part when using public transit. This study designs a high-freedom responsive feeder transit (HFRFT) system to serve at the transfer station, given vehicle routes, departure time, and service area based on demand. The proposed feeder transit system employs a travelling mode with multitype vehicles. In order to improve the operation of the HFRFT system, the optimization design methods are suggested for vehicle routes, scheduling, and service area. A mixed integer programming model and its hybrid of a metaheuristic algorithm are proposed to efficiently and integrally solve the vehicle routes and scheduling parameters according to the reservation requirements. A heuristic method is proposed to optimize the service area based on the equilibrium of system supply and demand. Case studies show that the mixed running mode of multiple models can significantly improve the seat utilization, which can also significantly reduce the number of departures and the average travel distance per passenger. The proposed service area optimization method is proved to be feasible to improve the last mile travel.

1. Introduction

In areas where passenger travel density is low, fixed route transit (FRT) has a low seat utilization rate due to low and scattered passenger flow. The bus company has a low operating efficiency, which is difficult to sustain. With the development of communication technology and automatic control technology, intelligent public transportation has gradually become a major trend in the development of traffic field [1–7]. Consequently, the public transportation field is largely affected by the Internet Plus Era, and passenger information can connect with bus company in real time. Using the Internet-of-things, passengers can indicate their own travel needs through online travel information platforms, and bus companies can also plan bus schedules based on passenger travel information and actual operating conditions in response to the travel demand. In recent years, a new kind of transportation mode called demand-responsive transit (DRT) has emerged. Compared with the conventional FRT, DRT can realize door-to-door service and effectively solve the last mile travelling problem with improved mobility and flexibility. Moreover, DRT is more suitable for low-density residents' travel areas and disadvantaged people (the elderly, the disabled, etc.) than FRT.

Responsive feeder transit (RFT) is a special form of demand-responsive transit (DRT), also known as demand-responsive connector (DRC), which is a flexible transit service because it operates in a demand responsive fashion within a service area and moves customers to/from a transfer point that connects to an RFT network [8]. RFT services have a significant advantage that they can increase the service coverage and accessibility compared with traditional fixed route transit, by offering a capacity with better alignment [9, 10].

Most of the current studies on DRT (shown in Table 1) focused on the vehicle's routing model, coordinated application of DRT and FRT, key parameters such as service performance, scheduling parameters including vehicle departure time, total number of vehicles, vehicle type etc., geometric features in the service area, critical demand density, and so on. However, limited studies have been conducted to study the coordination of vehicle routes and schedules.

In many cases of DRT, vehicle running mode can involve four types, as shown in Figure 1. Type A is the main running mode of mobility allowance shuttle transit (MAST) [51], high-coverage point-to-point transit [57], and demand-responsive feeder transit circulator [13]. Type B is the main running mode of dial-a-ride [18, 19]. Type C and D are both the main running modes of RFT [15, 30].

In Figure 1, the running route and bus stop of type A are basically fixed with only part of the route being adjusted according to the passengers' demands, and passengers are allowed to get on and off at the middle stops and the terminal stops. Meanwhile, the running route and service areas of type B are not fixed. After the vehicle completing one journey, it will continue the next journey; both type C and type D operate around the transfer station (the service area is relatively fixed), and the terminal stops are located at the transfer stations. Type C can only pick up passengers from the middle stops to the transfer station or from the transfer station to the middle stops, while type D can pick up passengers from the middle stops to the transfer station and from the transfer station to the middle stops simultaneously.

RFT can be divided into two categories based on the freedom of service areas, running route, and bus stops.

1.1. Low-Freedom RFT. Low-freedom RFT (LFRFT), whose terminal stops and some sections of the route and middle stops are fixed when others may be limitedly adjusted, can only respond to passengers' demands to a lower degree. LFRFT is formed by partially adjusting FRT (Nourbakhsh and Ouyang [28], Yu et al. [13], Guo et al. [10], Lu et al. [40, 41], Gschwender et al. [36], Alshalalfah and Shalaby [29], and so on) (shown in Table 2).

1.2. High-Freedom RFT. For the high-freedom RFT (HFRFT), the service area and terminal stops are fixed. HFRFT can respond to passengers' demands to a greater extent. In HFRFT, running route, vehicle departure time, middle stops, stay times, and so on are optimally determined according to passenger demand. HFRFT has attracted the attention of some researchers, such as Li and Quadrifoglio [8], Edwards and Watkins [32], Sun et al. [20], and so on (shown in Table 2).

Table 2 summarizes previous studies on RFT. One of the remarkable features of RFT is that the analytical model and planning model have been proposed to optimize the RFT service under the reservation demands. For example, Chandra and Quadrifoglio [30], Edwards and Watkins [32], and Nourbakhsh and Ouyang [28] came up with a structured flexible transit system and built the optimization model to seek the optimum network layout, service area of each bus, and bus headway. Pan et al. [15] proposed a mixed integer linear programming model to optimize the service area and transit route planning concurrently, which features a two-level structure with an upper level to maximize the number of served passengers by the feeder transit system and a lower level to minimize the operational cost for transit operators. However, several gaps still remain in previous studies:

(i) Research on optimizing path and scheduling simultaneously is quite limited, although some studies have determined departure time according to the routes required to reservation demands [15, 20, 36].

(ii) Optimization of service areas is rarely involved when the optimization design is carried out under fixed regular service area. In fact, the service area and road network are random because land use and travel demand are not evenly distributed. Optimization of service area can fully improve the system efficiency while system demands are met. At the same time, it can also facilitate to fulfill transfer services by subareas and avoid the crossover of service areas.

(iii) Some assumptions are inconsistent with reality, such as the uniform distribution of passengers, regular service area, only one type of vehicle, and unlimited vehicle capacity. In fact, vehicle capacity has a strong effect to vehicle route and scheduling. Moreover, the impact in the mixed running mode including both pick-up and deliver passengers is even greater because deliveries at each demand point will affect pickups in subsequent and pre-sequent stations, e.g., Kirchler and Wolfler Calvo [19]. As the demand for passengers increases in the service area, it may be necessary to choose larger capacity vehicles and operate multiple vehicles at the same time (Ceder [59], Guedes and Borenstein [60]).

In light of the above deficiencies, taking the mixed running mode with capacity constraint as the research object, this study proposes a integration optimization method for multitype vehicle routes and schedule on the basis of reservation requirements. The remainder of the paper is organized as follows. Section 2 presents a mixed integer programming (MIP) model and its solution algorithm for multitype vehicle routes and schedule problem. Section 3 puts forward a heuristic method to optimize service area according to the principle of capacity and demand equilibrium. Section 4 displays a case to illustrate the proposed model and algorithm. Some concluding remarks and possible future work are given in Section 5.

2. MIP Model and Its Solution Algorithm for Route and Scheduling

2.1. Problem Statement. The application scenarios of HFRFT under the mixed running mode, capacity constraint,

Table 1: A summary of existing studies on DRT.

Research content	Papers
Vehicle routing model	Saberi and Verbas [11]; Yan et al. [12]; Yao et al. [13]; Ma et al. [2, 3]; Dessouky et al. [14]; Pan et al. [15]; Núñez et al. [16]; Ghannadpour et al. [17]; Schilde et al. [18]; Kirchler and Calvo [19]; Sun et al. [20]; Yu et al. [13]; Tang et al. [21]
Coordinated application of DRT and FRT	Aldaihani et al. [22]; Sheu [23]; Guo et al. [10]; Rahimia et al. [24]; Chen and Nie [25]; Qiu et al [26]
Service performance	Alshalalfah and Shalaby [27]; Nourbakhsh and Ouyang [28]; Alshalalfah and Shalaby [29]; Chandra and Quadrifoglio [30, 31]; Edwards and Watkins [32]; Kelly et al. [33]; Engelen et al. [34]; Freia et al [35]; Gschwender et al. [36]; Rahimia et al. [24];
Scheduling parameters	Kikuchiand Donnelly [37]; Hsu [38]; Nair and Miller-Hooks [39]; Lu et al. [40, 41]; Zheng et al. [42]; Rahimi et al. [24]; Tang et al. [43]; Guo et al. [44]; Chandra and Quadrifoglio [45]; Dessouky and Adam [46]; Nourbakhsh and Ouyang [28]; Gschwender et al. [36]
Geometric features in the service area	Zhao and Dessouky [47]; Li and Quadrifoglio [48]; Errico et al. [49]; Chandra and Quadrifoglio [30, 31]; Li and Quadrifoglio [50];
Critical demand density	Quadrifoglio et al. [51–53]; Amiripour et al. [54]; Frei et al. [35]
Site selection	Pratelli and Schoen [55]; Razi and Muhammad [56]

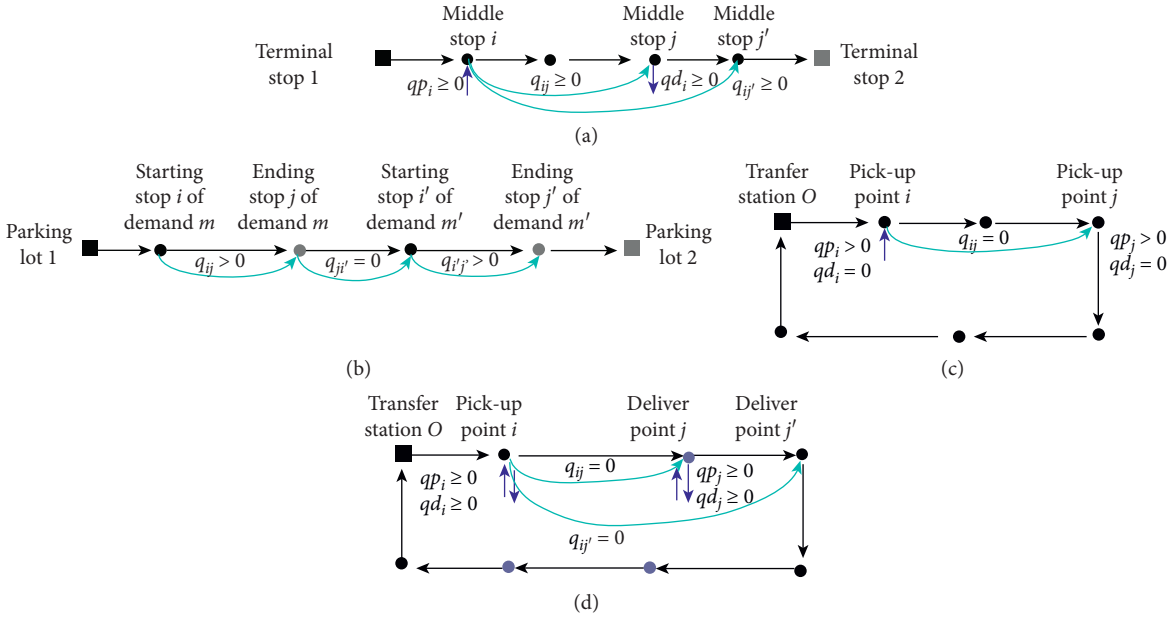


Figure 1: Four types of vehicle running modes (q_{ij}(i or $j \neq o$) are the demands from stop i to stop j, and qp_i and qd_i are the number of pick-up or deliver passengers in the stop i, respectively). (a) Type A. (b) Type B. (c) Type C. (d) Type D.

multi-type vehicle, and reservation requirements are listed as follows.

There are a certain number of multitype vehicles. For the actual bus operation, there are usually two passengers' types in the system, one is passengers from different places to the transfer station and the other is passengers from the transfer station to different places, both of which are limited by the time window. We define these two types of passengers as those who need to be picked up and need to be sent, respectively. According to the different types of passengers, the bus running mode can be divided into alone running mode (only pick-up or deliver) and mixed running mode (both pick-up and deliver), as shown in Figure 2. The same shift of the former is only allowed to serve the passengers need to be picked up or sent, while the latter is allowed to serve the passengers need to be picked up and sent at the same time. These passengers can send their own travelling information to the transit companies via customer service, mobile phone, or Internet, and the travelling information includes travel demand, boarding time/get-off time, pick-up location/deliver location, etc. After collecting the travelling information, the transit company arranges a team of vehicles to complete all passengers' transfer tasks and maximize the system's total benefit through the adjustment of departure time, vehicle type, and running route of all vehicles under passenger time window and vehicle capacity constraints.

Firstly, aiming at the maximum total benefit of the HFRFT system, the coordinated optimization model of vehicle routing and scheduling is constructed under the constraints of vehicle capacity, passenger time window, and

Table 2: A summary of some existing studies on RFT.

Reference	Research object	Model type	Vehicle number	Area shape	Departure time	Running mode	Demand type	Capacity constraint	Demand distribution	Solution method
Nourbakhsh and Ouyang [28]	LFRFT	Planning model	Multivehicles	Square	Fixed	Type A	Reservation	yes	Uniform	Steepest decent
Yu et al. [13]	LFRFT	Planning model	Multivehicles	Circular	Time windows	Type D	Reservation	yes	Random	Heuristic method
Guo et al. [10]	LFRFT	Analytical model	Multivehicles	Rectangle	Fixed	Type A	Stochastic	No	Random	-
Lu et al. [40, 41]	LFRFT	Planning model	One vehicle	Rectangle	Fixed	Type A	Reservation limited	No	Random	Heuristic method
Gschwende et al. [36]	LFRFT	Analytical model	Multivehicles	Rectangle	—	Type A	Determined	No	—	—
Alshalalfah and Shalaby [29]	LFRFT	Planning model	Multivehicles	Square	Fixed	Type C	Determined	yes	—	Heuristic method
Li and Quadrifoglio [8]	HFRFT	Analytical model	One vehicle	Rectangle	Circulation	Type D	Reservation	No	Uniform	Insertion algorithm
Quadrifoglio and Li [53]	HFRFT	Analytical model	Two vehicles	Rectangle	Circulation	Type C	Reservation	No	Uniform	Insertion algorithm
Edwards and Watkins [32]	HFRFT	Analytical model	One vehicle	Circular	Circulation	Type D	Reservation	No	Uniform	Insertion algorithm
Li and Quadrifoglio [48]	HFRFT	Analytical model	One vehicle each zone	Rectangle	Circulation	Type D	Reservation	No	Uniform	Insertion algorithm
Chandra and Quadrifoglio [31]	HFRFT	Analytical model	Multivehicles	Rectangle	Fixed	Type D	Reservation	No	Uniform	Insertion algorithm
Chandra and Quadrifoglio [30]	HFRFT	Analytical model	One vehicle	Rectangle	Fixed	Type C	Reservation	No	Uniform	Insertion algorithm
Chandra and Quadrifoglio [45]	HFRFT	Analytical model	Multivehicles	Rectangle	Fixed	Type D	Reservation	No	Uniform	Insertion algorithm
Li and Quadrifoglio [50]	HFRFT	Analytical model	2 vehicles	Rectangle	Fixed	Type C	Reservation	No	Uniform	Insertion algorithm
Sun et al. [20]	HFRFT	Planning model	Multivehicles	Rectangle	—	Type C	Reservation	yes	Uniform	Hybrid Bat
Chandra et al. [58]	HFRFT	Analytical model	Multivehicles	Square	Fixed	Type D	Reservation	No	Uniform	Insertion algorithm
	Call-n-Ride	Planning model			—	Type B	Reservation	No	Random	Heuristic method
Pan et al. [15]	HFRFT	Planning model	Multivehicles	Determined	—	Type C	Reservation	yes		Heuristic method

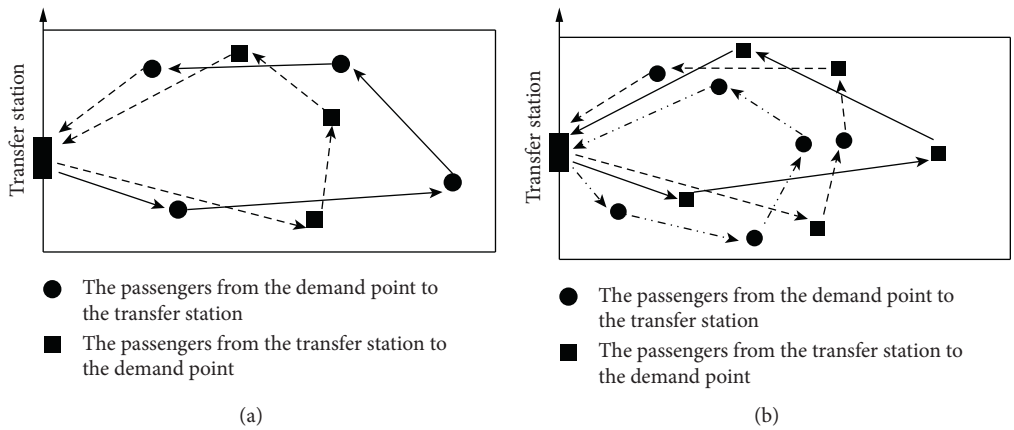


Figure 2: The vehicle route of different operation modes. (a) Alone running mode (only pick-up or deliver). (b) Mixed running mode (both pick-up and deliver).

maximum one way running time of vehicle. In the process of model construction, considering the problems of high repetition rate of running path and low average seat utilization rate of vehicles in the alone running mode, the mixed running mode with simultaneous pick up and deliver is proposed. In order to compare the performance of different running modes, it is assumed that there are three types of vehicles in RFT system, i.e., vehicles can only pick up, vehicles can only deliver, and vehicles can pick up and deliver simultaneously. Secondly, combined with the above coordination optimization model, the optimization method of service area is proposed based on the matching of transportation capacity and passenger demand.

2.2. Assumptions. The following assumptions are made to ensure that the proposed formulations are tractable and realistically reflect the real-world constraints:

(i) The feeder bus begins and ends each of its trips at the transfer station.

(ii) Pick-up/deliver requests sent by the passengers are collected and responded to before the beginning of each trip.

(iii) Running speed of all type vehicles is the same on all road sections.

(iv) There are enough vehicles to meet all passenger requests. All requests are responded in service area and passengers do not cancel requests.

(v) The pick-up passenger has only time window constraint on the boarding point and the deliver passenger has the drop-off point.

2.3. Notations. To facilitate model presentation, key parameter definitions and notations used subsequently are summarized in the notation list (Table 3).

2.4. Model Formulation. Given the above assumptions and definitions, the problem of deciding vehicle routes and scheduling for a HFRFT system can be formulated with the following MIP model which is similar to Wang et al. [61]. In the MIP model, the vehicle route determined by pick-up or deliver passengers is proposed (shown in Figure 3). This new route representation method can fully reflect the personality of the passenger time window and express accurately the route nonrepeated sections because each passenger can only pick up/deliver once. However, if the vehicle route is usually determined by demand points passed, there may be many demand points repeatedly on the routes and the routes are not easily distinguished.

In the MIP model, the optimization objective is to maximize system total benefit E, expressed as equation (1), which is the difference between operating income and operating costs. The former is fare income, E_1, of all passengers, expressed as equation (2), when the latter is made up of E_2, E_3, and E_4, expressed as equations (3)–(5), respectively:

$$E = w_1 E_1 - \left(w_2 E_2 + w_3 E_3 + w_4 E_4\right), \tag{1}$$

$$E_1 = \sum_v \sum_k \left|\mathbf{N}_{vk}\right| \cdot C, \tag{2}$$

$$E_2 = \sum_v \sum_k \left[C_v \cdot x_{vk} + \sum_m \sum_n L_{mn} \cdot x_{vkmn} \cdot G_v\right], \tag{3}$$

$$E_3 = \sum_v \sum_k \left[PE_m \cdot \left(D_{vk} + \sum_m x_{vkm} \cdot \theta_{vkm}\right)\right], \tag{4}$$

$$E_4 = \sum_v \sum_k \sum_m x_{vkm} \cdot PL_m, \tag{5}$$

where the RHS in equation (2) is fare income for all passengers when a one-ticket transit system is used. The RHS of equation (3) is vehicle operating cost made up of fixed use cost and travel cost of all shifts. The terms in the right hand side of equations (4) and (5) are the penalty cost of all shifts for early arrival and late arrival, respectively.

The proposed model is expressed as a MIP problem, as shown in the following equations:

Table 3: The notation list.

O	Transfer station
$\mathbf{N} = \{1,2,\ldots, a\}$	The set of passengers scheduled to go to the transfer point
$\mathbf{B} = \{a+1, a+2, \ldots, a+b\}$	The set of passengers scheduled to leave the transfer point
$\mathbf{NB} = \{1, 2, \ldots, a, a+1, a+b\}$	The set of all passengers
$m, n \in \mathbf{NB}$	The mth and nth passengers
V	The type of feeder vehicle
M, Mv	The set of feeder vehicles and the type v vehicle, respectively
$k \in \{1, 2, \ldots, K\}$	The kth shift, and each shift corresponds only to one vehicle type
K	The maximum shift
t_{vkm}	The time when the k th shift (belongs to type v vehicle) arrives at demand point of passenger m
UT_m, LT_m	Time window upper and lower bounds on the boarding point for pick-up passenger m or on the drop-off point for deliver passenger m
C_v, G_v	Fixed use cost per shift (\$/shift) and travel cost unit mileage (\$/mile) of type v, respectively
Q_v	The capacity of type v (people)
C	Ticket price (\$/people)
$\mathbf{N}_{vk}$	Passengers' set of the kth shift (belongs to type v vehicle)
L_{mn}, t_{mn}	Shortest path distance (miles) and time (hours) from demand point of passenger m to passenger n, respectively
V	Vehicle speed (miles/hour)
D_{vk}	The number of passengers in the feeder vehicle when the kth shift (belongs to type v vehicle) driving away from the transfer station
T	The maximum running time in one-way (hours)
T_{vk}	The departure time of the kth shift (belongs to type v vehicle)
DT_m	The reservation departure time from the transfer station of passenger m
x_{vkmn}	Binary variable $x_{vkmn} = 1$ if the kth shift (belongs to type v vehicle) is driving from demand point of passenger m to passenger n, otherwise 0
x_{vk}	Binary variable $x_{vkmn} = 1$ if the kth shift (belongs to type v vehicle) is used, otherwise 0
x_{vkm}	Binary variable $x_{vkm} = 1$ if the kth shift (belongs to type v vehicle) is passing the demand point of passenger m, otherwise 0
θ_{vkm}	Binary variable. $\theta_{vkm} = 1$ if passenger m is taking the kth shift (belongs to type v vehicle) to the transfer station, otherwise −1
x_{vkmn}	Binary variable. $x_{vkmn} = 1$ if the kth shift (belongs to type v vehicle) is driving from demand point of passenger m to passenger n, otherwise 0
PE_m, PL_m	Penalty function for early and late arrival at the demand point of passenger m, respectively
PE_{vkm}	Early arrival time of the kth shift (belongs to type v vehicle) at the demand point of passenger m
E_1	Fare income of all passenger(\$)
E_2	Vehicle operating cost of all shifts (\$)
E_3, E_4,	Penalty cost of all vehicles for early arrival and late arrival, respectively(\$)
α, β	Penalty parameters for early and late arrival, respectively
w_1, w_2, w_3, w_4,	Weight coefficients, respectively
ρ, ζ, ς	Connection degree, number of sections, number of nodes for road network
MP	Mutation probability in GA
cr, W_0, W_{end}, and CL	Cooling rate, initial temperature, end temperature, chain length in SA, respectively
Z	Iteration step in service area optimization
$z\zeta$	Service area of one block expanded in step ζ
S_ζ	Service area in step ζ
Sd_ζ	Expansion direction in step ζ
$\mathbf{SD}_\zeta$	Expansion direction set in step ζ

$$\text{Max E}, \tag{6}$$

$$\text{s.t. } D_{vk} + \sum_m x_{vkm} \cdot \theta_{vkm} \le Q_v, \quad \forall m \in \mathbf{NB}, \tag{7}$$

$$\mathbf{N}_{vk} \cap \mathbf{N}_{v_l k_l} = \varnothing, \quad \forall k \ne k' \text{ and } \forall v \ne v', \tag{8}$$

$$PE_m = \begin{cases} 0, & \text{if } T_m \le t_{vkm} \le UT_m, \\ \alpha\left(LT_m - t_{vkm}\right), & \text{if } t_{vkm} \le LT_m \text{ and } \theta_{vkm} = 1,\ m \in \mathbf{N}_{vk}, \\ 0, & \text{if } t_{vkm} \le LT_m \text{ and } \theta_{vkm} = -1, \end{cases} \tag{9}$$

$$PL_m = \begin{cases} 0, & \text{if } LT_m \le t_{vkm} \le UT_m, \\ \beta\left(t_{vkm} - UT_m\right), & \text{if } t_{vkm} > UT_m, \end{cases} \quad m \varepsilon \mathbf{N}_{vk}, \tag{10}$$

$$PE_{vkm} = \begin{cases} LT_m - t_{vkm}, & \text{if } \theta_{vkm} = 1, \\ 0, & \text{if } \theta_{vkm} = -1, \end{cases} \tag{11}$$

$$\sum_m \sum_n t_{mn} \cdot x_{vkmn} + \sum_m \max\left(PE_{vkm}, 0\right) \le T, \quad m, n \in \mathbf{N}_{vk}, \tag{12}$$

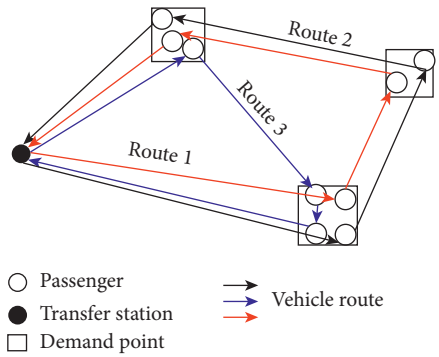


Figure 3: Vehicle route represented by passengers.

$$\sum_{n} x_{vkOn} = \sum_{m} x_{vkmO} = 1, \quad m, n \in \mathbf{NB}, \tag{13}$$

$$DT_m < T_{vk}, \quad \forall m \in \mathbf{N}_{vk}, \tag{14}$$

$$\sum_{v} \sum_{k} x_{vkm} = 1, \quad \forall m \in \mathbf{NB}. \tag{15}$$

In the model, equation (7) considers the vehicle capacity. Equation (8) determines each passenger can only be served by one vehicle. Equation (9) is the early arrival penalty cost only for pick-up passengers while the early arrival penalty cost is zero for deliver passengers because of vehicle waiting at the transfer station. Equations (10) and (11) are the late arrival penalty cost and the early arrival time, respectively. Equation (12) is the maximum run-time constraint. Equation (13) ensures vehicle depart from and return to the transfer station. Equation (14) ensures departure time is earlier than the reservation time for deliver passengers leaving the transfer station. Equation (15) ensures all reservation passengers will be served.

2.5. Solution Algorithm. The MIP model, constructed by equations (6)–(15), represents a more complicated vehicle routing problem (VRP) that is solved by heuristic algorithm, such as genetic algorithm (GA) [62–64], hyperheuristic algorithm [65], and hybrid of metaheuristic algorithm [66]. Ho et al. [67] argues that combining metaheuristics with other types of metaheuristics and mathematical programming is a growing trend to solve the DAR problem, and hybrids of metaheuristics usually come into two forms: (i) each metaheuristic is executed sequentially and (ii) a metaheuristic is executed within another metaheuristic. We adopt the second form, where GA is executed within simulated annealing (SA) to solve MIP model's abovementioned problem (as shown in Figure 4).

In Figure 4, firstly, based on GA, vehicle types and routes are optimized when departure intervals are known. Then, departure intervals are adjusted using the SA algorithm. GA is used again to optimize vehicle types and routes. This process is repeated until the ideal solution is obtained.

In the hybrid of metaheuristic, multichromosome encoding is used where both passengers and vehicles are encoded based on natural numbers, respectively. In initialization, r chromosomes are generated. Moreover, each chromosome contains a vehicle gene and a passenger gene. The code length of the former and the latter is $a + b$, and K, in this paper, is the number of reservation passengers and departure vehicles, respectively. When generating the initial populations, considering that the solution quality and efficiency of the heuristic algorithm depend on the quality of the initial solution, the generation of feasible solution itself is a NP-hard problem. Based on the initialization method proposed by Solomon [68], the improved nearest neighbor algorithm based on minimum cost (NNC algorithm) was designed to generate initial feasible populations. The algorithm flow is as follows:

Step 1: randomly set the departure time; the feeder bus starts from the transfer station (select the largest capacity vehicle)

Step 2: select an unserved passenger closest to the last served passenger; if the constraint conditions are satisfied, insert the passenger into the current running route

Step 3: repeat Step 2; if the capacity limit or maximum travel time limit of the current vehicle has been reached, go to Step 1 until all passengers have been served.

The "distance" (i.e., insertion cost) f_{mn} between two passengers in the algorithm flow is defined as follows:

$$f_{mn} = \alpha_1 t_{mn}^{vk} + \alpha_2 E_{mn}^{vk} + \alpha_3 L_{mn}^{vk} + \alpha_4 W_{mn}^{vk}, \tag{16}$$

where t_{mn}^{vk} denotes travel time of kth shift (belongs to type v vehicle); E_{mn}^{vk} is the difference between arrival time and start time, $E_{mn}^{vk} = |LT_n - t_{vkn}|$; L_{mn}^{vk} is the difference between the arrival time and the end time; $L_{mn}^{vk} = |UT_n - t_{vkn}|$; W_{mn}^{vk} is the sum of the travel time and the minimum possible waiting time; $W_{mn}^{vk} = t_{mn}^{vk} + \max\{0, (LT_n - t_{vkn})\}$; α_i denotes random parameter; and $\sum \alpha_i = 1$.

In GA, the fitness is the system total benefit; the selection method is roulette; the selection probability is proportional fitness; inner and interindividual crossovers are both adopted, respectively, for vehicle gene and passenger gene; the mutation probability is set to MP.

In SA, some parameters, such as cooling rate cr, initial temperature W_0, end temperature W_{end}, and chain length CL (= K in this paper), are determined; the code method is also a natural number; the new solution is based on the random disturbance and feasibility analysis; the Metropolis guideline is applied to determine whether the new solution is accepted.

3. Optimization Method of Service Area

The influence factors of the HFRFT service area include road network characteristics, passenger's demand distribution, and feeder facility performances. These factors together affect the shape and size.

Road network form is the important influence factor of service area. So far, the shape of service area is divided into

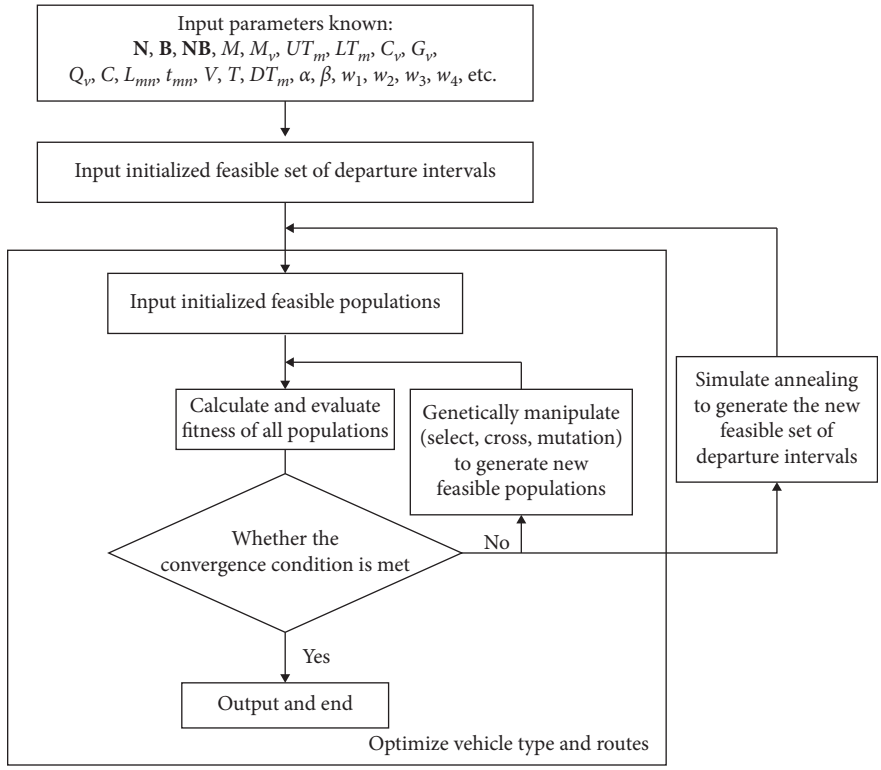


Figure 4: Flow chart of the hybrid of metaheuristic.

ideal regularity shape where road network form is assumed to be regular [8, 50] and actual irregularity shape [14, 15]. Moreover, the square road network form is suitable for the rectangular service areas while the radial network shape road network is suitable for the circular service areas. Other characteristics of the road network such as density and connectivity affect the size of service area to a certain extent. The bigger the road network density, the closer the feeder transit can be to the residential point. The higher the road network connectivity, the larger area the feeder system can serve because demand point can be reached more easily.

The temporal and spatial distribution characteristics of passenger demands are also the important factors. If other conditions are the same, the feeder transit in the low valley period can serve a larger area than the peak period; the feeder transit in low-density passenger area can serve larger scope than in the high-density passenger area.

The mode of route setting, the number and type of vehicles, and the running cycle of feeder facility performance are all important factors. The mode, in which the reference route exists, such as MAST, is generally used in the regular shape service area [51]. However, in HFRFT, whose vehicles have highly running freedom, the shape and size of the service area can be irregular and can be adjusted according to passenger demands. The more the pick-up and deliver capacity, determined by the number of vehicles and vehicle capacity, the larger the service area. At the same time, the longer the running cycle, the farther the feeder vehicle can travel, and the greater the service area can be.

The balance principle of capacity and demand is used to optimize service area. When road network and the number and type of vehicles are known, pick-up and deliver capacity can be obtained based on the MIP model. If demand distribution is also known in advance, the largest service area can be determined according to the balance principle. The optimization process is shown in Figure 5.

When determining the initial smaller service area S_0, S_0 should be included by the potential service area of HFRFT, determined by the reasonable attraction radius [69]. S_0 should be closed and be much smaller than the potential service area. If the MIP model is not solvable, the current system capacity cannot meet the system demand, that is, the system supply and system demand cannot be balanced.

In Figure 5, the expansion direction of service area is selected as follows:

(i) The expansion direction should have reservation requirements

(ii) After expanding a block, the service area will still be a closed area surrounded by roads

(iii) When there are multiple directions that can be expanded, the direction of the most demand points increased is preferentially chosen (other choice criteria such as the most passengers increased can also be used)

(iv) If there are multiple directions where demand points increased are the same, the direction of the largest connection degree ρ for road network after expansion is preferentially chosen, where

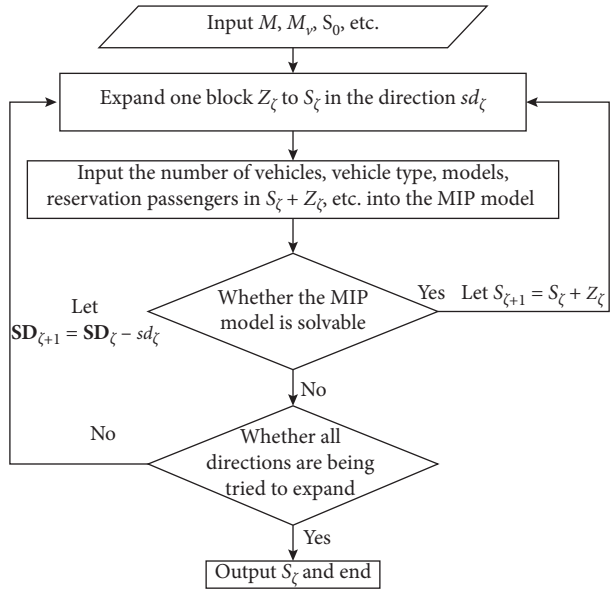


Figure 5: The optimization process of service area.

$$\rho = \frac{2\zeta}{\varsigma}. \tag{17}$$

4. Case Study

Shang Shuangtang Station of Changsha Metro Line 1 in Changsha City is selected as the sample. The models developed in this paper are used to obtain the optimal HFRFT system with an irregular area and road network. In the case study, the MIP model to coordinately optimize routes and scheduling is applied for the HFRFT system with the mixed running mode, and the optimization results are compared with the alone pick-up and alone deliver mode.

4.1. Basic Conditions. The scheduling cycle is 7 : 00–8 : 00 a.m.; the radius of the circular area is 2.5 miles; the uniform speed V is 21.7 miles/h; the coordinate of the transfer station O is (0, 0); there are two types of vehicles, $\mathbf{M}_1 = \{M_{11}, M_{12}, M_{13}, M_{14}, M_{15}\}$, $\mathbf{M}_2 = \{M_{21}, M_{22}, M_{23}\}$, $Q_1 = 10$ people, $Q_2 = 15$ people, $C_1 = 9.3$ \$/shift, $Q_2 = 15.5$ \$/shift, $G_1 = 1.9$ \$/mile, $G_2 = 2.4$ \$/mile; T is 45 min; C is 2.5 \$/people; w_1, w_2, w_3, and w_4 are 1.0, 1.0, 2/3, and 2/3 respectively; α, β are 9.3 \$/h and 13 \$/h, respectively; there are 1 transfer station and 25 demand points (shown in Figure 6 and Table 4) and 124 pick-up passengers to transfer station and 36 deliver passengers leaving the transfer station, whose time windows and locations are shown in Table 5. The distances between demand points are shown in Table 6.

4.2. Route and Scheduling Optimization in Initial Service Area. According to the land use, the light green dotted area, surrounded by Wanjiali Rd.(S.), Furong Rd.(S.), Huanbao Rd.(W.), and Huijin Rd., is set as the initial service area (shown in Figure 6). In the initial service area, there are 59 reservation passengers. The computer program of the Figure 5 based on Matlab is operated to optimize routes and scheduling, and the results are shown in Tables 7–9.

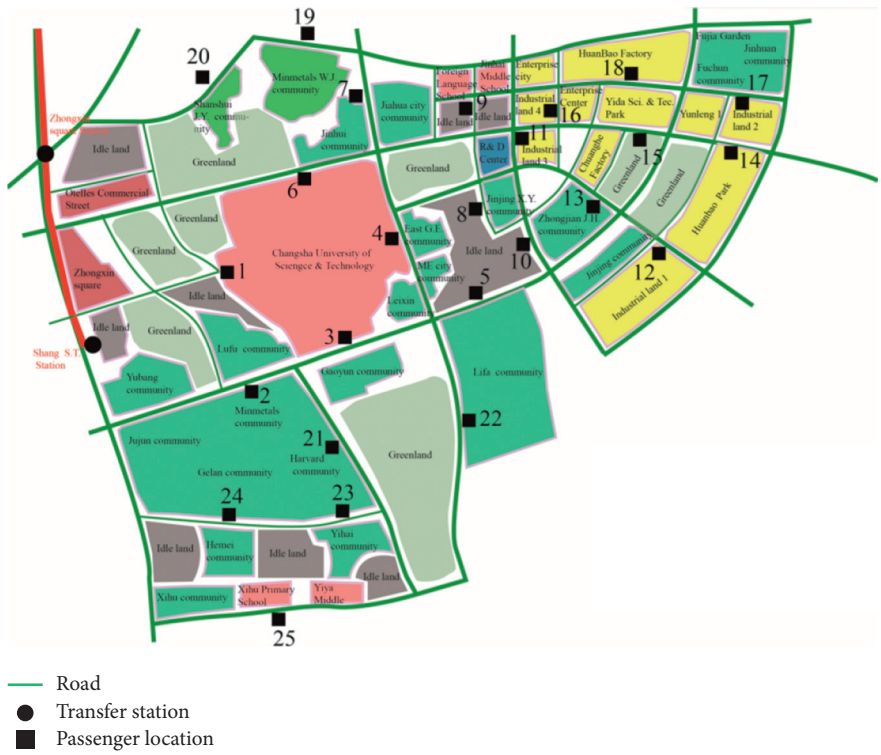
If the number of the vehicles is sufficient and other parameters in the initial service area are still used, routes and scheduling optimization under pick-up and deliver alone is performed based on the MIP model changed slightly constraints for the mixed running. The results are shown in Tables 8 and 9.

From Tables 7–9, comparing to the pick-up and deliver mode alone, the advantage of the mixed mode is obvious: the number of departures and vehicles required are reduced by 16.7% and 20%, respectively. Also, average seat utilization rate is increased by 8.3%. System total benefit E becomes positive.

4.3. Route and Scheduling Optimization in Optimal Service Area. In Table 7, only 4 vehicles are needed, so the service area can be expanded. There are three expandable directions (shown in Figure 7). The number of new demand points is the same when expanding a block in three directions, so road network connectivity in three directions should be compared. Road network connectivity that expanded to the north, to the east, to the south is 2.72, 2.61, and 2.66, respectively. Priority should be given to expanding the service area to the north. Loop this way until the optimal service area is obtained (dark green dotted line as shown in Figure 7).

In the service area surrounded by the dark green dotted line, all vehicles are used (according to the calculations followed); the number of departures is 9, and the number of passengers served is 105. Based on Figure 5, the service area under irregular road network and scope can be optimized.

There are 105 reservation passengers in the optimal service area. The computer program of Figure 4 is again utilized, and the results are shown in Tables 10–12.



Road
Transfer station
Passenger location

Figure 6: Possible service area and demand point distribution.

If the number of the vehicles is sufficient and the other parameters in the optimal service area are still used, routes and scheduling optimization under pick-up and deliver alone is performed based on the MIP model. The results are shown in Tables 11 and 12.

From Tables 10 and 12, comparing to the pick-up and deliver alone modes, the advantage of the mixed mode is also obvious: the number of departures, vehicles required, and vehicle mileage required per passenger is reduced by 10.0%, 11.1%, and 8%, respectively; average seat utilization rate is increased by 9.5%; system total benefit E is increased by more than 3 times. Compared with Tables 9 and 12, the number of passengers served by HFRFT system is increased by optimizing the service area. Therefore, regardless of whether the RFT system adopts the mixed running mode or alone running mode, the average seat utilization rate and the total system benefit E are improved significantly (as shown in Figure 8). Under the condition of certain transportation capacity, the optimization of service area has certain significance to improve vehicle utilization rate and reduce vehicle idle cost.

Table 4: Coordinates of demand points (miles).

Point	X	Y
0	0.00	0.00
1	0.52	0.33
2	0.66	−0.16
3	1.20	0.01
4	1.41	0.45
5	1.72	0.17
6	0.92	0.75
7	1.26	1.00
8	1.77	0.60
9	1.72	1.04
10	1.98	0.43
11	1.88	0.89
12	2.58	0.41
13	2.25	0.63
14	2.79	0.86
15	2.39	0.93
16	2.08	1.04
17	2.85	1.08
18	2.43	1.12
19	0.99	1.30
20	0.60	1.13
21	1.13	−0.37
22	1.62	−0.22
23	1.13	−0.76
24	0.61	−0.76
25	0.69	−1.44

5. Conclusion and Future Work

This paper proposes a design method for the HFRFT system with multitype vehicles, mixed running, and capacity constraints, where the vehicle routes, scheduling, and service area are optimized according to reservation requirements. A mixed integer programming model and its hybrid of metaheuristic algorithm is devised to efficiently and integrally solve for vehicle routes and scheduling parameters. A heuristic method is also put forward to optimize the service area based on the equilibrium of system supply and demand.

Table 5: Passenger locations and time windows.

Passenger number	Passenger type	Demand point	LT_m	UT_m	DT_m
1	Pick-up	1	7:02	7:07	
2	Pick-up	1	7:03	7:08	
3	Pick-up	1	7:05	7:10	
4	Pick-up	1	7:07	7:12	
5	Pick-up	1	7:10	7:15	
6	Pick-up	1	7:13	7:18	
7	Pick-up	1	7:15	7:20	
8	Pick-up	1	7:18	7:23	
9	Pick-up	1	7:20	7:25	
10	Pick-up	1	7:22	7:27	
11	Pick-up	1	7:24	7:29	
12	Pick-up	1	7:27	7:32	
13	Deliver	1	7:30	7:35	7:08
14	Deliver	1	7:40	7:45	7:12
15	Deliver	1	7:45	7:50	7:18
16	Pick-up	1	7:35	7:40	
17	Pick-up	1	7:40	7:45	
18	Pick-up	2	7:07	7:12	
19	Pick-up	2	7:10	7:15	
20	Pick-up	2	7:20	7:25	
21	Pick-up	2	7:27	7:32	
22	Deliver	2	7:33	7:38	7:08
23	Deliver	2	7:40	7:45	7:10
24	Pick-up	3	7:09	7:14	
25	Pick-up	3	7:12	7:17	
26	Pick-up	3	7:18	7:23	
27	Pick-up	3	7:20	7:25	
28	Pick-up	3	7:23	7:28	
29	Pick-up	3	7:26	7:31	
30	Pick-up	3	7:28	7:33	
31	Pick-up	3	7:30	7:35	
32	Pick-up	3	7:35	7:40	
33	Pick-up	3	7:38	7:42	
34	Pick-up	3	7:40	7:45	
35	Pick-up	3	7:42	7:47	
36	Deliver	3	7:30	7:35	7:05
37	Deliver	3	7:40	7:42	7:08
38	Pick-up	4	7:08	7:13	
39	Pick-up	4	7:10	7:15	
40	Pick-up	4	7:13	7:18	
41	Pick-up	4	7:15	7:20	
42	Pick-up	4	7:18	7:23	
43	Pick-up	4	7:23	7:28	
44	Pick-up	4	7:25	7:30	
45	Pick-up	4	7:27	7:32	
46	Pick-up	4	7:30	7:35	
47	Pick-up	4	7:32	7:37	
48	Pick-up	4	7:35	7:40	
49	Pick-up	4	7:37	7:42	
50	Pick-up	4	7:38	7:43	
51	Pick-up	4	7:40	7:45	
52	Deliver	4	7:45	7:50	7:12
53	Pick-up	5	7:10	7:15	

Table 5: Continued.

Passenger number	Passenger type	Demand point	LT_m	UT_m	DT_m
54	Pick-up	5	7:12	7:17	
55	Pick-up	5	7:15	7:20	
56	Pick-up	5	7:23	7:28	
57	Pick-up	5	7:25	7:30	
58	Pick-up	5	7:27	7:32	
59	Pick-up	5	7:28	7:33	
60	Pick-up	5	7:30	7:35	
61	Pick-up	5	7:32	7:37	
62	Pick-up	5	7:35	7:40	
63	Pick-up	5	7:37	7:42	
64	Deliver	5	7:35	7:40	7:05
65	Deliver	5	7:38	7:42	7:05
66	Deliver	5	7:42	7:47	7:12
67	Deliver	5	7:50	7:55	7:15
68	Pick-up	6	7:15	7:20	
69	Pick-up	6	7:17	7:22	
70	Pick-up	6	7:20	7:25	
71	Pick-up	6	7:23	7:28	
72	Pick-up	6	7:28	7:33	
73	Deliver	6	7:30	7:35	7:00
74	Deliver	6	7:35	7:40	7:00
75	Pick-up	7	7:15	7:20	
76	Pick-up	7	7:20	7:25	
77	Deliver	7	7:35	7:40	7:00
78	Deliver	7	7:40	7:45	7:05
79	Pick-up	8	7:20	7:25	
80	Pick-up	8	7:23	7:28	
81	Pick-up	8	7:25	7:30	
82	Pick-up	9	7:05	7:10	
83	Pick-up	9	7:10	7:15	
84	Pick-up	9	7:15	7:20	
85	Pick-up	10	7:15	7:20	
86	Pick-up	10	7:18	7:23	
87	Pick-up	10	7:20	7:25	
88	Deliver	10	7:25	7:30	7:00
89	Deliver	11	7:45	7:50	7:10
90	Deliver	11	7:45	7:50	7:10
91	Pick-up	12	7:10	7:15	
92	Pick-up	12	7:10	7:15	
93	Deliver	12	7:35	7:40	7:05
94	Deliver	12	7:35	7:40	7:05
95	Pick-up	13	7:30	7:35	
96	Pick-up	13	7:30	7:35	
97	Deliver	13	7:35	7:40	7:05
98	Pick-up	14	7:10	7:15	
99	Pick-up	14	7:15	7:20	
100	Deliver	14	7:35	7:40	7:00
101	Deliver	14	7:35	7:40	7:00
102	Deliver	15	7:40	7:45	7:05
103	Deliver	15	7:40	7:45	7:05
104	Pick-up	16	7:20	7:25	
105	Pick-up	16	7:20	7:25	
106	Pick-up	17	7:30	7:35	

Table 5: Continued.

Passenger number	Passenger type	Demand point	LT_m	UT_m	DT_m
107	Pick-up	17	7:35	7:40	
108	Deliver	18	7:35	7:40	7:00
109	Deliver	18	7:40	7:45	7:10
110	Deliver	18	7:40	7:45	7:15
111	Pick-up	19	7:10	7:15	
112	Pick-up	19	7:15	7:20	
113	Pick-up	19	7:20	7:25	
114	Pick-up	19	7:23	7:28	
115	Pick-up	19	7:30	7:35	
116	Pick-up	19	7:40	7:45	
117	Pick-up	20	7:12	7:17	
118	Pick-up	20	7:15	7:20	
119	Pick-up	20	7:25	7:30	
120	Pick-up	20	7:27	7:32	
121	Pick-up	20	7:35	7:40	
122	Pick-up	20	7:40	7:45	
123	Pick-up	21	7:12	7:17	
124	Pick-up	21	7:15	7:20	
125	Pick-up	21	7:20	7:25	
126	Pick-up	21	7:23	7:28	
127	Pick-up	21	7:25	7:30	
128	Pick-up	21	7:27	7:32	
129	Pick-up	21	7:30	7:35	
130	Pick-up	21	7:35	7:40	
131	Pick-up	21	7:38	7:43	
132	Deliver	21	7:35	7:40	7:05
133	Deliver	21	7:40	7:45	7:08
134	Pick-up	22	7:18	7:23	
135	Pick-up	22	7:18	7:23	
136	Pick-up	22	7:24	7:29	
137	Pick-up	22	7:24	7:29	
138	Pick-up	22	7:30	7:35	
139	Pick-up	23	7:13	7:18	
140	Pick-up	23	7:15	7:20	
141	Pick-up	23	7:20	7:25	
142	Pick-up	23	7:25	7:30	
143	Pick-up	23	7:40	7:45	
144	Deliver	23	7:35	7:40	7:05
145	Deliver	23	7:40	7:45	7:12
146	Pick-up	24	7:10	7:15	
147	Pick-up	24	7:13	7:18	
148	Pick-up	24	7:15	7:20	
149	Pick-up	24	7:18	7:23	
150	Pick-up	24	7:20	7:25	
151	Pick-up	24	7:23	7:28	
152	Pick-up	24	7:25	7:30	
153	Pick-up	24	7:28	7:32	
154	Pick-up	24	7:30	7:35	
155	Deliver	24	7:35	7:40	7:00
156	Deliver	24	7:42	7:45	7:08
157	Deliver	24	7:45	7:50	7:12
158	Pick-up	25	7:10	7:15	
159	Pick-up	25	7:12	7:17	
160	Pick-up	25	7:10	7:15	

Table 6: Distances between demand points (mile).

Point	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
0	0.00	0.69	0.87	1.42	2.11	1.98	1.57	2.11	2.60	2.50	2.42	2.57	2.98	2.96	3.42	3.04	2.83	3.63	3.24	2.22	1.74	1.54	2.15	1.63	1.12	2.08
1	0.69	0.00	0.74	1.30	1.69	1.82	0.94	1.48	2.05	1.87	2.03	1.93	2.81	2.68	2.79	2.41	2.20	3.00	2.60	1.59	1.12	1.37	1.98	2.06	2.06	3.02
2	0.87	0.74	0.00	0.55	0.74	1.08	1.68	1.78	1.87	2.17	1.55	2.04	2.07	1.94	2.44	2.31	2.23	2.65	2.50	2.26	1.85	0.63	1.23	1.19	1.31	2.28
3	1.42	1.30	0.55	0.00	0.68	0.53	1.44	1.23	1.31	1.53	1.00	1.46	1.52	1.39	1.89	1.75	1.68	2.10	1.95	1.71	2.18	0.59	0.68	1.15	1.66	2.34
4	2.11	1.69	0.74	0.68	0.00	0.54	0.75	0.55	0.75	0.84	1.01	0.81	1.45	1.08	1.53	1.31	1.11	1.90	1.51	1.02	1.49	1.28	0.70	1.84	2.35	2.62
5	1.98	1.82	1.08	0.53	0.54	0.00	1.29	1.09	0.86	1.29	0.48	0.93	1.00	0.86	1.36	1.31	1.26	1.57	1.43	1.56	2.03	1.12	0.54	1.68	2.19	2.46
6	1.57	0.94	1.68	1.44	0.75	1.29	0.00	0.55	1.11	0.94	1.46	1.00	1.90	1.53	1.85	1.48	1.26	2.05	1.64	1.03	1.50	2.03	1.45	2.59	2.68	3.37
7	2.11	1.48	1.78	1.23	0.55	1.09	0.55	0.00	0.91	0.73	1.25	0.79	1.69	1.32	1.64	1.26	1.05	1.87	1.46	0.48	0.95	1.82	1.24	2.38	2.89	3.16
8	2.60	2.05	1.87	1.31	0.75	0.86	1.11	0.91	0.00	0.43	0.38	0.39	1.04	0.66	1.28	0.91	0.76	1.51	1.10	1.32	1.79	1.90	1.33	2.47	2.98	3.24
9	2.50	1.87	2.17	1.53	0.84	1.29	0.94	0.73	0.43	0.00	0.82	0.32	1.22	0.85	1.18	0.79	0.46	1.14	0.73	0.89	1.36	2.12	1.54	2.68	3.19	3.47
10	2.42	2.03	1.55	1.00	1.01	0.48	1.46	1.25	0.38	0.82	0.00	0.46	0.78	0.52	1.14	0.76	0.68	1.35	0.95	1.66	2.13	1.59	1.01	2.15	2.67	2.93
11	2.57	1.93	2.04	1.46	0.81	0.93	1.00	0.79	0.39	0.32	0.46	0.00	0.90	0.53	0.92	0.55	0.33	1.13	0.71	1.20	1.67	2.05	1.69	2.61	3.12	3.39
12	2.98	2.81	2.07	1.52	1.45	1.00	1.90	1.69	1.04	1.22	0.78	0.90	0.00	0.37	0.60	0.82	0.87	0.81	1.16	2.34	2.81	2.11	1.54	2.67	3.19	3.45
13	2.96	2.68	1.94	1.39	1.08	0.86	1.53	1.32	0.66	0.85	0.52	0.53	0.37	0.00	0.74	0.55	0.43	0.95	0.74	1.77	2.24	1.98	1.40	2.54	3.05	3.32
14	3.42	2.79	2.44	1.89	1.53	1.36	1.85	1.64	1.28	1.18	1.14	0.92	0.60	0.74	0.00	0.38	0.78	0.28	0.63	2.06	2.53	2.45	1.87	3.01	3.52	3.79
15	3.04	2.41	2.31	1.75	1.31	1.31	1.48	1.26	0.91	0.79	0.76	0.55	0.82	0.55	0.38	0.00	0.40	0.59	0.51	1.68	2.15	2.42	1.85	2.99	3.50	3.77
16	2.83	2.20	2.23	1.68	1.11	1.26	1.26	1.05	0.76	0.46	0.68	0.33	0.87	0.43	0.78	0.40	0.00	0.83	0.42	1.28	1.75	2.28	1.70	2.84	3.35	3.62
17	3.63	3.00	2.65	2.10	1.90	1.57	2.05	1.87	1.51	1.14	1.35	1.13	0.81	0.95	0.28	0.59	0.83	0.00	0.41	1.97	2.44	2.69	2.11	3.25	3.76	4.03
18	3.24	2.60	2.50	1.95	1.51	1.43	1.64	1.46	1.10	0.73	0.95	0.71	1.16	0.74	0.63	0.51	0.42	0.41	0.00	1.55	2.02	2.30	1.72	2.86	3.34	3.64
19	2.22	1.59	2.26	1.71	1.02	1.56	1.03	0.48	1.32	0.89	1.66	1.20	2.34	1.77	2.06	1.68	1.28	1.97	1.55	0.00	0.47	2.54	2.19	3.33	2.86	3.83
20	1.74	1.12	1.85	2.18	1.49	2.03	1.50	0.95	1.79	1.36	2.13	1.67	2.81	2.24	2.53	2.15	1.75	2.44	2.02	0.47	0.00	2.49	2.19	3.04	2.72	3.68
21	1.54	1.37	0.63	0.59	1.28	1.12	2.03	1.82	1.90	2.12	1.59	2.05	2.11	1.98	2.45	2.42	2.28	2.69	2.30	2.54	2.49	0.00	1.28	0.56	1.07	1.75
22	2.15	1.98	1.23	0.68	0.70	0.54	1.45	1.24	1.33	1.54	1.01	1.69	1.54	1.40	1.87	1.85	1.70	2.11	1.72	2.19	2.19	1.28	0.00	1.77	2.29	1.92
23	1.63	2.06	1.19	1.15	1.84	1.68	2.59	2.38	2.47	2.68	2.15	2.61	2.67	2.54	3.01	2.99	2.84	3.25	2.86	3.33	3.04	0.56	1.77	0.00	0.51	1.55
24	1.12	2.06	1.31	1.66	2.35	2.19	2.68	2.89	2.98	3.19	2.67	3.12	3.19	3.05	3.52	3.50	3.35	3.76	3.34	2.86	2.72	1.07	2.29	0.51	0.00	1.64
25	2.08	3.02	2.28	2.34	2.62	2.46	3.37	3.16	3.24	3.47	2.93	3.39	3.45	3.32	3.79	3.77	3.62	4.03	3.64	3.83	3.68	1.75	1.92	1.55	1.64	0.00

Table 7: Optimization results under mixed running mode in the initial service area.

Departure time	Vehicle number	Route	Mileage (miles)	Cumulative number of passengers serviced	Seat utilization	End time	Running mode
7:13	M_{21}	0-68-69-1-3-6-9-25-27-38-39-40-43-29-45-19-0	7.45	15	100	7:33	Alone pick-up
7:14	M_{11}	0-26-18-5-70-41-44-42-24-46-21-0	7.37	10	100	7:36	Alone pick-up
7:30	M_{22}	0-13-11-10-71-72-73-47-30-32-48-49-51-35-0	5.85	13	86.6	7:47	Mixed
7:35	M_{12}	0-8-4-7-20-22-34-28-31-33-36-50-74-0	4.99	12	120	7:49	Mixed
7:45	M_{11}	0-23-52-37-12-14-2-15-16-17-0	4.28	9	90	7:57	Mixed

Table 8: Optimization results under pick-up and deliver alone in the initial service area.

Departure time	Vehicle number	Route	Mileage (miles)	Cumulative number of passengers serviced	Seat utilization	End time	Running mode
7:03	M_{21}	0-3-9-20-24-25-26-28-39-42-44-19-0	4.28	11	73.3	7:30	Alone pick-up
7:13	M_{11}	0-1-2-4-5-6-7-53-40-38-18-0	3.94	10	100	7:24	Alone pick-up
7:18	M_{12}	0-58-59-13-14-15-22-23-36-37-52-0	6.58	10	100	7:37	Alone deliver
7:18	M_{13}	0-54-56-8-10-11-21-41-43-45-27-0	6.09	10	100	7:36	Alone pick-up
7:28	M_{14}	0-57-46-47-48-55-12-16-17-0	4.70	8	80	7:41	Alone pick-up
7:36	M_{11}	0-29-30-31-32-34-49-50-51-33-35-0	4.21	10	100	7:48	Alone pick-up

Table 9: Comparisons of two running modes in the initial service area.

Running mode	K	Number of vehicles required			Average seat utilization rate	Vehicle mileage required per passenger(miles)	E ($)
		$v=1$	$v=2$	Sum			
Alone	6	4	1	5	90.76%	0.51	−59.0
Mixed	5	2	2	4	98.33%	0.51	14.5
Increase rate(%)	−16.7	—	—	−20	8.3	—	—

Case studies show that the mixed running mode with multitype vehicle can significantly increase the seat utilization and consequentially reduce the vehicle ownership and average vehicle travel distance per passenger significantly. From the case studies, the following conclusions are obtained:

(i) The MIP model can optimize the running route and departure time of each shift under reservation requirements with the mixed running, multitype vehicles, and capacity constraints and also optimize the number of departures and vehicle possession.

(ii) In the same situation of reservation requirements, the mixed running mode can significantly improve the seat utilization rate compared with that of the pick-up and deliver alone modes. It can also significantly reduce the total system cost, number of vehicle ownership and departures, and vehicle running mileage required per passenger. The results indicated that the mixed running mode is superior.

(iii) The proposed service area heuristic optimization method based on the MIP model for HFRFT running routes and scheduling coordination model can determine an optimal service area that matches the capacity with the demand.

The optimization design method can provide certain practical guidance for the transit operators to determine the operation strategy and also contribute to the promotion and implementation of the RFT technology. In this study, we focus on reservation requirements. In real world, passengers may make a pick-up or deliver request during the running of

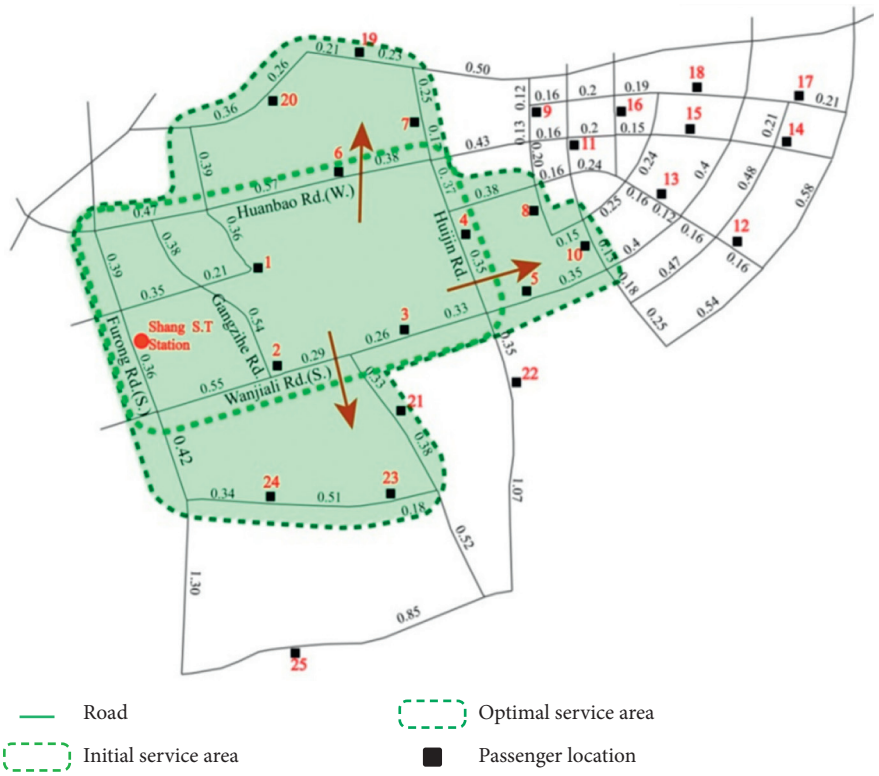


Figure 7: Service area optimization process.

Table 10: Optimization results under mixed running in the optimal service area.

Departure time	Vehicle number	Route	Mileage (miles)	Cumulative number of passengers serviced	Seat utilization	End time	Running mode
7:09	M_{11}	0-111-112-76-68-70-41-42-28-124-127-0	6.81	10	100	7:32	Alone pick-up
7:10	M_{21}	0-139-140-123-39-20-4-5-128-119-143-141-0	13.49	11	73.3	7:48	Alone pick-up
7:14	M_{22}	0-113-7-118-75-27-29-12-116-0	12.17	8	53.3	7:47	Alone deliver
7:16	M_{23}	0-2-3-8-10-146-148-126-26-31-121-16-0	8.41	11	73.3	7:43	Alone pick-up
7:22	M_{12}	0-1-6-9-11-18-24-25-142-151-153-0	4.77	10	100	7:35	Alone pick-up
7:30	M_{13}	0-13-19-38-40-43-44-46-114-115-69-71-0	5.80	11	110	7:46	Mixed
7:30	M_{14}	0-30-36-129-45-47-48-49-50-32-33-34-0	5.39	11	110	7:45	Mixed
7:40	M_{15}	0-72-73-74-77-78-117-120-122-51-130-131-145-147-150-154-0	8.03	15	150	8:02	Mixed
7:45	M_{13}	0-14-15-125-132-133-144-152-155-156-157-149-35-37-52-21-22-23-17-0	7.66	18	120	8:07	Mixed

Table 11: Optimization results under pick-up and deliver alone in the optimal service area.

Departure time	Vehicle number	Route	Mileage (miles)	Cumulative number of passengers serviced	Seat utilization (%)	End time	Running mode
7 : 07	M_{11}	0-19-41-38-39-24-25-95-76-28-21-0	7.04	10	100	7 : 31	Alone pick-up
7 : 10	M_{12}	0-71-40-77-94-97-99-44-45-46-0	10.06	9	90	7 : 38	Alone pick-up
7 : 10	M_{21}	0-53-60-20-96-98-88-78-90-80-79-101-89-102	10.61	13	86.6	7 : 40	Alone deliver
7 : 12	M_{13}	0-13-14-85-86-92-93-103-104-105-0	4.26	9	90	7 : 24	Alone deliver
7 : 16	M_{22}	0-64-65-66-67-61-54-55-56-2-3-4-6-12-0	4.44	13	86.6	7 : 30	Alone pick-up
7 : 18	M_{14}	0-58-59-36-37-22-23-52-62-63-15-0	7.02	10	100	7 : 38	Alone deliver
7 : 18	M_{15}	0-5-7-8-9-26-27-87-70-72-74-0	7.95	10	100	7 : 41	Alone pick-up
7 : 20	M_{23}	0-1-10-42-73-84-34-43-47-49-91-100-0	11.10	11	73.3	7 : 52	Alone pick-up
7 : 22	M_{16}	0-11-18-29-57-68-16-75-69-50-51-0	10.75	10	100	7 : 52	Alone pick-up
7 : 30	M_{13}	0-81-82-83-30-31-32-33-48-35-17-0	5.49	10	100	7 : 48	Alone pick-up

Table 12: Comparisons of two running modes in the optimal service area.

Running mode	K	Number of vehicles required			Average seat utilization rate	Vehicle mileage required per passenger(miles)	E ($)
		$v = 1$	$v = 2$	Sum			
Alone	10	6	3	9	91.3%	0.75	23.1
Mixed	9	5	3	8	100%	0.69	119.3
Increase rate (%)	−10.0	—	—	−11.1	9.5	−8.0	416

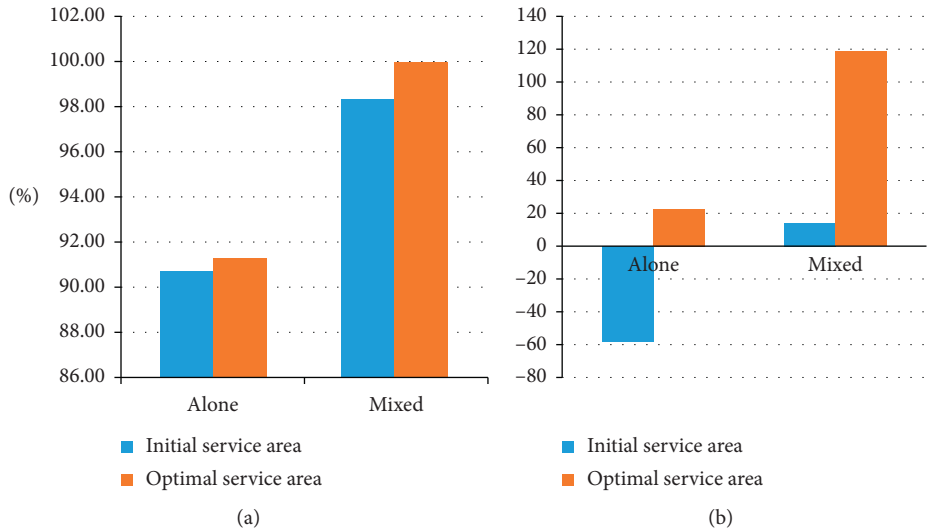


Figure 8: Comparisons of two operation modes before and after optimization. (a) Comparisons of average seat utilization. (b) Comparisons of number of vehicles required.

the vehicle, which means it will be more practical to consider the real-time requirements. In addition, the travel time of each road section is not a fixed value, it changes with the different traffic flow. Therefore, future research will focus on designing HFRFT system under the mixed demand (with reservation and real-time requirements at the same time) and dynamic road network.

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 paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (No. 51678075). In addition, this work was also supported by the Hunan Key Area R&D Program (2019SK2171) and Hunan Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems (2017TP1016). In addition, this work was also supported by the Hunan Provincial Natural Science Foundation of China (No. 2019JJ30026), the Young Elite Scientists Sponsorship Program by Hunan Province of China (2018RS3074), the Young Elite Scientists Sponsorship Program by Hunan Provincial Department of Education (18B142), Changsha Science and Technology Bureau Project (kq1801056), Open Fund of Hunan Key Laboratory of Smart Roadway and Cooperative Vehicle-Infrastructure Systems (kfj180702; kfj190702; kfj190704; and 180402), Innovation Team Project for Transportation Engineering in CSUST, CSUST International Cooperation Project (2019IC11), and Ministry of Science Project with Montenegro (3-2).

References

[1] T. Liu and A. Ceder, "Analysis of a new public-transport-service concept: customized bus in China," *Transport Policy*, vol. 39, pp. 63–76, 2015.

[2] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment," *IEEE Access*, vol. 6, pp. 52471–52478, 2018.

[3] C. Ma, W. Hao, W. Xiang, and W. Yan, "The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents," *Journal of Advanced Transportation*, vol. 2018, Article ID 9841498, 10 pages, 2018.

[4] Y. Li, W. Chen, S. Peeta, and Y. Wang, "Platoon control of connected multi-vehicle systems under V2X communications: design and experiments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1891–1902, 2020.

[5] L. Hu, X. Hu, J. Wan, M. Lin, and J. Huang, "The injury epidemiology of adult riders in vehicle-two-wheeler crashes in China, Ningbo, 2011–2015," *Journal of Safety Research*, vol. 72, pp. 21–28, 2020.

[6] L. Hu, J. Ou, J. Huang, Y. Chen, and D. Cao, "A review of research on traffic conflicts based on intelligent vehicles," *IEEE Access*, vol. 8, pp. 24471–24483, 2020.

[7] K.Y. Ma, S. Chen, A. J. Khattak, and Y. Pan, "On-line aggressive driving identification based on in-vehicle kinematic parameters under naturalistic driving conditions," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 554–571, 2020.

[8] X. Li and L. Quadrifoglio, "Feeder transit services: choosing between fixed and demand responsive policy," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 770–780, 2010.

[9] Y. S. Sun, Q. W. Guo, P. Schonfeld, and Z. F. Li, "Evolution of public transit modes in a commuter corridor," *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 84–102, 2017.

[10] Q.-W. Guo, J. Y. J. Chow, and P. Schonfeld, "Stochastic dynamic switching in fixed and flexible transit services as market entry-exit real options," *Transportation Research Procedia*, vol. 23, pp. 380–399, 2017.

[11] M. Saberi and İ. Ö. Verbas, "Continuous approximation model for the vehicle routing problem for emissions minimization at the strategic level," *Journal of Transportation Engineering*, vol. 138, no. 11, pp. 1368–1376, 2012.

[12] Y. Yan, Z. Liu, Q. Meng, and Y. Jiang, "Robust optimization model of bus transit network design with stochastic travel time," *Journal of Transportation Engineering*, vol. 139, no. 6, pp. 625–634, 2013.

[13] Y. Yao, B. M. Randy, and X. Chi, "Demand-responsive transit circulator service network design," *Transportation Research Part E*, vol. 76, pp. 160–175, 2015.

[14] M. Dessouky, M. Rahimi, and M. Weidner, "Jointly optimizing cost, service, and environmental performance in demand-responsive transit scheduling," *Transportation Research Part D: Transport and Environment*, vol. 8, no. 6, pp. 433–465, 2003.

[15] S. Pan, J. Yu, X. Yang et al., "Designing a flexible feeder transit system serving irregularly shaped and gated communities: determining service area and feeder route planning," *Journal of Urban Planning and Development*, vol. 141, no. 3, Article ID 04014028, 2015.

[16] A. Núñez, C. E. Cortés, D. Sáez, B. De Schutter, and M. Gendreau, "Multiobjective model predictive control for dynamic pickup and delivery problems," *Control Engineering Practice*, vol. 32, pp. 73–86, 2014.

[17] S. F. Ghannadpour, S. Noori, R. Tavakkoli-Moghaddam, and K. Ghoseiri, "A multi-objective dynamic vehicle routing problem with fuzzy time windows: model, solution and application," *Applied Soft Computing*, vol. 14, no. 1, pp. 504–527, 2014.

[18] M. Schilde, K. F. Doerner, and R. F. Hartl, "Integrating stochastic time-dependent travel speed in solution methods for the dynamic dial-a-ride problem," *European Journal of Operational Research*, vol. 238, no. 1, pp. 18–30, 2014.

[19] D. Kirchler and R. Wolfler Calvo, "A granular tabu search algorithm for the dial-a-ride problem," *Transportation Research Part B: Methodological*, vol. 56, no. 10, pp. 120–135, 2013.

[20] B. Sun, M. Wei, and S. L. Zhu, "Optimal design of demand-responsive feeder transit services with passengers' multiple time windows and satisfaction," *Future Internet*, vol. 30, pp. 1–15, 2018.

[21] J. Tang, Y. Wang, W. Hao, F. Liu, H. Huang, and Y. Wang, "A mixed path size logit-based taxi customer-search model considering spatio-temporal factors in route choice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1347–1358, 2019.

[22] M. M. Aldaihani, L. Quadrifoglio, M. M. Dessouky, and R. Hall, "Network design for a grid hybrid transit service," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 7, pp. 511–530, 2004.

[23] J.-B. Sheu, "A fuzzy clustering approach to real-time demand-responsive bus dispatching control," *Fuzzy Sets and Systems*, vol. 150, no. 3, pp. 437–455, 2005.

[24] M. Rahimi, M. Amirgholy, and E. J. Gonzales, "System modeling of demand responsive transportation services: evaluating cost efficiency of service and coordinated taxi usage," *Transportation Research Part E: Logistics and Transportation Review*, vol. 112, pp. 66–83, 2018.

[25] P. Chen and Y. Nie, "Optimal design of demand adaptive paired-line hybrid transit: case of radial route structure," *Transportation Research Part E: Logistics and Transportation Review*, vol. 110, pp. 71–89, 2018.

[26] F. Qiu, W. Li, and J. Zhang, "A dynamic station strategy to improve the performance of flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 229–240, 2014.

[27] B. Alshalalfah and A. Shalaby, "Development of important relationships for the planning of flex-route transit services," in *Proceedings of Transportation Research Board 89th Annual Meeting*, pp. 1–21, Washington DC, USA, 2010.

[28] S. M. Nourbakhsh and Y. Ouyang, "A structured flexible transit system for low demand areas," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 204–216, 2012.

[29] B. Alshalalfah and A. Shalaby, "Feasibility of flex-route as a feeder transit service to rail stations in the suburbs: case study in Toronto," *Journal of Urban Planning and Development*, vol. 138, no. 1, pp. 90–100, 2012.

[30] S. Chandra and L. Quadrifoglio, "A model for estimating the optimal cycle length of demand responsive feeder transit services," *Transportation Research Part B: Methodological*, vol. 51, no. 2, pp. 1–16, 2013.

[31] S. Chandra and L. Quadrifoglio, "A new street connectivity indicator to predict performance for feeder transit services," *Transportation Research Part C: Emerging Technologies*, vol. 30, no. 3, pp. 67–80, 2013.

[32] D. Edwards and K. Watkins, "Comparing fixed-route and demand-responsive feeder transit systems in real-world settings," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2352, no. 1, pp. 128–135, 2013.

[33] R. Kelly, N. Ronald, and M. Wallace, "Exploring the effects of mixed request schemes for demand-responsive feeder services," in *Proceedings of 21st International Congress on Modeling and Simulation*, pp. 1731–1737, Gold Coast, Australia, November 2015.

[34] M. V. Engelen, O. Cats, and H. Post, "Enhancing flexible transport services with demand-anticipatory insertion heuristics," *Transportation Research Part E*, vol. 110, no. 1, pp. 110–121, 2018.

[35] C. Frei, M. Hyland, and H. S. Mahmassani, "Flexing service schedules: assessing the potential for demand-adaptive hybrid transit via a stated preference approach," *Transportation Research Part C: Emerging Technologies*, vol. 76, pp. 71–89, 2017.

[36] A. Gschwender, S. Jara-Díaz, and C. Bravo, "Feeder-trunk or direct lines? economies of density, transfer costs and transit structure in an urban context," *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 209–222, 2016.

[37] S. Kikuchi and R. A. Donnelly, "Scheduling demand-responsive transportation vehicles using fuzzy-set theory," *Journal of Transportation Engineering*, vol. 118, no. 3, pp. 391–409, 1992.

[38] S. C. Hsu, "Determinants of passenger transfer waiting time at multi-modal connecting stations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 3, pp. 404–413, 2010.

[39] R. Nair and E. Miller-Hooks, "Equilibrium network design of shared-vehicle systems," *European Journal of Operational Research*, vol. 235, no. 1, pp. 47–61, 2014.

[40] X. Lu, J. Yu, X. Yang, S. Pan, and N. Zou, "Flexible feeder transit route design in complexroad network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 6, pp. 128–134, 2016.

[41] X. Lu, J. Yu, X. Yang, S. Pan, and N. Zou, "Flexible feeder transit route design to enhance service accessibility in urban area," *Journal of Advanced Transportation*, vol. 50, no. 4, pp. 507–521, 2016.

[42] Y. Zheng, W. Li, and F. Qiu, "A slack arrival strategy to promote flex-route transit services," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 442–455, 2018.

[43] J. Tang, Y. Yang, and Y. Qi, "A hybrid algorithm for urban transit schedule optimization," *Physica A: Statistical Mechanics and Its Applications*, vol. 512, pp. 745–755, 2018.

[44] Y. Guo, J. Ma, C. Xiong, X. Li, F. Zhou, and W. Hao, "Joint optimization of vehicle trajectories and intersection controllers with connected automated vehicles: combined dynamic programming and shooting heuristic approach," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 54–72, 2019.

[45] S. Chandra and L. Quadrifoglio, "Critical street links for demand responsive feeder transit services," *Computers & Industrial Engineering*, vol. 66, no. 3, pp. 584–592, 2013.

[46] M. Dessouky and S. Adam, "Real-time schedulingrules for demand responsive transit systems," in *Proceedings of 1998 IEEE International Conference on Systems, Man, and Cybernetics*, vol. 3, pp. 2956–2961, San Diego, CA, USA, October 1998.

[47] J. Zhao and M. Dessouky, "Service capacity design problems for mobility allowance shuttle transit systems," *Transportation Research Part B: Methodological*, vol. 42, no. 2, pp. 135–146, 2008.

[48] X. G. Li and L. Quadrifoglio, "Optimal zone design for feeder transit services," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 1, pp. 100–108, 2009.

[49] F. Errico, T. G. Crainic, F. Malucelli, and M. Nonato, "A survey on planning semi-flexible transit systems: methodological issues and a unifying framework," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 324–338, 2013.

[50] X. Li and L. Quadrifoglio, "2-vehicle zone optimal design for feeder transit services," *Public Transport*, vol. 3, no. 1, pp. 89–104, 2011.

[51] L. Quadrifoglio, M. M. Dessouky, and F. Ordóñez, "Mobility allowance shuttle transit (MAST) services: MIP formulation and strengthening with logic constraints," *European Journal of Operational Research*, vol. 185, no. 2, pp. 481–494, 2008.

[52] L. Quadrifoglio, M. M. Dessouky, and F. Ordóñez, "A simulation study of demand responsive transit system design,"

Transportation Research Part A: Policy and Practice, vol. 42, no. 4, pp. 718–737, 2008.

[53] L. Quadrifoglio and X. Li, "A methodology to derive the critical demand density for designing and operating feeder transit services," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 922–935, 2009.

[54] S. M. M. Amiripour, A. Ceder, and A. S. Mohaymany, "Designing large-scale bus network with seasonal variations of demand," *Transportation Research Part C*, vol. 48, pp. 332–338, 2014.

[55] A. Pratelli and F. Schoen, "A mathematical programming model for the bus deviation route problem," *Journal of the Operational Research Society*, vol. 52, no. 5, pp. 494–502, 2001.

[56] I. Razi and G. U. Muhammad, "Intelligent bus stops in the flexible bus systems," *Journal of Engineering Science and Technology Review*, vol. 7, no. 4, pp. 59–65, 2014.

[57] J. Jaeyoung, R. Jayakrishnan, and D. Nam, "High coverage point-point transit: hybrid evolutionary approach to local vehicle routing," *KSCE Journal of Civil Engineering*, vol. 19, no. 6, pp. 1882–1891, 2015.

[58] S. Chandra, M. E. Bari, P. C. Devarasetty, and S. Vadali, "Accessibility evaluations of feeder transit services," *Transportation Research Part A: Policy and Practice*, vol. 52, pp. 47–63, 2013.

[59] A. Ceder, "Optimal multi-vehicle type transit timetabling and vehicle scheduling," *Procedia-Social and Behavioral Sciences*, vol. 20, pp. 19–30, 2011.

[60] P. C. Guedes and D. Borenstein, "Real-time multi-depot vehicle type rescheduling problem," *Transportation Research Part B: Methodological*, vol. 108, pp. 217–234, 2018.

[61] Z. W. Wang, T. Chen, and M. Q. Song, "Coordination optimization of operating lines and scheduling for Responsive Feeder Transit under the mixed operation mode," *Journal of Traffic and Transportation Engineering*, vol. 19, no. 5, pp. 1–9, 2019.

[62] C. P. Chen, S. J. Han, J. S. Lu et al., "A multi-chromosome Genetic algorithm for multi-depot and multi-type vehicle routing problems," *China Mechanical Engineering*, vol. 29, no. 2, pp. 218–223, 2018.

[63] M. Wei, T. Liu, B. Sun, and B. Jing, "Optimal integrated model for feeder transit route design and frequency-setting problem with stop selection," *Journal of Advanced Transportation*, vol. 2020, pp. 1–12, 2020.

[64] Z. Wang, J. Yu, W. Hao et al., "Two-step coordinated optimization model of mixed demand responsive feeder transit," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 3, Article ID 04019082, 2020.

[65] L. Ahmed, C. Mumford, and A. Kheiri, "Solving urban transit route design problem using selection hyper-heuristics," *European Journal of Operational Research*, vol. 274, no. 2, pp. 545–559, 2019.

[66] W. Hao, Y. Lin, Y. Cheng, and X. Yang, "Signal progression model for long arterial: intersection grouping and coordination," *IEEE Access*, vol. 6, pp. 30128–30136, 2018.

[67] S. C. Ho, W. Y. Szeto, Y.-H. Kuo, J. M. Y. Leung, M. Petering, and T. W. H. Tou, "A survey of dial-a-ride problems: literature review and recent developments," *Transportation Research Part B: Methodological*, vol. 111, pp. 395–421, 2018.

[68] M. M. Solomon, "Algorithms for the vehicle routing and scheduling problems with time window constraints," *Operations Research*, vol. 35, no. 2, pp. 254–265, 1987.

[69] S. W. Wang, L. S. Sun, and J. Rong, "Catchment area analysis of Beijing transit Stations," *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, no. 3, pp. 183–188, 2013.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 3236841, 16 pages
https://doi.org/10.1155/2020/3236841

WILEY | Hindawi

Research Article

Coordinated Headway-Based Control Method to Improve Public Transit Reliability considering Control Points Layout

Hu Zhang, Shidong Liang, Jing Zhao, Shengxue He, and Tianyu Zhao

Department of Traffic Engineering, University of Shanghai for Science and Technology, Shanghai, China

Correspondence should be addressed to Shidong Liang; sdliang@hotmail.com

Received 21 November 2019; Revised 22 March 2020; Accepted 4 June 2020; Published 9 July 2020

Academic Editor: Tao Liu

The headway-based control method is usually used to regulate the bus headways and improve reliability of public transit. In general, the holding control strategy is applied at the control point, because enough space for dwell longer at the control point is required, while the stop-skipping control strategy can be used at any bus stop. However, in the headway-based control method, too much stop-skipping will bring longer waiting time and make the passengers impatient. The number and distribution of control points for stop-skipping are not considered in previous self-equalizing bus headway control works. Therefore, in this paper, the control points selection rules for stop-skipping involving their number and distribution on the bus route are discussed. A second by second discrete system is formulated to describe the bus operation. In the proposed control method, the threshold value for activating stop-skipping strategy is raised, avoiding provoking much additional waiting time because of boarding rejected. In the numerical analysis, a set of cases are conducted to evaluate the performance of control method under different number and distribution of control points for stop-skipping. The numerical results show that distribution of control points for stop-skipping has a greater influence on the public transit than the number.

1. Introduction

Bus bunching is a phenomenon when two or more buses encounter the same bus stop on a bus line, and it commonly appears in an unstable high-frequency bus line system. If a bus running on a bus route is delayed by an incidental disturbance, this bus should pick up more passengers than expected at the downstream bus stop. Therefore, the bus should dwell longer than expected and further gets delayed because of more boarding passengers at the bus stop. Meanwhile, its following bus runs closer to the leading bus and meets less number of passengers at the bus stop; therefore, it is speeded up because of shorter dwell time than expected. The leading bus gradually slows down, while the following bus speeds up, resulting in the two successive buses arriving at a same bus stop at the same time or in a very short time interval inevitably.

Bus bunching results in longer waiting and travel times for the passengers and the increased environment pollution due to inefficient operation of unevenly loaded buses. In addition, the unbalanced load of passengers on successive buses wastes bus capacity, because the leading bus is quite crowded while the trailing bus is relatively empty.

To resist bus bunching, many control methods were proposed in previous related works. With the advent of new technologies, some holding strategies have been proposed to take advantage of real-time information so as to reduce passengers' waiting times [1, 2]. Using real-time information, many headway-based holding strategies have been proposed to adaptively control the system [3–13]. A "two-way-looking control" bus headway control method was proposed by Daganzo [14]. This method attempts to make the headways presettled static values. Later, Daganzo and Pilachowski [15] proposed a control strategy that continuously adjusts the bus's cruising speed on the route based on a cooperative two-way system to achieve proper spacing between the successive buses on the line. As an extension of this control concept, Xuan et al. [16] and He [6] proposed a holding strategy to regularize headways while maximizing the commercial speeds, as well as considering both the forward and backward headways. Bartholdi and Eisenstein

[4] proposed a self-adaptive control method to automatically equalize bus headways on the bus route. To integrate the advantages of the "two-way-looking control" and the "self-adaptive equalizing bus headway," Liang et al. [17] proposed a self-equalizing control strategy based on the two-way-looking control method (the headways between the bus at the control point and both its leading and following buses) with zero slack. Zhang and Lo [18] proposed two-way-looking self-equalizing headway control that considered multifarious variables, enriching the headway-based bus holding control method system. More recently, He et al. [19] and Liang et al. [20] proposed a dynamic target headway-based control method to resist bus bunching with holding strategy, which is another version of self-equalizing bus headway control method.

However, in the self-adaptive equalizing bus headway control method system, only one holding control means was used. Bus holding is based on slowing the faster leading bus and increasing the slack to balance headways, thus decreasing the average operating speed of the total bus system to resist bus bunching. Therefore, other control means at the control point can be integrated with bus holding to enhance the performance of control method based on the self-adaptive equalizing bus headway control concept. Stop-skipping strategy is used in which a bus is asked to skip a stop as long as nobody inside the bus requests to stop, even though there are passengers at the bus stop waiting to board the bus. Different criteria such as the operational costs, average waiting time, and passenger awareness are considered to determine which stops should be skipped. In the early work on rail systems by Suh et al. [21], a stop-skipping strategy was utilized to increase the speed of subway service. Sun and Hickman [22] proposed a stop-skipping strategy to minimize the passenger waiting times but allowed the passengers to alight at stops in the skipped segment. A multiobjective optimization approach for stop-skipping strategies was addressed by Sidi et al. [23]. Cortés et al. [24] and Sáez et al. [25] developed a hybrid predictive control formulation that jointly estimates the bus speed, while implementing control strategies such as skipping stations or holding buses. Delgado et al. [5] addressed the combination of holding strategies with boarding limits. The authors developed a deterministic optimization model capable of executing two strategies: holding and boarding limits (limiting the amount of passengers that can board a bus even if the bus is not at full capacity); this control strategy was also used by [26]. In recent years, Zhang et al. [27, 28] and Wu et al. [29, 30] proposed a set of limited-stop control methods to deal with stochastic travel time and uncertain demand.

The holding control means should be implemented at the bus stops selected as control points, because the control point should provide enough space for bus to dwell longer, while the stop-skipping control means can be used at any bus stop. However, the stop-skipping control method will provoke more waiting time. Therefore, in this paper, the integrated control method involving bus holding and stop-skipping control means is proposed based on the self-equalizing bus headway control concept. The method that integrated "two-way-looking control" with "self-adaptive equalizing bus headway" has been proven to perform well in reducing bus bunching and improving the service level for passengers. However, the control concept cannot be implemented completely in the proposed integrate control method, because the stop-skipping is a strong control means which will provoke dissatisfaction of passengers who are rejected to board. Therefore, the priority of activating stop-skipping is reduced, with higher threshold value, giving consideration to both performance of equalizing bus headways and feeling of passengers.

The remainder of this paper is organized as follows. In Section 2, the coordinated control algorithm is proposed to dynamically select the two control means based on predictive control. In Section 3, the buses' operation process has been described by formulated as a discrete system. The operation rules are divided into operation at normal bus stop and control point, respectively. In Section 4, two kinds of performance indexes are proposed to reflect the regulation of bus operation and the condition of passengers on the buses. A set of numerical tests are conducted in Section 5, using the proposed coordinated control method under different number and distribution of control points for stop-skipping.

2. Self-Equalizing Bus Headways Control Strategy with Two Control Means

In this section, a dynamic control strategy was developed to equalize bus headways, resisting bus bunching. A dynamic selection algorithm of control means including bus holding, stop-skipping, and no control means was implemented. The control variables, e.g., holding time, can be calculated by analysing both the real-time bus running status and passenger demand. In order to distinguish the control point for bus holding only, stop-skipping only, and coordinated control, CPH (control point for holding), CPS (control point for stop-skipping), and CPC (control point for coordinated control), respectively, are named. The coordinated control method, used at CPC, involves holding control strategy and stop-skipping control strategy. In practice, these control strategies should be selected dynamically according to the algorithm shown in Section 3.3.

2.1. Motivation of Proposing the Coordinated Control Method. The coordinated control method proposed in this paper is derived from the control concept of self-adaptive equalizing bus headways proposed in Liang et al. [17, 31], which is suitable to be applied to a high-frequency bus route. A self-adaptive method to equalize headways based on bus holding was tested and proved to perform well to reduce slack time and enhance operation efficiency. However, successive bus headways at control point cannot be equalized completely using bus holding only, because bus holding is insufficient when the forward headway is larger than the backward headway; thus the bus holding cannot speed up the slow bus. Recently, Liang et al. [32] fixed this gap by integrating the two control means into self-adaptive equalizing bus headways framework. According to the test results compared with bus holding used only, the coordinated control method can shorten total travel time and make

bus headways more regular. However, the waiting time increased because of stop-skipping control means. Therefore, in order to decrease the negative effects brought by stop-skipping control means, in this paper, it is treated as a supplementary means. When the headways between the bus at control points and the following bus are much smaller than the bus headway with leading bus, the stop-skipping control means shall be used. In other word, the stop-skipping will be activated with higher threshold value, giving consideration to regular bus operation and passengers' experience.

In fact, other control strategies, i.e., speed control and schedule recovery, can achieve similar control result. However, the speed control is usually used on bus lanes, without disturbance of cars on the street [33]. Therefore, in this paper, the speed control is not considered as supplement to holding control strategy. Although schedule recovery control concept is quite useful in practice to regulate bus operation, it is not suitable to be integrated as a supplement strategy in the self-equalizing control concept.

In traditional research, the stop-skipping control is a static schedule to improve operation efficiency of bus system instead of a dynamical control means to control the bus. Boarding limits can accelerate the bus at control point, while the following bus can be slowed down by leaving more passengers to it. However, in practice, it is not easy to refuse part of passengers waiting at the bus stop to get on the bus even when the bus is not full. Therefore, stop-skipping in this paper refers to the bus only providing service for alighting passengers when the control point is skipped, proposed by Sun and Hickman [22]. This kind of stop-skipping control method can speed up the slower bus through dwelling shorter time and meanwhile slow down the faster bus by giving it more passengers to pick up. In addition, the passengers on the bus can alight at any bus stops, avoiding missing the destination bus stop and causing dissatisfaction of passengers.

If the bus headways are quite equalized, the bus at the CPC should not be controlled with any control means. We term such a condition as "no control means." Therefore, three control means can be used at the CPC which are bus holding, stop-skipping, and no control means. The core of the coordinated control method is to dynamically select proper control means to equalize the bus headways, according to the prediction of control results. Because studies that predict the related parameters in bus systems are relatively mature [34–41], the evolution of bus systems can be predicted with these methods in practice. In addition, the estimations of urban traffic parameters are also sufficiently mature, e.g., travel speed and queue length at signalized intersection [42–45]. Therefore, the main objective of this section is to propose a method that combines the bus holding control means and stop-skipping control means based on the self-adaptive equalizing headways control concept. The detailed control strategy and algorithm are introduced in the following subsection.

2.2. Control Strategy Coordinated Control Means at CPC. When a bus has just arrived at the CPC, it should judge whether the stop-skipping control means is suitable to be used to equalize the bus headways. If the stop-skipping control means is selected to be used, the bus can leave the CPC directly after letting the passengers alight. If the stop-skipping control means is not suitable to be used, the bus can provide normal service for the passengers on the bus and wait at the CPC. After the normal service, it should be judged whether the bus holding control means is suitable to be used to equalize the bus headways. If the holding control means is not suitable to be used, the bus can leave the CPC. Otherwise the bus should dwell at the CPC for suitable additional time.

Note that the stop-skipping control means should be considered when the bus has just arrived at the CPC, and the bus holding should be considered after the bus providing service, because the predictive values of headways can be more accurate due to shorter prediction step length.

According to the previous research conclusion by Liang et al. [17, 31, 32], the bus headways can be convergent to a common value after continually equalizing the two bus headways between the bus at CPC and its leading and following buses. Therefore, we only focus on three buses (i.e., the bus arriving at CPC and its leading and following buses) at a moment, as shown in Figure 1. We define the bus headway as the time between two successive buses leaving the same bus stop. According to this definition, the bus headway between the bus at CPC and its leading bus can be predicted after the bus at the CPC leaving current bus stop, named $\widetilde{h}_l$. The bus headway between the bus at CPC and its following bus can be predicted after the following bus leaving CPC, named $\widetilde{h}_f$. The selection of variable control means can be obtained through comparing the $\widetilde{h}_l$ and $\widetilde{h}_f$ under different control means. Concretely, the control process at CPC for both bus holding and stop-skipping is shown in Figure 2.

Note that, according to the research results of Liang et al. [32], based on this control principle, the stop-skipping control means was used relatively frequently increasing the waiting time of passengers. In order to decrease the additional waiting time of passengers and maintain the regular bus headway of public transit, in this paper stop-skipping is treated as a supplementary means whose priority is lower than bus holding and no control means. If the difference of two bus headways at the CPC is relatively large, the stop-skipping control means can be used. More exactly, in this paper if the following bus headway $\widetilde{h}_f^{\mathrm{ss}}$ controlled by stop-skipping means is still smaller than the leading bus headway $\widetilde{h}_l^{\mathrm{ss}}$, the stop-skipping control means should be used. According to this control principle, the stop-skipping control means was used less and ensured large difference of successive bus headways can be resisted, considering both efficiency of public transit and experience of passengers. The stop-skipping can be activated with higher threshold value.

According to the dynamic control means selection procedure, the bus at the CPC can make proper decision to equalize the two successive bus headways. The bus headways on the whole route can be regulated by continually equalizing the successive bus headways.

2.3. Control Strategy for Bus at CPS. As mentioned in Section 2.1, the stop-skipping control means is a strong way to

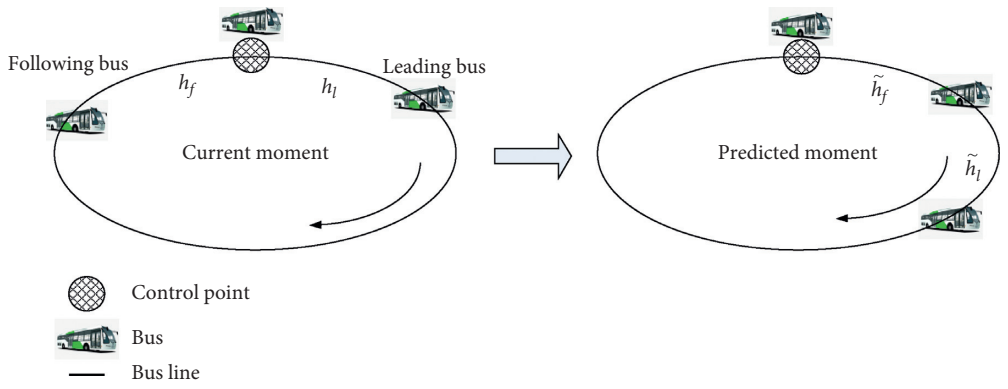


Figure 1: Prediction step length in the coordinated control method.

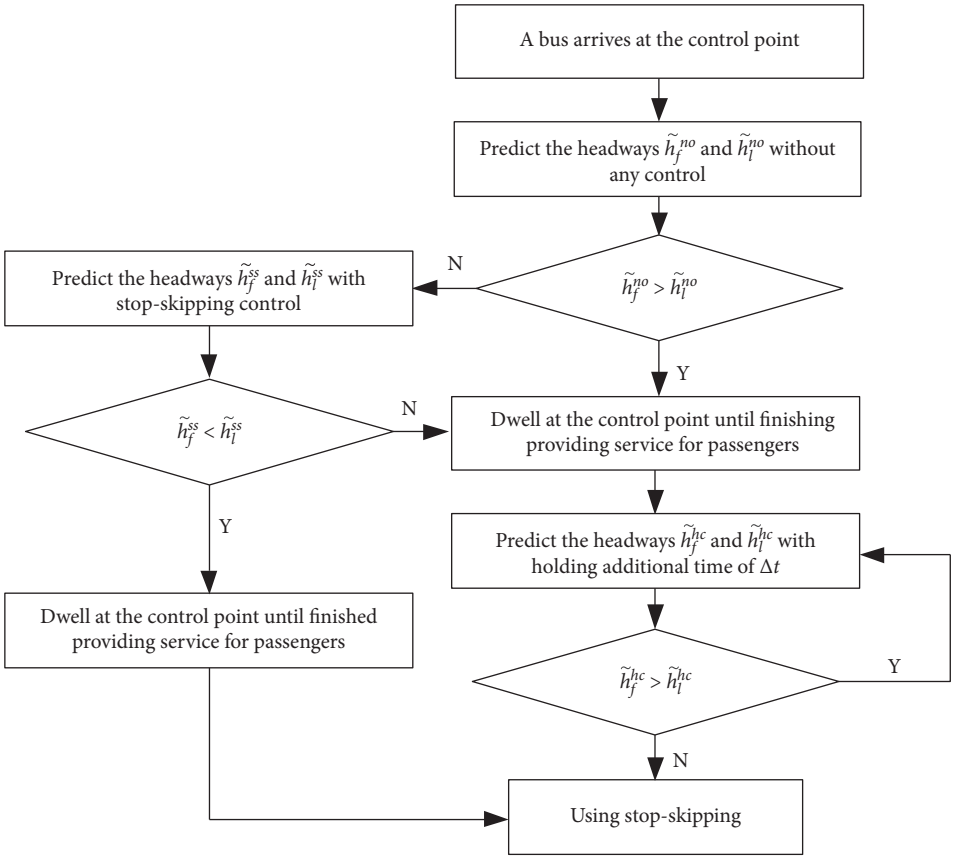


Figure 2: Control procedure for buses at CPC.

balance the bus headway. Meanwhile, stop-skipping control mean will provoke dissatisfaction of passengers waiting at the bus stop. At the CPS instead of CPC, this control means can be used according to the control concept shown in Figure 3.

3. Bus Operation and Passengers' Evolution under Different Control Scenarios

In this section, a second by second discrete system for description of bus operation process is formulated. On a bus line, the bus always runs on the road between two adjacent bus stops or dwells at the bus stop to provide service to the passengers. The bus stop can be classified into normal bus stop and control point.

3.1. Bus Operation at Normal Bus Stops for No Control. A discrete system is designed to describe the operation of buses. The running time is divided into a serious of uniform short time intervals Δt, e.g., 1 second.

A bus can move forward between two adjacent bus stops with distance $\Delta t \cdot v(t)$; we assume the velocity $v(t)$ of bus is constant during short time interval from t to $t + \Delta t$. Therefore, the location $l_n(t + \Delta t)$ of bus n at time $t + \Delta t$ on the bus line between two adjacent bus stops can be shown as

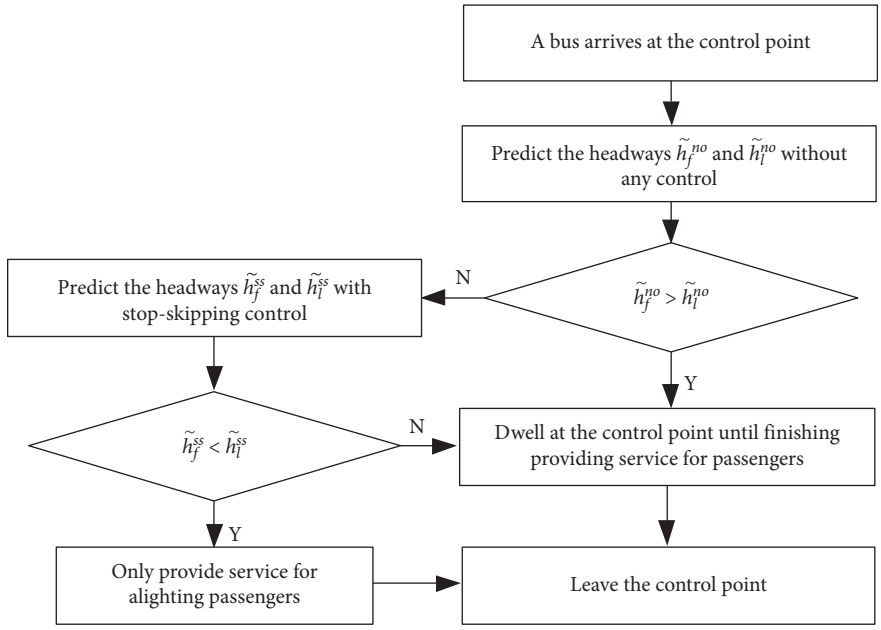


Figure 3: Principal control strategy procedure for buses at CPS.

$$l_n(t+\Delta t) = l_n(t) + \Delta t \cdot v(t). \tag{1}$$

If a bus n encounters bus stop s, the passengers waiting at the bus stop s begin to get on bus n. The criteria of bus n arriving at bus stop s, whose location is expressed by l_s, can be written as

$$l_n(t+\Delta t) - l_s \in [0, \Delta t \cdot v(t)). \tag{2}$$

The boarding time of one passenger is assumed constant, expressed by α, and the alighting time of one passenger is constant as well, expressed by β. Therefore, the dwell time of one bus at bus stop for providing service to passengers can be written as

$$t_d = \max(\alpha \cdot P_b, \beta \cdot P_a) \tag{3}$$

where P_b and P_a refer to the number of boarding and alighting passengers, respectively.

In the description system, the $P_n(t)$ represents number of passengers on the bus n at time t; $W_s(t)$ refers to number of passengers waiting at the bus stop s at time t; $A_s(t)$ means number of passengers that newly arrived at bus stop s during the period from t to $t+\Delta t$. In one time interval Δt, the process of alighting and boarding can be expressed by

$$\begin{cases} P_n(t+\Delta t) = P_n(t) + \dfrac{\Delta t}{\alpha} - \dfrac{\Delta t}{\beta}, \\ \\ W_s(t+\Delta t) = W_s(t) + A_s(t) - \dfrac{\Delta t}{\alpha}. \end{cases} \tag{4}$$

In fact, equation (4) only gives description of alighting and boarding process without considering bus capacity limit. If the bus capacity C is considered, equation (4) is only suitable to be used when $P_n(t) + (\Delta t/\alpha) - (\Delta t/\beta)$ is not larger than the C. Otherwise, the passengers cannot get on the bus because of capacity limit, although there still are passengers on the bus stop s. Therefore, the passengers at the bus stop cannot get on bus n. The alighting and boarding process for passengers can be expressed by

$$\begin{cases} P_n(t+\Delta t) = P_n(t) - \dfrac{\Delta t}{\beta}, \\ \\ W_s(t+\Delta t) = W_s(t) + A_s(t). \end{cases} \tag{5}$$

If the total number of alighting passengers $\sum \Delta t/\beta$ during dwell time has been equal to P_a, the part $\Delta t/\beta$ in line one of equations (4) and (5) should be deleted.

The bus n cannot leave bus stop s until meeting one of following two constraint conditions:

$$\begin{cases} W_s(t) = 0, \quad P_a = \sum \dfrac{\Delta t}{\beta}, \\ \\ P_n(t) = C, \quad P_a = \sum \dfrac{\Delta t}{\beta}. \end{cases} \tag{6}$$

3.2. Bus Operation at CPS. If bus n has determined to skip bus stop s, the bus still should dwell at the bus stop s to provide service for the alighting passengers, if any. Therefore, the location of bus is the same as the bus stop, l_s.

$$l_n(t+\Delta t) = l_s \tag{7}$$

Meanwhile, the passengers on bus who would like to alight can get off, while the passengers waiting at the bus stop cannot get on the bus. Therefore, the bus cannot leave the bus stop until all the passengers who would like to get off have finished. The process of alighting and the number of passengers waiting at bus stop s can be expressed by

$$\begin{cases} P_n(t+\Delta t) = P_n(t) - \dfrac{\Delta t}{\beta}, \\ \\ W_s(t+\Delta t) = W_s(t) + A_s(t). \end{cases} \tag{8}$$

When the stop-skipping control means is used, the capacity of bus should not be considered as constraint. The bus can leave the CPS when all the passengers wanting to alight at the bus stops have finished, which can be written as

$$P_{\mathrm{a}} = \sum \frac{\Delta t}{\beta}. \tag{9}$$

After providing service for the alighting passengers, the bus can leave the bus stop, and the location evolution of bus can be written as

$$l_n(t+\Delta t) = l_s + \Delta t \cdot v(t). \tag{10}$$

3.3. Bus Operation at CPC. If the encountered bus stop is a CPC, the bus can leave the CPC when all the passengers wanting to alight at the bus stop s have finished, which can be written as equation (9).

If the stop-skipping control means is abandoned, the bus can provide service for passengers at the CPC. The evolution of passengers on bus and the CPC is same as equations (4) and (5). When the bus has finished the normal service for alighting and boarding passengers, it should consider whether the bus holding control means should be used or not. As mentioned in Section 2, the passengers that newly arrived at the CPC can still get on the holding bus. Therefore, the boarding process and the passengers' number during the holding time can be expressed by

$$\begin{cases} P_n(t+\Delta t) = P_n(t) + \dfrac{\Delta t}{\alpha}, \\ \\ W_s(t+\Delta t) = W_s(t) + A_s(t) - \dfrac{\Delta t}{\alpha}. \end{cases} \tag{11}$$

The bus can leave the CPC when the predicted bus headway $\widetilde{h}_f^{\mathrm{hc}}$ is no larger than $\widetilde{h}_l^{\mathrm{hc}}$, and all the newly arrived passengers have boarded the bus. The condition should meet

$$\begin{cases} W_{\mathrm{s}}(t) = 0, \\ \widetilde{h}_f^{\mathrm{hc}} \le \widetilde{h}_l^{\mathrm{hc}}. \end{cases} \tag{12}$$

If the number of passengers on the bus has reached the capacity of the bus, the passengers cannot get on the bus any longer. Therefore, if the bus is full, the passenger evolution during holding time can be written as

$$\begin{cases} P_n(t+\Delta t) = P_n(t), \\ W_s(t+\Delta t) = W_s(t) + A_s(t). \end{cases} \tag{13}$$

The bus can leave the CPC when the condition shown in equation (14) is achieved:

$$\widetilde{h}_f^{\mathrm{hc}} \le \widetilde{h}_l^{\mathrm{hc}}. \tag{14}$$

4. Performance Index for Headways and Travel Time of Passengers

4.1. Total Travel Time of Passengers' Calculation Model. In this paper, the bus operation process and the passengers' arrival process are described as a discrete system. The number of passengers at the bus stops may not be integer in every time. Therefore, only integer part of the passengers at the bus stops should be added up.

Therefore, the total waiting time T_w at the bus stops can be calculated by the formula written as

$$T_w = \Delta t \cdot \sum_{t=o}^{T} \sum_{s=1}^{S} \lceil W_s(t) \rceil. \tag{15}$$

In equation (15), the $\lceil W_s(t) \rceil$ means the integer part of passengers' number at the bus stops, and the T means the end of implement time of t. Adding up all the passengers at bus stops at each time t and multiplying time interval Δt, the total waiting time for passengers at the bus stops can be obtained.

The number of passengers on buses is not always an integer, because the operation of buses and passengers' description system are discrete. When the bus dwells at the bus stop, the passenger may not finish the boarding process in one interval Δt. Therefore, the travel time of passengers on buses cannot be the summation of passengers' number in every interval Δt. In addition, the waiting time at the bus stops adds up all the integer part, so that the decimal part should be counted into the travel time on buses. Therefore, the total travel time T_p on buses can be calculated by

$$T_p = \Delta t \cdot \sum_{t=o}^{T} \sum_{n=1}^{N} \lfloor P_n(t) \rfloor. \tag{16}$$

In equation (17), N refers to the total number of buses running on the bus route, and $\lfloor P_n(t) \rfloor$ means the passengers on buses rounding up to an integer. Therefore, adding up all the passengers on every bus during all the time and multiplying the time interval Δt can obtain the total travel time of passengers on buses. Therefore, the total travel time TT for passengers including the waiting time at bus stops and travel time on buses can be written as

$$\mathrm{TT} = T_w + T_p. \tag{17}$$

4.2. Bus Headways Calculation Model. The other index reflecting the performance of public transit is the bus headway. In the transit corridor, the bus operation pays more attention on the deviation of bus headways instead of a schedule. Because the average bus headways value is relatively small (about 3 min to 5 min), the bus headways are expected to be equalized with each other.

In this paper, the bus headway is defined as the time between two successive buses leaving the same bus stop. Therefore, the bus headway at the bus stop s can be expressed by

$$H_s = \begin{cases} t_{n+1,s} - t_{n,s}, & \text{if } n < N, \\ t_{1,s} - t_{n,s}, & \text{if } n = N. \end{cases} \quad (18)$$

In equation (18), $t_{n,s}$ means the time bus n leaving bus stop s. The bus n is the leading bus, followed by bus $n+1$. Therefore, the bus headway should be $t_{n+1,s} - t_{n,s}$. It should be mentioned that if the value of n is equal with the total number of buses on the route N, $n+1$ is not meaningful. According to the actual physical meaning, the bus $N+1$ should be bus 1. The standard deviation of bus headways can be expressed by

$$\sigma_s = \sqrt{\frac{1}{N_n - 1} \sum_{1}^{N_n - 1} H_s}. \quad (19)$$

In equation (19), N_n means the times buses passed the bus stop, so that the number of bus headways is $N_n - 1$.

Therefore, the index values of travel time of passengers and bus headways can be calculated by the equations formulated in this section.

5. Numerical Analysis

In this section, numerical tests are conducted to analyze the performance under variable CPS selected. In Section 5.1, the input data for the bus route is introduced including the arrival process of passengers at bus stops and the distribution of bus stops on the bus line. In Section 5.2, the test results under variable distribution of CPS are presented with index of performance mentioned in Section 4. Finally, the result analysis is conducted and several conclusions and suggestions about the setting of CPS are obtained.

5.1. Input Data of the Bus Route. The length of the simulation bus route is 12 km, with 15 bus stops on the bus route. Average distance between two adjacent bus stops is 0.8 km. There are 11 buses running on the bus route whose capacity is 100 pax. The average travel speed between two successive bus stops is 5 m/s. Considering the random disturbance caused by passengers' arrival process, the travel speed of buses on the bus route is stochastic, because of random dwell time at bus stops. The simulation time in one test is 4 h. To evaluate the performance of the proposed control method under high traffic demands, the average arrival rate of the bus route is 2970 pax/h, and the distributions at the bus stops are same. The alighting proportions at each bus stop are shown in Figure 4. Since the arriving process is stochastic, special cases may occur. According to the Monte Carlo method, repeating the tests for enough times can obtain a more general conclusion. Therefore, to eliminate effects of random factors, the program was run 100 times to generate passenger arrival processes at bus stops. We built a simulation platform based on the running process of buses described in Section 3 and imbedded the control method into it. In Figure 5, the relationship between simulation times and the convergence tendency of performance indices is presented, illustrating that repeating 100 times is enough.

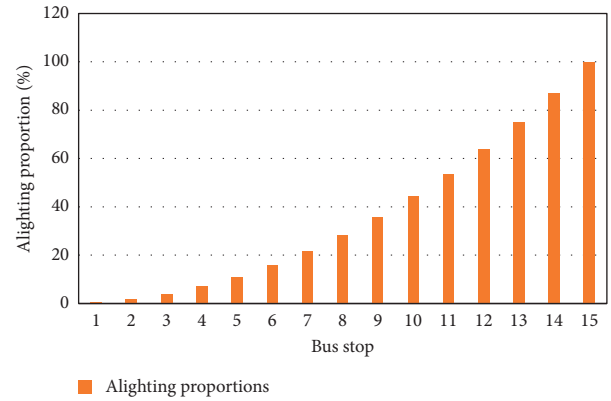


Figure 4: Percentage of average passenger alighting at each stop.

There are two CPH which are fixed at bus stop 6 and bus stop 11, while the variable setting of CPS is shown in Table 1.

As shown in Table 1, 10 groups are set to select variable locations of CPS. Among them, S1 and S2 have only one CPS where the bus stop 6 locates at the beginning section of bus route, while bus stop 11 locates at the end section of bus route. In group S3, the two CPS are same with CPH. The groups S4 and S5 have three CPS while the groups S6, S7, S8, and S9 have four CPS, and their spatial distributions are different. Group S10 have the most CPS with five.

5.2. Test Results under Variable Number and Location of CPS. In this section, the raw test results are presented first to give preliminary impression of bus operation under different number of CPS. More detailed indices and insight can be seen in Section 5.3.

Firstly, the trajectories of buses in groups S1, S3, S4, S7, and S10 are presented in Figure 6, where the five groups mean different number of CPS from 1 to 5. According to Figures 6(a) and 6(b), the bus operation is not quite regular as several buses tend to be very close to each other. In addition, the bus holding time at the two CPH, at 4 km and 8 km, is relatively long. According to Figures 6(c)–6(e), the trajectories of buses are smoother than those in the first two figures, in which the bus headways are more equalized with each other. It is obvious that the holding time at the two CPH is shorter than that in the first two figures as well. However, the bus headway cannot be more regular when the number of CPS is relatively large, because the situations presented in the last three figures are quite same.

Figure 7 shows the passengers on buses when the bus just left the bus stop under different number of CPS with five groups. According to Figure 7, the number of passengers grows gradually along with the buses running on the bus route, reaching peak values at bus stop 5 to bus stop 8 and then decreases gradually at the end section of the bus route. In more detail, the situations of passengers on buses presented in Figures 7(a) and 7(b) illustrate that the passengers on buses at bus stop 5 fluctuate greatly. The maximum values of passengers on buses at bus stop 4 to bus stop 7 reach the capacity of the bus 100 pax, while the minimum values are

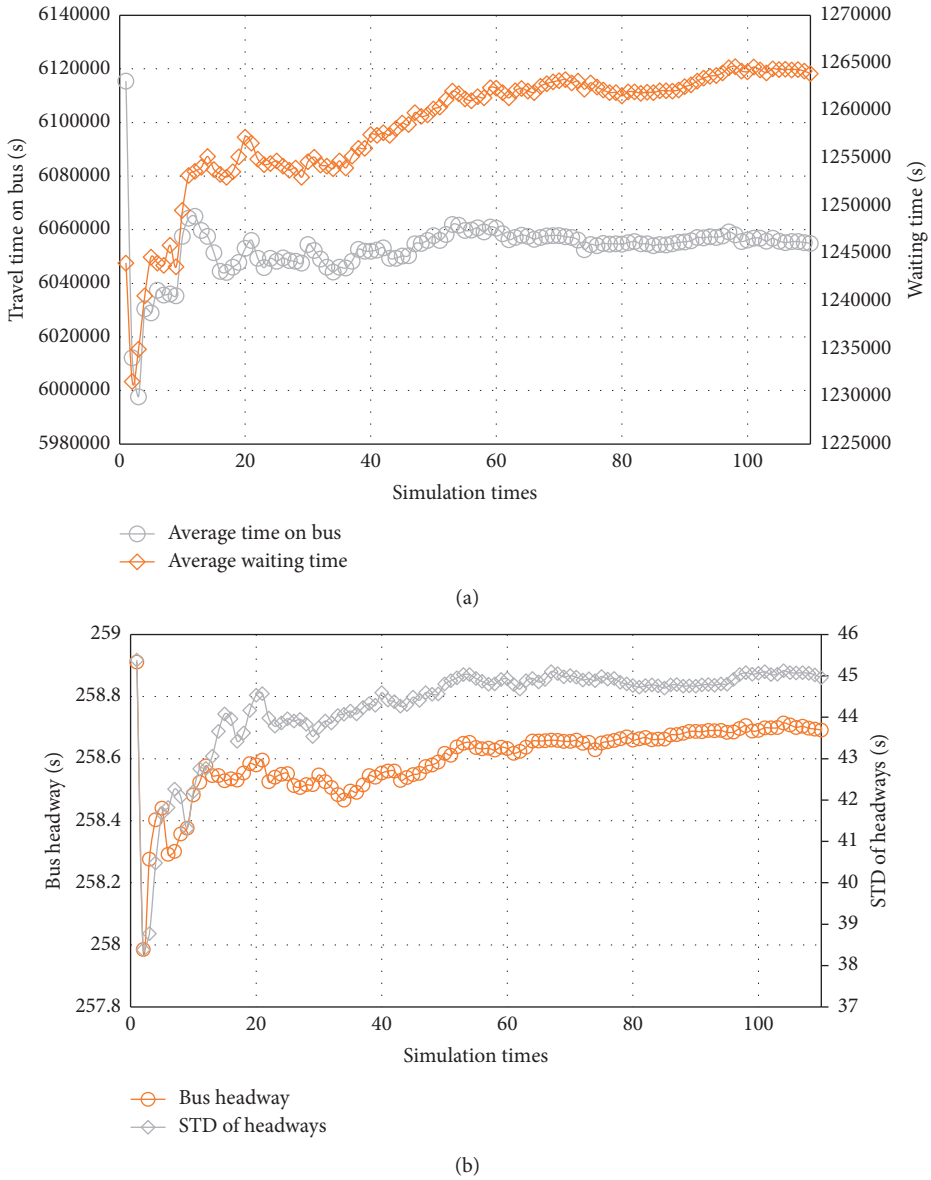


Figure 5: The convergent tendency with increasing of simulation times. (a) Travel time on bus and waiting time at bus stops. (b) Bus headways and STD of bus headways.

Table 1: Bus stops selected for stop-skipping.

Group	Location of CPS
S1	6
S2	11
S3	6, 11
S4	3, 6, 11
S5	6, 9, 11
S6	6, 11, 14
S7	2, 4, 6, 11
S8	6, 8, 10, 11
S9	6, 11, 13, 15
S10	3, 5, 8, 10, 13

quite small. It means the capacity of some buses is not fully used while the other buses are quite crowded. This phenomenon has been greatly improved in Figures 7(c)–7(e), where the fluctuating range of passengers on buses is relatively small, which means the capacity of buses fully used.

Similar to the conclusion obtained from Figure 6, the passengers on the bus feel more comfortable along with the increasing of CPS number, because the numbers of passengers in different buses at the same bus stop become more equalized. However, the brought benefits become inconspicuous when the number of CPS is large enough.

According to the intuitionistic raw test results, a preliminary conclusion can be achieved; the bus headway can be equalized and the bus operation process can be regulated along with the increasing of CPS number. However, the benefits will not increase persistently, especially when the number of CPS is relatively large. Therefore, the relationship between the location distribution of the CPS and the performance of the control method should be explored further.

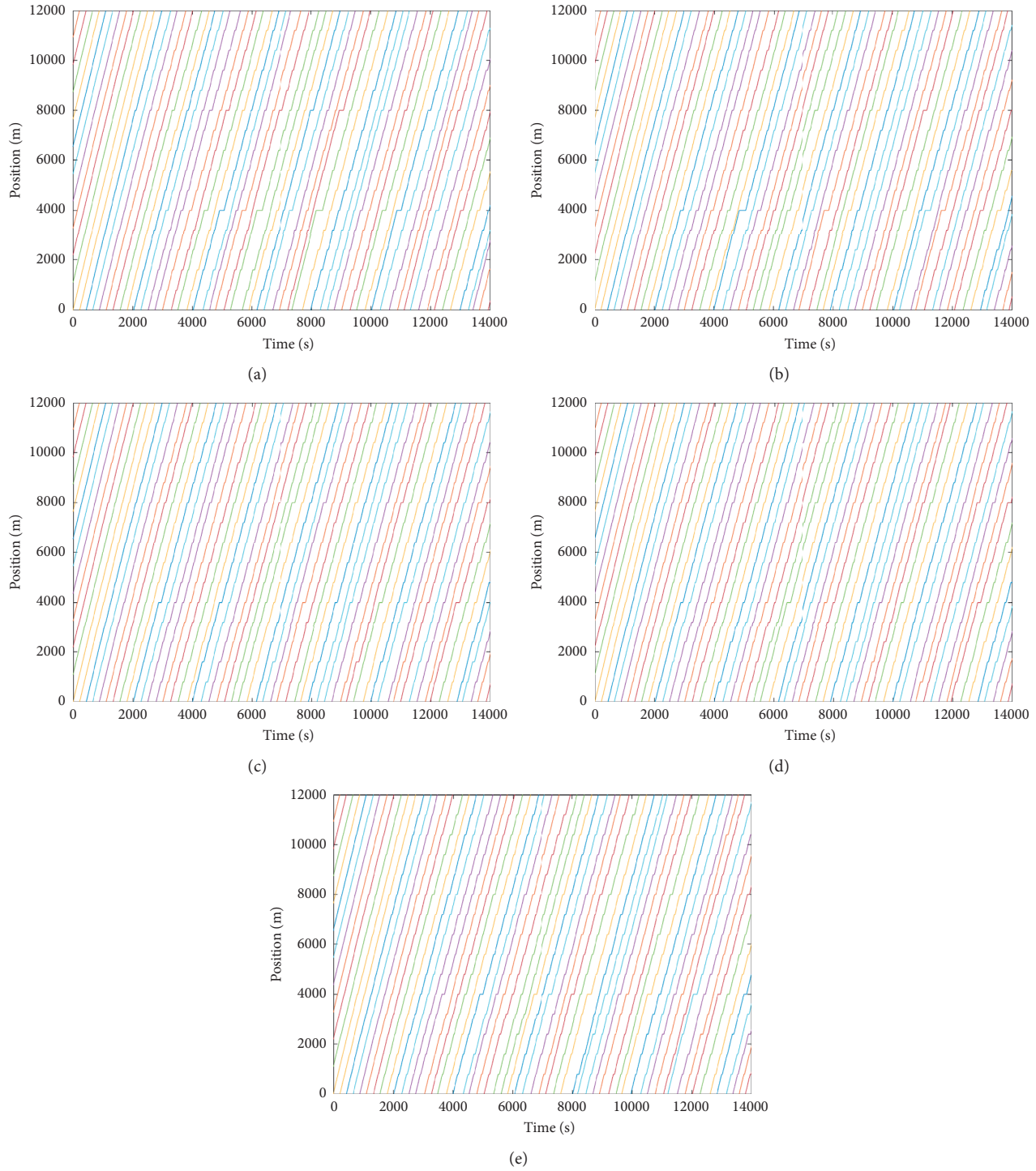


Figure 6: The trajectories of buses under different number of CPS. (a) S1. (b) S3. (c) S4. (d) S7. (e) S10.

5.3. Results Analysis under Variable Number and Distribution of CPS. In this section, the indices of the control method performance are presented including bus operation and the experience of passengers. The relationship between the performance of the control method and the number and distribution of CPS is presented followed by the explanation for these phenomena. Furthermore, several conclusions and suggestions are proposed for application in practice.

Because the passengers' arriving process is stochastic, only one simulation and test cannot obtain an impartial result. Therefore, we conduct many times simulations and tests and obtain an average result to reflect the true performance of the control method. However, the times should be discussed. More simulations can obtain more impartial result, while too many simulations may lead to waste of time. Therefore, we should obtain the convergence process along

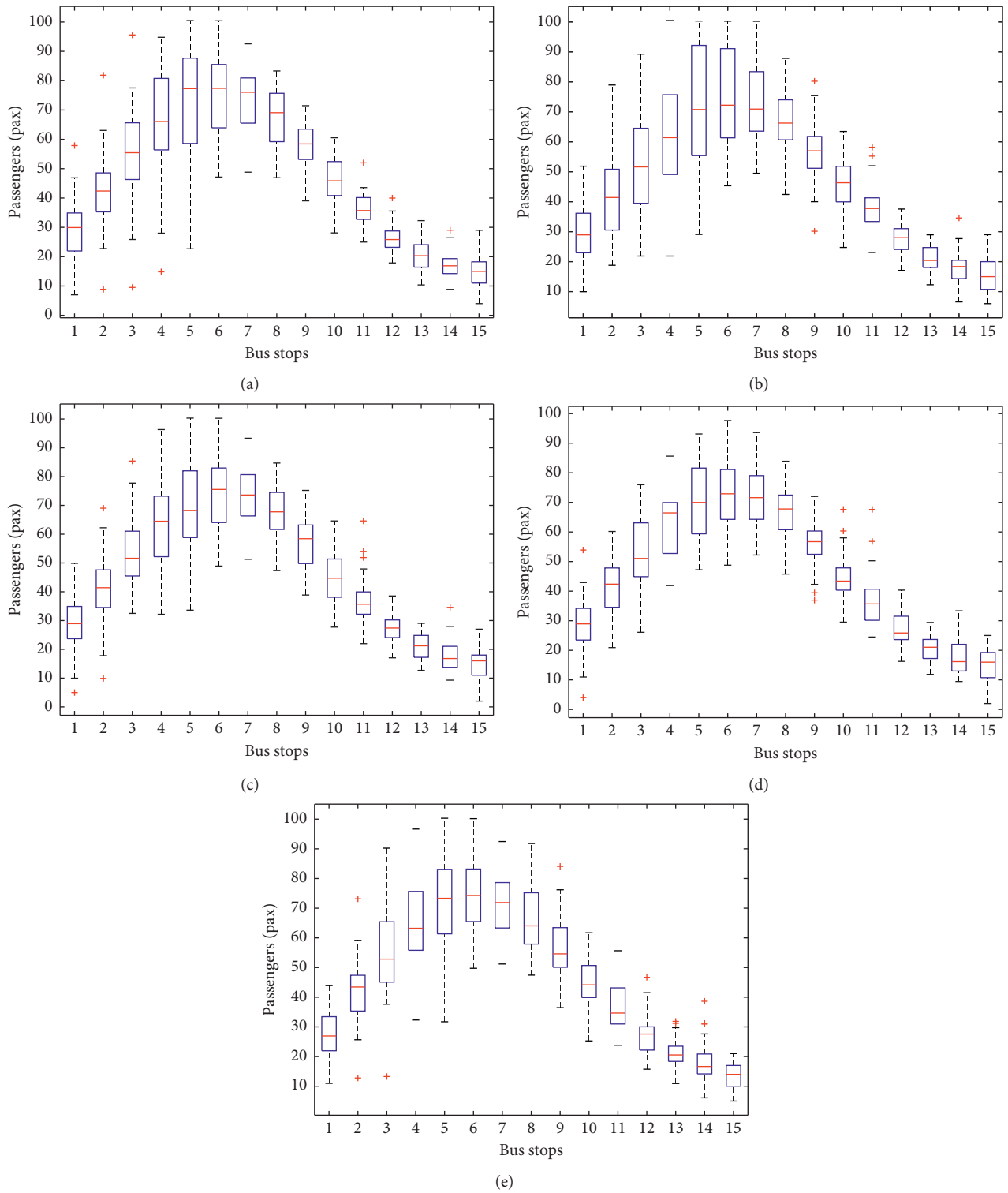


Figure 7: The passengers on buses under different number of CPS. (a) S1. (b) S3. (c) S4. (d) S7. (e) S10.

with the increasing of simulation times. Figure 5 shows that the performance result can be convergent, when the simulation repeated for 100 times. In this section, each test group has been repeated for 100 times. The relationship between average values of main indices and the simulation times is presented in Figure 7. According to this figure, it is obvious that the values tend to stablize and fluctuate a little. Therefore, the 100 times simulation can represent a general result for one fixed setting.

The indices values extracted from the test results are presented in Table 2, involving the total travel time, total waiting time, travel time on bus, average bus headway, STD

Table 2: The performance indices under variable number of CPS.

Group	Total travel time (s)	Total waiting time	Time on bus (s)	Average headway (s)	STD of headway (s)	Skipping times
S1	7310775	1252021	6058754	260.91	60.10	10.68
S2	7300540	1257292	6043249	261.53	63.14	5.71
S3	7341375	1261543	6079832	260.54	59.32	15.20
S4	7285720	1256509	6029211	259.03	48.16	21.81
S5	7351393	1261139	6090254	260.16	57.60	17.70
S6	7342283	1261077	6081206	260.10	56.35	18.27
S7	7267743	1256691	6011052	258.51	42.55	24.45
S8	7331101	1253641	6077460	259.86	55.95	18.13
S9	7375701	1245524	6130177	259.55	49.74	18.50
S10	7318832	1263890	6054942	258.69	44.96	25.38

of headway, and skipping times. The first three indices in Table 2 can reflect the travel time cost including at bus stops and on buses. The middle two indices average bus headways and STD of headways reflect the operating efficiency and the reliability of public transit, respectively. The last column means skipping times occurred at the CPS. According to the raw values shown in Table 2, the performance of the control method will not become monotone better along with the increasing of number of CPS. Therefore, in order to reflect both distribution and the number of CPS influence on the performance of the control method, more detailed comparisons are shown in Tables 3–6. The classification of the ten groups is shown in Table 7. For example, the group S1 has one CPS which locates at the front (head) section of the bus route, so that S1 locates at the first line and first column in Table 7. In addition, the group S3 is set as the normal control group, which has two CPS at bus stop 6 and bus stop 11. The relative difference can be calculated by S3 minus S_i and divided by S3. Therefore, if the values of the relative difference indices are positive, it means the value of S_3 is larger. Otherwise, the values of performance indices of the S3 are smaller.

As shown in Table 3, the values of total travel time in S1, S4, and S7 are smaller than that in S3, which means the CPS locating at the front section of the bus route is beneficial to reduce the total travel time for passengers. In addition, the saved travel time increased from 0.42% to 1.00% along with increasing of CPS number. The increasing tendency also occurred when the CPS are set at the middle section of the bus route, shown in the second line of Table 3. However, the improvement values are lower than those in line one of Table 3 where the CPS locate at the front of the bus route. On the contrary, the improvement values of travel time decreased along with the increasing of CPS number when the CPS locate at the rear of the bus route. The reasons will be given in detail in the following content. As shown in Tables 4 and 5, both the headways and their STD decreased along with the increasing of CPS number, which means the operation efficiency and the regularity improved. The shorter bus headway and more regular bus headway mean that the passengers have less waiting time. Because the number of the buses is fixed on the bus line, the shorter bus headways also mean faster travel speed for buses and shorter travel time for passengers. However, the total travel time does not always decrease along with the increasing of CPS number. The difference of these two conclusions comes from the additional waiting time of passengers due to stop-skipping. According to Table 6, along with the increasing of CPS number, stop-skipping control means are applied more times to control the bus, and more passengers are denied to board. Therefore, if the benefits brought by the shorter and more regular bus headways cannot cover the negative influence by boarding limits, the performance of the control method will be bad and worse along with the increasing of CPS number. Combined with Table 3, if CPS locates at front or middle section of bus route, the total travel time will decrease along with increasing of stop-skipping times. However, if the CPS locates at the behind section of bus route, the total travel time will increase along with increasing of stop-skipping times, causing worse control result. Therefore, this means, in the framework of self-equalizing control concept, the distribution of the CPS has more influence on the performance of control method. In order to discover the total and distribution of waiting time for passengers, the box figures of passengers at each bus stop are presented in Figure 8.

Table 3: Travel time relative difference compared with group S2.

Travel time	1	3	4	5
Front (%)	0.42	0.76	1.00	—
Middle (%)	—	−0.14	0.14	0.31
Behind (%)	0.56	−0.01	−0.47	—

Table 4: Bus headways relative difference compared with group S2.

Bus headways	1	3	4	5
Front (%)	−0.14	0.58	0.78	—
Middle (%)	—	0.15	0.26	0.71
Behind (%)	−0.38	0.17	0.38	—

Table 5: STD of bus headways relative difference compared with group S2.

STD of headways	1	3	4	5
Front (%)	−1.32	18.81	28.27	—
Middle (%)	—	2.91	5.69	24.22
Behind (%)	−6.43	5.00	16.15	—

Table 6: Stop-skipping times relative difference compared with group S2.

Stop-skipping times	1	3	4	5
Front (%)	29.75	−43.45	−60.83	—
Middle (%)	—	−16.43	−19.27	−66.96
Behind (%)	62.43	−20.15	−21.70	—

Table 7: Classification of CPS number and distribution.

Number of CPS	1	3	4	5
Front	S1	S4	S7	—
Middle	—	S5	S8	S10
Behind	S2	S6	S9	—

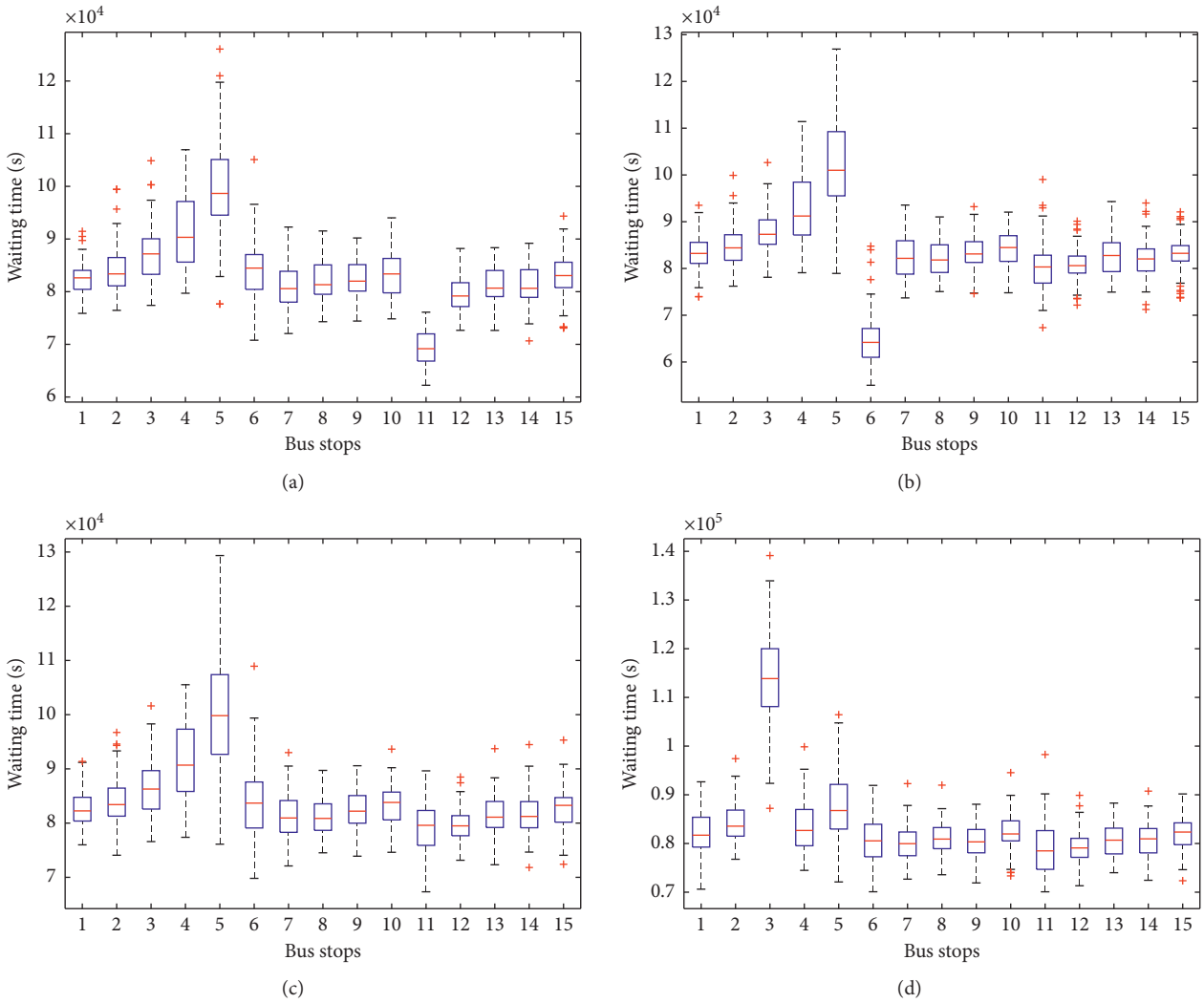


Figure 8: Continued.

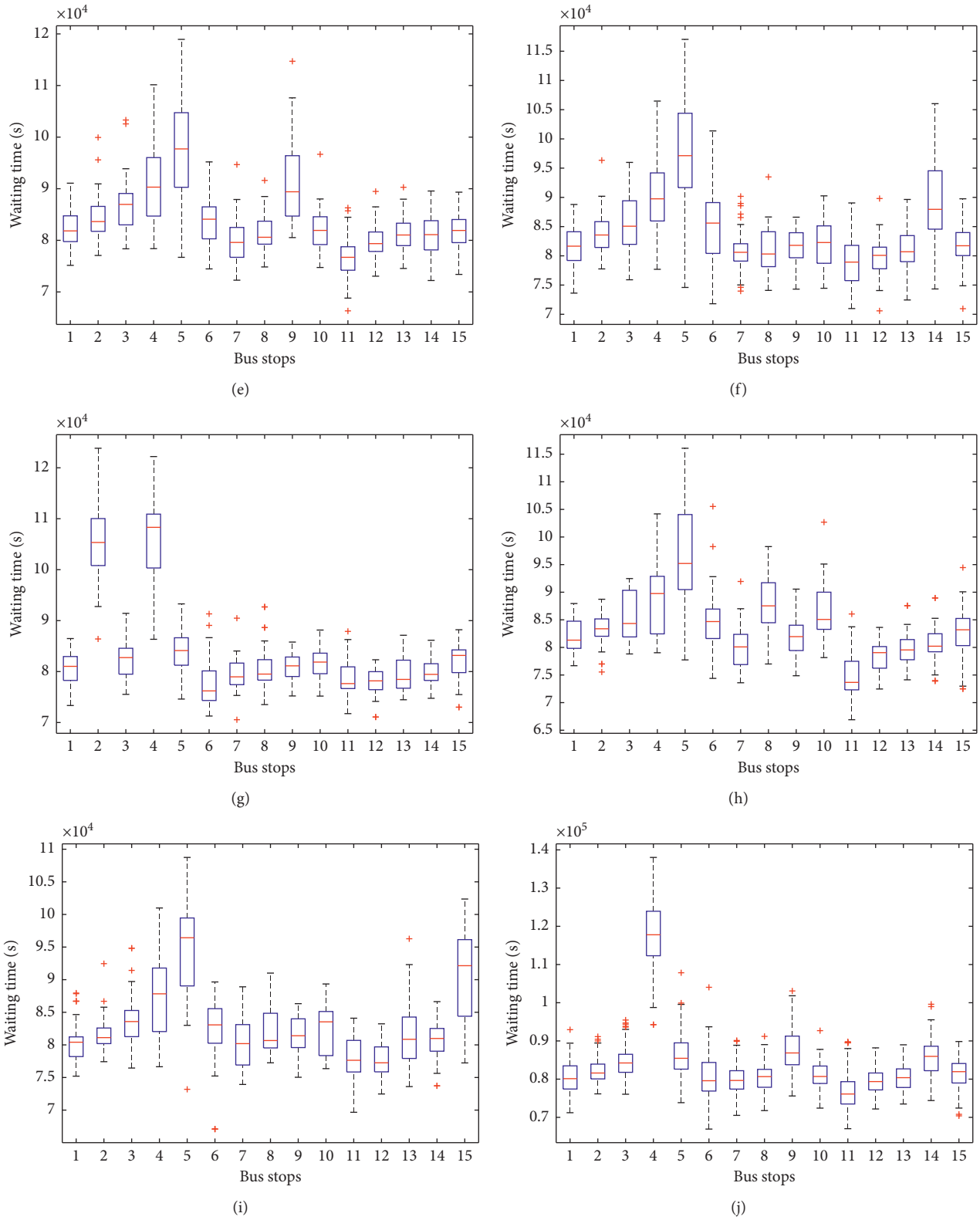


Figure 8: Waiting time of passengers at bus stops. (a) S1. (b) S2. (c) S3. (d) S4. (e) S5. (f) S6. (g) S7. (h) S8. (i) S9. (j) S10.

Figure 8(a) presents the waiting time of passengers at bus stops when the CPH locate at bus stop 6 and bus stop11, and the CPS locates at bus stop 6. At beginning of the five bus stops from 1 to 5, the waiting time increased gradually; at bus stop 5 the waiting time was the largest. The first reason is that the bus headways become more chaotic gradually when running at the bus route, and unequal bus headways lead to more waiting time. However, this is not the ultimate reason

that the waiting time will not change so greatly if only affected by the unregular bus headways. The other important reason for large waiting time at bus stop 5 is that the passengers waiting at the bus stop cannot board because of capacity limit of the bus. According to Figure 7(a), the maximum number of passengers on bus has reached at the capacity of the bus (100 pax) while the minimum number of passengers on bus is only a little more than 20 pax. Therefore, in the group S1, the bus capacity is not well utilized, so some buses are relatively full while other are relatively empty because of unequal bus headways. Therefore, the bus headways should be equalized before the buses encounter busy section of the bus route where the buses are relatively full.

Therefore, the groups S1, S4, and S7 perform best because the bus headways are more equal before arriving at bus stop 5. Combining Figure 7(d) with Figure 8(g), the bus headways are quite equal before arriving at bus stop 5, and the number of passengers on each bus is relatively equal too. Therefore, the travel time on buses and the waiting time at bus stops of passengers are relatively small, and the total travel time is the smallest among these 10 groups.

The groups S5, S8, and S10 set the CPS at the bus stops of the middle bus line. Although the bus headway can be relatively equal at the busy section of bus route, the stop-skipping may lead to the resisted boarding passengers still not being able to board next arriving bus, because all the buses at the busy section are relatively full. In addition, the bunching tendency will be severe when the influence of initial disturbance is not eliminated timely. Therefore, the performance of control method in these group is worse than the "front groups."

The groups S2, S6, and S9 set the CPS at the bus stops of the bus line rear. The test results are worse than the "middle groups," because the bus headways are not equal when the buses arrived at bus stop 5. In addition, the stop-skipping control means increasing the waiting time of passengers. Therefore, the benefits brought by the equal bus headways are not large enough to cover the additional waiting time because of stop-skipping. The negative influence will be larger along with increasing of CPS number, while the equal bus headways are not benefit to the passengers most needed.

Therefore, several conclusions can be obtained according to the analysis mentioned above. The CPS should be located before the busy section of bus route, and the bus capacity can be fully utilized and avoid additional waiting time because of bus capacity limit. In addition, stop-skipping control means can equalize the bus headways and adjust the passengers on successive buses making the passengers on buses more equal. The CPS should be set at the bus stops where passengers' demand is relatively small, avoiding part of the passengers that cannot board next arriving bus again if the number of resisting passengers is relatively large and reducing the number of passengers who are unsatisfied because of stop-skipping control means.

6. Conclusion

In this paper, the coordinated bus headway control method is designed based on the adaptive self-equalized concept. Furthermore, a procedure is designed to select the bus that should be skipping or held as well as its corresponding holding time. The discrete operation system is formulated, imbedding the coordinated control method, and considers the limitation of the bus capacity. However, the coordinated control method is not the core contribution of this paper. The main work of this paper is to discuss the setting of CPS to obtain better performance of the coordinated control method based on the self-equalizing bus headway control strategy. Therefore, four main indices are formulated including the waiting time of passengers at bus stops, the travel time on buses, the STD of bus headways, and the stop-skipping times.

In Numerical Analysis, a set of numerical tests is conducted under variable CPS settings. Ten groups are formulated representing different number and distribution of CPS. According to the test results of these 10 groups, several conclusions can be obtained. The control performance in groups S1, S4, and S7 is the best among these 10 groups, which means the CPS should be set at the beginning of the bus route, before the busy section of bus route, and the passengers on each bus are more equal, while less passengers are limited by the bus capacity. The increasing number of CPS will make the stop-skipping times larger, although the saved total travel times are larger as well. The distribution of the control points has larger influence on the performance of coordinated control method. If the distribution of the control points is not well set, a larger number of control points have more negative effects on the public transit in terms of total travel time of passengers. In addition, an extended conclusion can be obtained: the CPS is not suitable set at the large demand bus stops, because the rejected passengers, due to stop-skipping, may miss next arriving bus when there are relatively full passengers on bus and relatively large number of passengers at the bus stop. Therefore, the waiting time will increase and the passengers who resisted twice will be greatly unsatisfied.

Although this paper discussed the CPS setting to obtain better performance of the control method by a set of numerical analyses, the CPH settings are not considered, and deeper insight and theoretical proof are not given. More detailed and precise experiment and theoretical proof will be presented in future work. In addition, the input of travel speed between two adjacent bus stops is assumed as instant, which can be formulated as a variable in the future work.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 71801153 and 71801149), Natural Science Foundation of Shanghai (Grant no.

18ZR1426200), and Shanghai Innovation Training Program (Grant nos. XJ2020124 and SH2020069).

References

[1] M. Dessouky, R. Hall, A. Nowroozi, and K. Mourikas, "Bus dispatching at timed transfer transit stations using bus tracking technology," *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 4, pp. 187–208, 1999.

[2] X. J. Eberlein, N. H. M. Wilson, and D. Bernstein, "The holding problem with real-time information available," *Transportation Science*, vol. 35, no. 1, pp. 1–18, 2001.

[3] C. F. Daganzo, "A cheap and resilient way to eliminate bus bunching," in *Proceedings of the 4th International Conference on Future Urban Transport*, Gothenburg, Sweden, 2009.

[4] J. J. Bartholdi III and D. D. Eisenstein, "A self-coördinating bus route to resist bus bunching," *Transportation Research Part B: Methodological*, vol. 46, no. 4, pp. 481–491, 2012.

[5] F. Delgado, J. C. Muñoz, and R. Giesen, "How much can holding and/or limiting boarding improve transit performance?" *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1202–1217, 2012.

[6] S.-X. He, "An anti-bunching strategy to improve bus schedule and headway reliability by making use of the available accurate information," *Computers & Industrial Engineering*, vol. 85, pp. 17–32, 2015.

[7] S.-X. He, J. Dong, S.-D. Liang, and P.-C. Yuan, "An approach to improve the operational stability of a bus line by adjusting bus speeds on the dedicated bus lanes," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 54–69, 2019.

[8] J. Argote-Cabanero, C. F. Daganzo, and J. W. Lynn, "Dynamic control of complex transit systems," *Transportation Research Part B: Methodological*, vol. 81, no. 1, pp. 146–160, 2015.

[9] L. Lynn, O. Cats, J. Gama, J. Mendes-Moreira, and J. F. de Sousa, "An online learning approach to eliminate Bus Bunching in real-time," *Applied Soft Computing*, vol. 47, pp. 460–482, 2016.

[10] W. Wu, R. Liu, and W. Jin, "Designing robust schedule coordination scheme for transit networks with safety control margins," *Transportation Research Part B: Methodological*, vol. 93, pp. 495–519, 2016.

[11] W. Wu, R. Liu, and W. Jin, "Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour," *Transportation Research Part B: Methodological*, vol. 104, pp. 175–197, 2017.

[12] W. Wu, R. Liu, and W. Jin, "Integrating bus holding control strategies and schedule recovery: simulation-based comparison and recommendation," *Journal of Advanced Transportation*, vol. 2018, Article ID 9407801, 13 pages, 2018.

[13] G. Laskaris, O. Cats, E. Jenelius, M. Rinaldi, and F. Viti, "Multiline holding based control for lines merging to a shared transit corridor," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1062–1095, 2019.

[14] C. F. Daganzo, "A headway-based approach to eliminate bus bunching: systematic analysis and comparisons," *Transportation Research Part B: Methodological*, vol. 43, no. 10, pp. 913–921, 2009.

[15] C. F. Daganzo and J. Pilachowski, "Reducing bunching with bus-to-bus cooperation," *Transportation Research Part B: Methodological*, vol. 45, no. 1, pp. 267–277, 2011.

[16] Y. Xuan, J. Argote, and C. F. Daganzo, "Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1831–1845, 2011.

[17] S. Liang, S. Zhao, C. Lu, and M. Ma, "A self-adaptive method to equalize headways: numerical analysis and comparison," *Transportation Research Part B: Methodological*, vol. 87, pp. 33–43, 2016.

[18] S. Zhang and H. K. Lo, "Two-way-looking self-equalizing headway control for bus operations," *Transportation Research Part B: Methodological*, vol. 110, pp. 280–301, 2018.

[19] S.-X. He, S.-D. Liang, J. Dong, D. Zhang, J.-J. He, and P.-C. Yuan, "A holding strategy to resist bus bunching with dynamic target headway," *Computers & Industrial Engineering*, vol. 140, Article ID 106237, 2020.

[20] S. Liang, H. Zhang, M. Ma, and S. He, "Sensitivity analysis of fleet size for dynamic headway-based control method performance in terms of passengers' experience," *Journal of Advanced Transportation*, vol. 2020, Article ID 5070347, 16 pages, 2020.

[21] W. Suh, K.-S. Chon, and S.-M. Rhee, "Effect of skip-stop policy on a Korean subway system," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1793, no. 1, pp. 33–39, 2002.

[22] A. Sun and M. Hickman, "The real-time stop-skipping problem," *Journal of Intelligent Transportation Systems*, vol. 9, no. 2, pp. 91–109, 2005.

[23] M. M. O. Sidi, S. Hammadi, S. Hayat, and P. Borne, "Urban transport network regulation and evaluation: a fuzzy evolutionary approach," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 2, pp. 309–318, 2008.

[24] C. E. Cortés, D. Sáez, F. Milla, A. Núñez, and M. Riquelme, "Hybrid predictive control for real-time optimization of public transport systems' operations based on evolutionary multi-objective optimization," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 757–769, 2010.

[25] D. Sáez, C. E. Cortés, F. Milla, A. Núñez, A. Tirachini, and M. Riquelme, "Hybrid predictive control strategy for a public transport system with uncertain demand," *Transportmetrica*, vol. 8, no. 1, pp. 61–86, 2012.

[26] F. Delgado, J. C. Muñoz, R. Giesen, and A. Cipriano, "Real-time control of buses in a transit corridor based on vehicle holding and boarding limits," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2090, no. 1, pp. 59–67, 2009.

[27] H. Zhang, S. Zhao, Y. Cao, H. Liu, and S. Liang, "Real-time integrated limited-stop and short-turning bus control with stochastic travel time," *Journal of Advanced Transportation*, vol. 2017, Article ID 2960728, 9 pages, 2017.

[28] H. Zhang, S. Liang, S. Zhao, and Q. Shang, "Advantages of bus stop skipping and holding control in reducing schedule deviation," *Proceedings of the Institution of Civil Engineers—Municipal Engineer*, pp. 1–10, 2019.

[29] W. Wu, R. Liu, W. Jin, and C. Ma, "Simulation-based robust optimization of limited-stop bus service with vehicle overtaking and dynamics: a response surface methodology," *Transportation Research Part E: Logistics and Transportation Review*, vol. 130, pp. 61–81, 2019.

[30] W. Wu, R. Liu, W. Jin, and C. Ma, "Stochastic bus schedule coordination considering demand assignment and rerouting of passengers," *Transportation Research Part B: Methodological*, vol. 121, pp. 275–303, 2019.

[31] S. Liang, M. Ma, and S. He, "Multiobjective optimal formulations for bus fleet size of public transit under headway-based holding control," *Journal of Advanced Transportation*, vol. 2019, Article ID 2452348, 14 pages, 2019.

[32] S. Liang, M. Ma, S. He, H. Zhang, and P. Yuan, "Coordinated control method to self-equalize bus headways: an analytical method," *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 1175–1202, 2019.

[33] Y. Bie, X. Xiong, Y. Yan, and X. Qu, "Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 4–25, 2020.

[34] Y. Bie, D. Wang, and H. Qi, "Prediction model of bus arrival time at signalized intersection using GPS data," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 12–20, 2012.

[35] Y. Bie, X. Gong, and Z. Liu, "Time of day intervals partition for bus schedule using GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 443–456, 2015.

[36] M. Ma and S. Liang, "An integrated control method based on the priority of ways in a freeway network," *Transactions of the Institute of Measurement and Control*, vol. 40, no. 3, pp. 843–852, 2018.

[37] M. Ma and S. Liang, "An optimization approach for freeway network coordinated traffic control and route guidance," *PLoS One*, vol. 13, no. 9, Article ID e0204255, 2018.

[38] H. Yu, D. Chen, Z. Wu, X. Ma, and Y. Wang, "Headway-based bus bunching prediction using transit smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 72, pp. 45–59, 2016.

[39] H. Yu, Z. Wu, D. Chen, and X. Ma, "Probabilistic prediction of bus headway using relevance vector machine regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1772–1781, 2017.

[40] S. Liang, M. Ma, S. He, and H. Zhang, "Short-term passenger flow prediction in urban public transport: kalman filtering combined *K*-nearest neighbor approach," *IEEE Access*, vol. 7, pp. 120937–120949, 2019.

[41] G. Ma, M. Ma, S. Liang, Y. Wang, and Y. Zhang, "An improved car-following model accounting for the time-delayed velocity difference and backward looking effect," *Communications in Nonlinear Science and Numerical Simulation*, vol. 85, Article ID 105221, 2020.

[42] H. X. Liu, X. Wu, W. Ma, and H. Hu, "Real-time queue length estimation for congested signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 4, pp. 412–427, 2009.

[43] S. Liang, M. Ma, S. He, H. Zhang, and Z. Tang, "Influence of bus stop location on traffic flow," *Proceedings of the Institution of Civil Engineers—Municipal Engineer*, pp. 1–8, 2019.

[44] M. Ma, S. Liang, H. Guo, and J. Yang, "Short-term traffic flow prediction using a self-adaptive two-dimensional forecasting method," *Advances in Mechanical Engineering*, vol. 9, no. 8, 2017.

[45] M. Ma, S. Liang, and Y. Qin, "A bidirectional searching strategy to improve data quality based on *K*-nearest neighbor approach," *Symmetry*, vol. 11, no. 6, p. 815, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 4294195, 18 pages
https://doi.org/10.1155/2020/4294195

WILEY | Hindawi

Research Article

A Matheuristic Iterative Approach for Profit-Oriented Line Planning Applied to the Chinese High-Speed Railway Network

Di Liu,[1,2,3] Javier Durán Micco,[3] Gongyuan Lu,[1,2] Qiyuan Peng,[1,2] Jia Ning,[1,2] and Pieter Vansteenwegen[3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China*
[3]*KU Leuven Mobility Research Center-CIB, KU Leuven, Celestijnenlaan 300-box 2422, Leuven 3001, Belgium*

Correspondence should be addressed to Qiyuan Peng; qiyuan-peng@swjtu.edu.cn

Received 25 October 2019; Revised 13 March 2020; Accepted 26 May 2020; Published 1 July 2020

Academic Editor: Tao Liu

In this paper, a matheuristic iterative approach (MHIA) is proposed to solve the line planning problem, also called network design problem, and frequency setting on the Chinese high-speed railway network. Our optimization model integrates the cost-oriented and passenger-oriented objectives into a profit-oriented objective. Therefore, the passenger travel time is incorporated in the ticket price using a travel time value. As a result, transfers and detours will result in lower ticket prices and thus lower revenues for the operator. When evaluating the performance of a given line plan, the way in which passengers will travel through the network needs to be modelled. This passenger assignment is typically a time-consuming calculation. The proposed line planning approach iteratively improves the line plan using easy-to-determine indicators. During the process, a mixed integer linear programming model addresses the passenger assignment and optimizes the frequency setting in order to maximise the operational profit. Extensive computational experiments are executed to show the effectiveness of the proposed approach to deal with the real-world railway network line planning problem. Through extensive computational experiments on the small example network and real-world-based instances, the results show that the proposed model can improve the profits by 22.4% on average comparing to their initial solutions. When comparing to an alternative iterative approach, our proposed method has advantage of obtaining high quality of solutions by improving the profit 10.8% on average. For small, medium, and large size networks, the obtained results are close to the optimal solutions, when available.

1. Introduction

The Chinese high-speed railway (HSR) network has developed rapidly during the past decade. More railway lines will be constructed in 2020 to accomplish a comprehensive connection between 80% of the cities of China. Currently, the basic backbone of the HSR network contains 4 "vertical" and 4 "horizontal" tracks (4V4H). The practical HSR operation in China is different from Europe and Japan because, in China, a large number of long-distance HSR trains operate every day to satisfy as many passenger travel demand as possible. However, the average passenger travel distance is usually much shorter compared to the HSR line lengths. For instance, the average passenger travel distance on the two main HSR lines is about 558 km (Beijing-Guangzhou HSR) and 621 km (Beijing-Shanghai HSR), while the lengths of the lines are 2281 km and 1318 km, respectively [1]. This might lead to the inefficient use of railway resources such as train capacities and line capacities. For example, the average passenger load factor of HSR trains is less than 40% in some extreme cases [2].

The line planning for the HSR network is a complicated task because of the large-scale size, the high transportation demand, and the limited network capacity. Developing an efficient line plan to improve the whole network's operational performances is becoming urgent.

This paper aims to design a line plan for the 4V4H HSR network and to determine the frequency of the lines, optimizing operational cost and passenger travel time, while considering transfers when necessary. In order to obtain this, a profit-oriented objective function is applied. When the passenger demand (or potential) is considered as given and fixed, the operator's profit is determined by the operator costs and the revenue from selling train tickets. The (variable) operator costs are determined by the lines that are operated. The selling price of the train tickets is assumed to decrease when passengers need to make a transfer or detour to reach their destination. This is explained in detail in Section 3.3. Therefore, a trade-off will have to be made during line planning between, on the one hand, operating more and longer lines and, on the other hand, transfers and longer trips for the passengers.

This paper proposes an iterative approach combined with a mixed integer linear programming (MILP) model for maximizing the operator's profit during line planning. The iterative approach aims to determine better lines by heuristically modifying the current set of lines based on a fast evaluation of the current line plan. The MILP optimizes the frequency setting of the lines based on the expected routes the passengers will take, the so-called "passenger assignment" (or "transit assignment"). The two stages are optimized iteratively by what we call a matheuristic iterative approach.

The detailed contributions of this paper are as follows:

(i) A profit-oriented objective is proposed using a time value parameter in order to consider the travel time in the ticket price.

(ii) A matheuristic iterative approach is designed to solve the line planning problem.

(iii) Different local search improvements are considered to improve the current set of lines, such as extending a line, reducing a line, inserting a line, and removing a line. Fast and heuristic evaluation methods are designed to choose the most promising neighbourhood solution in order to obtain a better line plan.

(iv) MILP optimizes the frequency setting of the lines based on the expected passenger assignment.

(v) An alternative solution approach is also developed in order to illustrate the effectiveness of our approach.

(vi) A number of benchmark instances of different sizes are designed and made available together with detailed information about the best available solutions.

The remainder of this paper is structured as follows. In Section 2, the existing literature concerning the LPP is reviewed. After that, we present a mathematical model to define our profit-oriented line planning problem in detail in Section 3. This model will also be used to solve the passenger assignment and frequency setting problem. In Section 4, our matheuristic iterative approach is proposed. Several case studies and numerical experiments are shown in Section 5 to evaluate the performance and effectiveness of the proposed approach. Finally, the performance of our approach, our results for the Chinese HSR network, and our further work are summarized in Section 6.

2. Literature Review

The planning process in public transportation is typically divided into consecutive planning phases. Desaulniers and Hickman [3] consider network design, i.e., building the infrastructure, as the first phase, usually followed by line planning, timetabling, and then vehicle and crew scheduling. During operations, disturbances and disruptions might occur. Therefore, real-time rescheduling is required in order to minimize passenger inconvenience. As a crucial component of public transportation planning, the line planning problem (LPP) has attracted more and more attention recently [4–7]. Basically, the LPP decides which stops will be served by which line and in which order. Then, the frequency setting is about determining how often each line is operated. Many different variants of line planning, with different assumptions and objectives, are available in the state of the art. We will discuss a selection of the most relevant papers in this section and define the variants we will tackle mathematically in the next section.

Canca et al. [7] probably describe a problem closest to the problem discussed in this paper. However, the main differences are the components of the objective function and the planned time period. In [7], the objective function consists of ticket revenue, operational cost, and network infrastructure construction cost which are all based on the operator's point of view. Conversely, our model combines passenger-oriented and cost-oriented objectives into a profit-oriented objective. The method we used in our model considers a ticket price that depends on the passenger travel time which is not included in [7]. In terms of planning period, Canca et al. [7] consider revenue over a long period of time (i.e., years), while the LPP in this paper considers a much shorter duration (i.e., per day). Since the passenger demand scenario per day used in our model represents the regular pattern of the demand during a long period of time, our objective represents the profit over a longer period of time, typically three months, six months, or a year. If also the investment of building railway infrastructure is considered, as in [7], the considered time horizon is typically multiple years at least.

When comparing to [7], the similar components in the objective function are the revenue function, the variable operation cost, and the acquisition cost (i.e., fixed cost in our model). Our revenue function considers a ticket price that depends on the travel time, and this is not considered in [7]. Besides, the variable cost in both models include operation costs related to the line length, but crew cost is added in [7]. The crew cost shown in [7] is related to the line (i.e., frequency) and the yearly crew cost per train. We consider the crew cost as a fixed cost per train in our model. In addition, the acquisition cost in [7] formulates the purchase of each train model, while the fixed cost in our MILP includes the

depreciation expenses, material expense, fuel expense, crew cost, and other related cost per train.

In summary, our model not only considers ticket revenue but also includes the passenger travel time into the profit-oriented objective so as to minimize the passenger travel time and to make a better trade-off between the cost-oriented and passenger-oriented objectives. We consider this model to be more useful for profit-oriented line planning in practice.

Line planning models can be divided into two groups according to the objective function: a cost-oriented objective and a passenger-oriented objective. For the cost-oriented line planning problems, the objective function aims to minimize the operational cost [8–11]. The passenger-oriented line planning problems focus on maximizing the number of direct travellers [12, 13] or reducing the passenger travel time and/or the number of transfers [14, 15]. In those problems, the operator costs are considered as constraints, such as a limited number of lines. This is confirmed by Nachtigall and Jerosch [16], who propose that the cost-oriented objective and passenger-oriented objective can be considered by transforming one of them into constraints.

Obviously, some approaches try to combine the two aspects into a single problem. Pfetsch and Borndörfer [17] presented a weighted sum of the cost-oriented and passenger-oriented objective function. Rosalia [18] proposed an approach to optimize the operational cost and passenger travel time iteratively on a city road network.

A crucial challenge in passenger-oriented line planning is that, for evaluating the performance of the line plan, the passenger route choice or passenger assignment needs to be modelled. This will also determine the number of passengers on each line. This leads to a bilevel optimization problem which solves the line planning problem on the upper level and optimizes passenger assignment on the lower level [1, 5]. Friedrich et al. [19] investigated a cost-oriented line planning model with a passenger assignment evaluation process. Since the objective is cost-oriented, the solutions focus more on the operational cost, which has a negative effect on the service quality. Borndörfer and Karbstein [20] integrated line planning and passenger routing optimization by using a direct connection approach, which encourages direct connections and penalizes transfers. In addition, Karbstein [21] applied a variant of the 2-terminal Steiner connectivity problem to handle the transfers when integrating line planning and passenger routing. The complexity of the integrated line planning and passenger routing is investigated in Schmidt and Schöbel [22]. It is shown that the resulting problem is NP-hard even in very special cases.

When it comes to solving the LPP, an early approach uses a skeleton model, described by Silman et al. [23], which assembles routes from short pieces iteratively. After adding stops, the short pieces are connected by the shortest paths to form the line plan. Many approaches to line planning assume that a limited pool of possible lines is available or calculated beforehand, e.g., [14, 19]. Many models and algorithms to construct the line pool have been proposed and can be found in Kepaptsoglou and Karlaftis [24]. Other approaches to solve the LPP construct and modify lines instead of using a pool of lines [25–28]. These methods usually consider a (minimum and) maximum line length for each line. Considering heuristic approaches in the LPP becomes a tendency to reduce computation time while promising high quality solutions. The analysis of different heuristic methods applied on urban line planning is studied by Ahmed et al. [29]. Schmid [30] decomposed the bus rapid transit line planning using a large neighbourhood search to calculate the line planning design subproblem and using a linear programming model to obtain the results of passenger assignment and frequency setting. Several instances considering a single corridor are tested to show the efficiency of the proposed approach. Goerigk and Schmidt [31] used a bilevel optimization to model the line planning with passenger route choice and proposed two different techniques, binary variables and "big-M-constraints," to transfer the bilevel model into a single level model. But when the instance becomes large, say 250 stations, a genetic algorithm is required and performs well in determining a trade-off between passenger travel time and operational costs. In [7], an ALNS metaheuristic method is proposed to solve the line planning problem in railway rapid transit. At each iteration, a branch-and-cut algorithm is called to solve the passenger assignment given the results of operation information such as frequencies, train types, and fleet sizes. More details on the LPP can be found in the overviews by Schöbel [32] and Schmidt [33].

In this paper, instead of choosing between a cost-oriented and a customer-oriented objective, we propose a profit-oriented line planning model which maximizes the ticket price income minus the operational cost. The ticket price (and thus the operator revenues) is reduced when passengers need a transfer or a detour and have no direct train from their origin to their destination. This is to compensate a passenger facing a detour or transfer and an incentive for the railway operator to offer direct services as much as possible. Moreover, the operational costs consider fixed and length-dependent costs for operating the different lines. The profit-oriented line planning also turns the bilevel problem of line planning and passenger route choice into a single level problem. We have published preliminary results for profit-oriented line planning in a conference paper [34]. However, in that paper, passenger assignment is done heuristically, while this paper applies MILP to optimize the passenger assignment. Moreover, in the current paper, a better structure for implementing the local search operators allows to further improve the performance of the solution approach.

3. Profit-Oriented Line Planning

This section starts by discussing the assumptions of the line planning problem considered in this paper as well as the input data required. Then, the problem is defined mathematically with a mixed integer linear programming model. Finally, a small example network is introduced to illustrate the profit-oriented line planning problem.

3.1. Assumptions and Input Data. In this paper, the proposed profit-oriented line planning focuses on optimizing the line plan while considering passenger assignment and line plan design in a single model. The integrated model aims to make a trade-off between a cost-oriented objective, related to the number and length of the lines operated, and the passenger-oriented objective, related to minimising the travel inconvenience. This travel inconvenience is defined here as the additional travel time compared to the travel time of having a direct connection along the shortest path in the infrastructure network. In order to formulate the integrated model, the following assumptions are made throughout this paper:

(i) Stopping pattern: since only major stations are considered as nodes in the network, which attract the majority of the HSR passenger demand and are the backbone of the HSR network, the stopping pattern of the line plan is considered as an all-stop pattern for these major stations. The passenger demand of small stations can be assigned to the major stations in a precalculation phase.

(ii) Demand: passenger demand is assumed symmetrically. All demand in the network must be served with at most two transfers. In this network of limited size (only considering the major stations), two transfers should be more than enough. Transfers are penalized by a penalty time value.

(iii) Train type: trains are considered homogeneous, i.e., a double train set with 1000 seats, and its operation speed is 300 km/h. Including trains with different speeds is considered as future work.

(iv) Passenger route choice: passengers will always choose the shortest travel time path no matter what the price of the path is. Passengers of the high-speed railway normally pay more attention to the travel time rather than the ticket price. Moreover, many research papers on railway line planning [12, 15, 25, 31] assume that passengers travel according to their shortest path (with or without transfers).

(v) Line attributes: there is no limitation on the line length considered and lines can start and end in any station. Each line operates in both directions.

(vi) Feasible lines: all paths on the infrastructure network are allowed to be lines (except for cycles) and the possible passenger travel paths or train lines are not fixed in advance but determined during the line planning process. We consider no limitation on the number of lines (or their frequency) that can use a certain edge, but we take into account the capacity of the trains.

Actually, there are two types of trains operating on HSR. However, only few papers in literature [1, 7, 11, 34] consider a heterogeneous fleet during the line planning stage. Therefore, we will consider the different types of trains in future work. Moreover, the trend of HSR is to operate a single train speed so that the possibility of improving the HSR capacity utilization can be enhanced in the near future.

When considering the shortest path assignment for passenger route choice, more realistic models [35–37] have been proposed, but the LPP is already very challenging even with this simplification [22]. The additional assumption that passenger will not select a longer path (with detours and/or transfers) to get a reduction on the ticket price is based on the corresponding behaviour of most high-speed train customers. We assume they buy a more expensive high-speed ticket in order to get a shorter trip. In future work, these assumptions can be relaxed.

This is considered as input:

(i) Passenger OD matrix: the number of passengers traveling between any origin station and destination station is given in a symmetrical OD matrix. The passenger OD matrix represents the daily passenger demand.

(ii) HSR network topology: the available stations (nodes) and tracks (edges) are fixed, and the length of each edge between two stations is known beforehand.

The line planning solution is represented as a set of lines associated with certain frequencies. A line consists of a sequence of nodes.

3.2. Mathematical Model for Profit-Oriented Line Planning. A MILP model is presented to address the profit-oriented line planning. In order to keep the model understandable (and solvable in reasonable time), it assumes that a limited pool of possible lines is given. So the model will determine which lines from the pool to operate and at which frequency. It should be noted that the LPP we are solving for the Chinese HSR, and also the approach presented in Section 4, does not require such a limited pool of lines and allow to operate any feasible line.

3.2.1. Variables and Notations. The physical network topology is considered as the undirected graph $G = (S, E)$. The node set is defined as $S = \{s_1, s_2 \ldots, s_{|S|}\}$ and represents the stations. The edge set is described as $E = \{e, e \in S \times S\}$ and represents the connections of two stations in the network. When solving the LPP, a train service network (TSN) (Fu et al. [1]) is also needed to take the transfer times into account and to depict the itineraries of passengers. This is also called the Change & Go network in Schöbel and Scholl [14]. In this network, each station is duplicated per line it serves. See Figure 1(c) for an example of TSN. In the model, $[s_i, l]$ indicates the duplication of station s_i online l.

Let $W = \{w_1, \ldots, w_{|W|}\} \subseteq S \times S$ be the set of OD pairs $w_i = (s_i^{\text{org}}, s_j^{\text{des}})$. The number of passengers for a certain OD pair is denoted as d_{w_i}. A pool of possible lines L_{cur} is given as input to the MILP model. The length k_l of line $l \in L_{\text{cur}}$ is assumed to be known.

These variables are used in the model:

Inc: the total operational income

Cos: the total operational cost

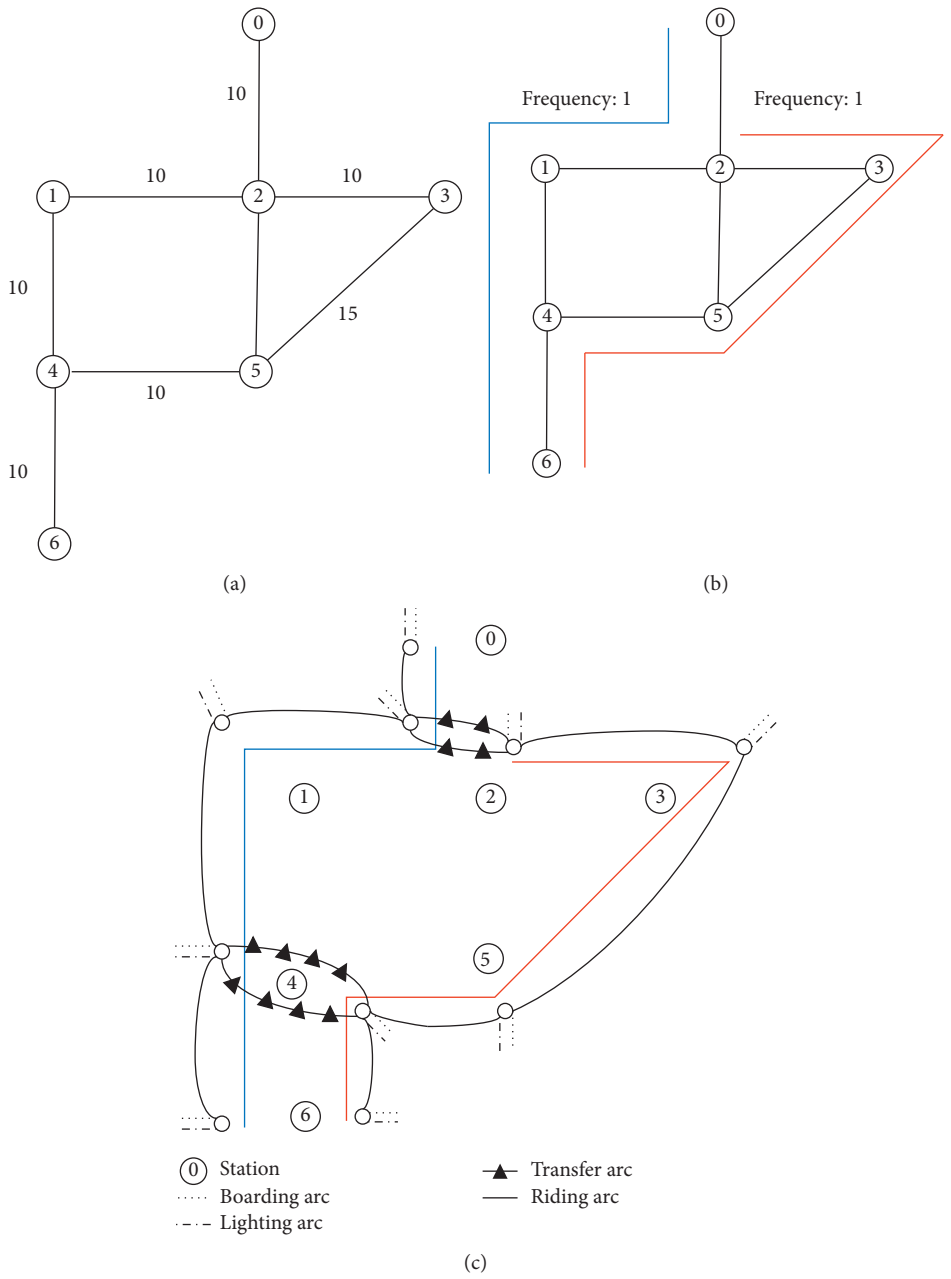


FIGURE 1: The infrastructure of the small example network (a); an example line plan with two lines (b); the corresponding TSN for this line plan (c).

Cap: the capacity of each train represented by the number of seats

L_{cur}: the pool of possible lines

C^{Fix}: the fixed cost for operating a line with frequency one

C^{Var}: the variable cost per line per kilometre for frequency one

IdeInc: the ideal income, if each passenger would have a direct connection on his/her shortest path in the physical network; i.e., IdeInc = $\sum_{w_i \in W} T^{\text{Phy}}_{w_i} * V_t * d_{w_i}$

d_{w_i}: the number of passengers for a certain OD pair w_i

$T^{\text{Phy}}_{w_i}$: the shortest path travel time of each OD pair w_i with respect to the physical network independent of the line plan

V_t: the travel time value (the ticket price per unit of time) to convert the passenger travel time into the ticket price

$T^{\text{TSN}}_{w_i}$: the shortest path travel time of each OD pair w_i on the TSN, including a fixed time penalty for each transfer

$T^{\text{TSN(dri)}}_{w_i}$: the total driving time and stopping time in $T^{\text{TSN}}_{w_i}$

$T_{w_i}^{\mathrm{TSN}(ts)}$: the total transfer time in $T_{w_i}^{\mathrm{TSN}}$

t_s: fixed stopping time at a station

t_{ts}: fixed transfer time between two lines at the same station

V_p: the penalty time value: the value of time for detours or transfers

The train service network (TSN) notations are as follows:

A_{dri}: set of driving arcs, $A_{\mathrm{dri}} = \{a = ([s_i, l], [s_j, l]),\ l \in L_{\mathrm{cur}}, s_i \text{ and } s_j \text{ in } S, \text{ index } i < j\}$; the cost of each driving arc equals the travel time of the edge, i.e., $t_{e(a)}$

$t_{e(a)}$: the driving time of arc a on edge e

A_{ts}: set of transfer arcs, $A_{ts} = \{a = ([s_i, l], [s_i, l']): l \neq l',\ s_i \epsilon l,\ s_i \epsilon l',\ s_i \text{ in } S\}$; the cost of each transfer arc is set to a fixed penalty cost; i.e., t_{ts}

A_{org}: set of origin arcs, $A_{\mathrm{org}} = \{a = (s_i^{\mathrm{org}}, [s_i, l]): s_i \text{ in } S\}$; the travel time cost of each origin arc is 0

A_{des}: set of destination arcs, $A_{\mathrm{dest}} = \{a = ([s_j, l], s_j^{\mathrm{des}}),\ s_j \text{ in } S\}$; the travel time cost of each destination arc is 0

$A(w)$: arc set $A = \left\{A_{\mathrm{dri}} \cup A_{ts} \cup A_{\mathrm{org}} \cup A_{\mathrm{des}}\right\}$, the arcs from A used by OD pair w_i

$A_{\mathrm{dri}}(l)$: the driving arcs from A_{dri} used by line l

$y_{s_i}^l$: this parameter equals 1 when station s_i is covered by line l; otherwise, 0

$y_{e(a)}^l$: this parameter equals 1 when edge e is covered by line l as arc a; otherwise, 0

The decision variables are as follows:

f_l: the frequency of line l (which can be zero if the line is not actually operated)

$x_a^{w_i}$: binary variable equals 1 when the passenger OD pair $w_i = (s_i^{\mathrm{org}}, s_j^{\mathrm{des}})$ uses arc a; otherwise, 0

3.2.2. Objective. The objective is to maximise the operational profit, which equals the difference between revenues and operational costs. The operational cost consists of a fixed cost per line (of frequency one) and a variable cost related to the length of the line and its frequency. By introducing the travel time value parameter, the passenger travel time is converted into operational income. Thus, the operational income can be formulated as the passenger total travel time multiplied with the travel time value and minus the penalties for transfers and detours:

$$\max Z = \mathrm{Inc} - \mathrm{Cos}, \tag{1}$$

$$\mathrm{Inc} = \mathrm{IdeInc} - \sum_{w_i \in W} \left(T_{w_i}^{\mathrm{TSN}} - T_{w_i}^{\mathrm{Phy}}\right) * V_p * d_{w_i}, \tag{2}$$

$$\mathrm{Cos} = \sum_{l \in L_{\mathrm{cur}}} \left(C^{\mathrm{Fix}} + C^{\mathrm{Var}} * k_l\right) * f_l. \tag{3}$$

Equation (2) gives specific information about the revenue considering the penalties. The left side of the minus is the ideal income, obtained when each passenger travels along the shortest path in the physical network, with a direct connection. It should be noted that the ideal income is a fixed value and could be omitted from the objective function. However, since the ideal income is part of the profit components, we want to include it in order to make the objective function more readable and understandable. The right side of the minus is the penalty for transfers and detours the passengers require in the TSN. This assumes that the ticket price is reduced to compensate for the discomfort of having a longer travel time (than the ideal shortest path). The operational cost is presented as equation (3), which is related to the number of lines and its associated frequencies. An example of the objective function calculation will be presented in Section 3.3.

3.2.3. Constraints. We assume that all the lines in the given pool of lines are selected, but some lines might have a frequency of zero. The constraints included in the MILP model are listed below:

$$y_{e(a)}^l \geq x_a^{w_i}, \quad \forall l \in L_{\mathrm{cur}}, a \in A_{\mathrm{dri}}, w_i \in W, \tag{4}$$

$$y_{s_i}^{l_k} + y_{s_i}^{l_j} \geq 2x_a^{w_i}, \quad k \neq j, \forall s_i \in S, a \in A_{ts}, w_i \in W, \tag{5}$$

$$y_{s_i}^l \geq x_a^{w_i}, \quad \forall s_i \in S, l \in L_{\mathrm{cur}}, a \in A_{\mathrm{org}} \cup A_{\mathrm{des}}, w_i \in W, \tag{6}$$

$$\sum_{a \in A_{\mathrm{org}} \cap A(w)} x_a^{w_i} = 1, \quad \forall w_i \in W, \tag{7}$$

$$\sum_{a \in \delta^+(s_i) \cap A(W)} x_a^{w_i} - \sum_{a \in \delta^-(s_i) \cap A(w)} x_a^{w_i} = 0, \quad \forall s_i \in S \backslash \left\{s_i^{\mathrm{org}}, s_j^{\mathrm{des}}\right\}, a \in A(w), w_i \in W, \tag{8}$$

$$\sum_{a \in A_{\mathrm{des}} \cap A(w)} x_a^{w_i} = 1, \quad \forall w_i \in W, \tag{9}$$

$$\sum_{a \in A_{ts} \cap A(W)} x_a^{w_i} \leq 2, \quad \forall w_i \in W, \tag{10}$$

$$\sum_{a \in A_{\text{dri}} \cap A(W)} x_a^{w_i} \left(t_{e(a)} + t_s\right) - t_s = T_{w_i}^{\text{TSN(dri)}}, \quad \forall w_i \in W, \tag{11}$$

$$\sum_{a \in A_{ts} \cap A(W)} x_a^{w_i} \left(t_{ts} - t_s\right) = T_{w_i}^{\text{TSN}(ts)}, \quad \forall w_i \in W, \tag{12}$$

$$T_{w_i}^{\text{TSN(dri)}} + T_{w_i}^{\text{TSN}(ts)} = T_{w_i}^{\text{TSN}}, \quad \forall w_i \in W, \tag{13}$$

$$\sum_{w_i \epsilon W} \left(x_a^{w_i} \times d_{w_i}\right) \leq f_l \times \text{Cap}, \quad \forall a \in A_{\text{dri}}(l), \, l \in L_{\text{cur}}, \tag{14}$$

$$f_l \in \mathbb{N}, \quad \forall l \in L_{\text{cur}}, \tag{15}$$

$$x_a^{w_i} \in \{0, 1\}, \quad \forall a \in A, \, w_i \in W. \tag{16}$$

Equations (4)–(6) assure the TSN construction based on a line plan. The three constraints indicate that only nodes and edges that are covered by the possible lines are considered as the nodes and arcs of the TSN; i.e., edges belonging to the lines can be selected as the driving arc of passenger routes (4) and nodes (stations) covered by more than one line are chosen as potential transfer arcs (5). Origin and destination arcs are connected to each node covered by a line in the TSN (6). These constraints assume that driving arc a in the TSN corresponds to edge e in the physical network (4) that transfer arc a corresponds to a transfer between l_i and l_j (5) and that station s_i is an endpoint of an origin or destination arc a (6).

Equations (7)–(9) are network flow conservation constraints, which require that each passenger OD pair should have a feasible path on the TSN based on the given line plan. Equations (7) and (9) assure that only one origin arc and destination arc can be selected for each passenger OD pair. On the contrary, for each passenger OD pair, there must be one origin arc and one destination arc. Equation (8) ensures the conservation of passenger flows on the intermediate nodes. If there is no travel path for any of the OD pairs or the number of transfers is more than two, the model will turn out to be infeasible.

Equation (10) limits the number of transfers for each OD pair to 2. Equations (11)–(13) calculate the actual travel time for each OD pair. Equations (11) and (12) compute the actual total driving time and total transfer time for a certain OD pair. Equation (14) ensures the capacity of a line on a driving arc is sufficient to meet the passenger demand on that driving arc; i.e., all passengers are served taking into account the capacity of the trains. The frequency of the train is calculated based on the number of passengers taking that train. Constraint (15) and constraint (16) are variable value constraints.

The main purpose of this model is to clarify the profit-oriented line planning problem considered in this paper. Moreover, this model will be used in the line planning approach we present in Section 4. However, the pool of possible lines L_{cur} considered in the MILP model above will be replaced by the lines of the line plan under evaluation. As a result, it will optimize the passenger assignment and the frequencies of the line plan under evaluation.

3.3. Small Example Network. Here, a small network is introduced in Figure 1(a) which will be used to illustrate the calculation of the objective function. In this small example, the number besides each arc corresponds to both the distance (km) and the travel time (min). In this example, the data of Table 1 are assumed to be given.

A brief explanation of the objective function calculation is presented based on the example line plan in Figure 1(b). In order to consider the transfers and depict the itinerary of passengers, the TSN (see Figure 1(c)) is constructed based on the given line plan in Figure 1(b). As indicated in the figure, both lines have frequency 1 per unit time (for example, 1 per hour or 1 per day). The operational cost of the blue line is $(15,000 + 150 \times (10 + 10 + 10 + 10)) \times 1 = 21,000$ RMB. For the red line, this is 21,750 RMB, so the total operational cost is 42,750 RMB. The revenue is calculated by the ideal income minus the penalties caused by transfers and detours. In order to simplify the calculation process, we assume that the passenger demand of each OD pair is 100. The ideal income is calculated as the price that all the passengers pay for their direct connections on their shortest paths. For instance, the ideal income of passengers from nodes 1 to 3 (the path (1, 2, and 3)) is $(10 + 1 + 10) \times 2.5 \times 100 = 5\,250$ RMB. The total ideal income can be calculated as 105,750 RMB. When computing the penalty fees of transfers and detours, the penalty time equals the difference between the actual travel time and the shortest possible travel time. For example, the penalty fee of passengers from node 1 to node 3, requiring a transfer in node 2 (instead of a stop), equals $((10 + 5 + 10) - (10 + 1 + 10)) \times 0.55 \times 100 = 220$ RMB. Another example includes the passengers from node 2 to

Table 1: Data of the small example network.

Name	Value	Unit
Transfer time penalty	5	min
Station stopping time	1	min
Travel time value	2.5	RMB/person, min
Penalty time value	0.55	RMB/person, min
Fixed line operating cost	15,000	RMB/train
Variable line operating cost	150	RMB/train, km
Train capacity	1000	Seats/train

node 5, requiring a detour through node 3. The penalty time value multiplied with the detour time and the passenger demand equals $((10+1+15)-10)\times 0.55\times 100 = 880$ RMB. The total penalty fee of the passengers with transfers and detours (OD pairs 0-3, 0-5, 1-3, 1-5, and 2-5) is 2640. The final profit of this line plan is $105,750-2640-42,750 = 60,360$ RMB.

4. A Matheuristic Iterative Approach for Line Planning

After generating an initial line plan, our approach decomposes the LPP into two subproblems, improving the design of the lines and the passenger assignment together with frequency setting. The two subproblems are optimized iteratively based on a heuristic evaluation of possible neighbourhood solutions in the first subproblem and a MILP model to address the second subproblem. We call this approach MHIA, and it is illustrated in Figure 2 and discussed in detail in the next sections.

4.1. Initial Line Plan Generation. The basic idea of the initial line plan generation is to select those lines that serve directly as much passenger demand as possible. According to the classical approach for line planning introduced in Bussieck et al. [12], we first search the shortest paths for each OD pair on the physical network by the Floyd–Warshall algorithm [38]. This results in a set of candidate lines. The initial line plan generation consists of two stages, namely, heuristic construction and repair, determining which lines to include in the initial solution. During the heuristic construction process, the line set is built by selecting one line at a time from the candidate lines until all the nodes are covered. The choice of the lines obeys the rule that the selected line serves the most passengers on their shortest paths without transfers, not only from the starting towards the ending station of the line but also from and towards all stations in between on that line. When selecting the next line to include, all passengers already served directly by the already selected lines are no longer considered. For example, the first line is selected based on the calculation of the direct passengers served by each line. Then, the passengers served by that line are eliminated from the OD matrix. The next line is then selected from the candidate lines based on the remaining unserved passengers. The selection procedure continues until all the nodes of the network are covered.

Since the construction stage focuses on direct connections, it does not guarantee that all passengers have a path and have less than two transfers. For this reason, the MILP model of the previous section is used to check the line plan feasibility by using the Floyd–Warshall algorithm to calculate all the shortest paths for the OD pairs. If there exists any OD pair for which no path with two or less transfers is available in the TSN, then the shortest path on the physical network of that OD pair is selected as a new line and added to the initial line plan.

To illustrate the initial line plan generation, we return to the small example network of Figure 1(a). The list with the shortest path of each passenger OD pair (in one direction) is given in Table 2.

We assume each OD pair corresponds to 100 passengers. Based on serving directly as much demand as possible, the first line will be (0, 2, 1, 4, 6) or (0, 2, 5, 4, 6), both serving 1000 passengers. Then, those served passengers are removed from the OD matrix, and the second line is selected from the rest of the paths in the same way. If we choose (0, 2, 1, 4, 6) as the first line, then (3, 5, 4, 6) becomes the second line with 500 passengers served directly. Now, all the nodes on the network are covered by the selected lines. After the feasibility check (travel path, number of transfers, and node coverage), the selected lines meet all the constraints and an initial line plan is generated.

4.2. Line Plan Evaluation and Modification. In this line plan modification and evaluation process, illustrated in Algorithm 1, two subproblems are considered: line plan design (modification) and passenger assignment and frequency setting (evaluation). Clearly, the passenger assignment is a crucial part of the line plan evaluation. However, passenger assignment typically requires a lot of calculation time, and the smallest change in the line plan could significantly change the passenger assignment and thus the passenger travel times. Therefore, we try to limit the number of times the passenger assignment is calculated. It would be, for instance, too time-consuming to calculate the passenger assignment for every possible modification to the current line considered below. Therefore, we try to improve the line plan based on easier-to-calculate heuristic evaluation indicators. Only for the most promising modifications, the passenger assignment and frequency setting are applied using the MILP model discussed in Section 3.

In the plan design or modification part, a framework with four modification operators is developed to iteratively improve the current line plan. The detailed modification and evaluation process are illustrated in Figure 2. The four modification operators are reducing a line (*Reduction*), extending a line (*Extension*), removing a line (*Removal*), and inserting a line (*Insertion*). *Reduction* and *Extension* work as intensification or improvement modifications, while *Removal* and *Insertion* are used as diversification of the search. Each type of modification leads to a set of possible line plans, the neighbourhood of the current line plan.

The *Reduction* neighbourhood of a current line plan contains all line plans where one terminal node of one line is

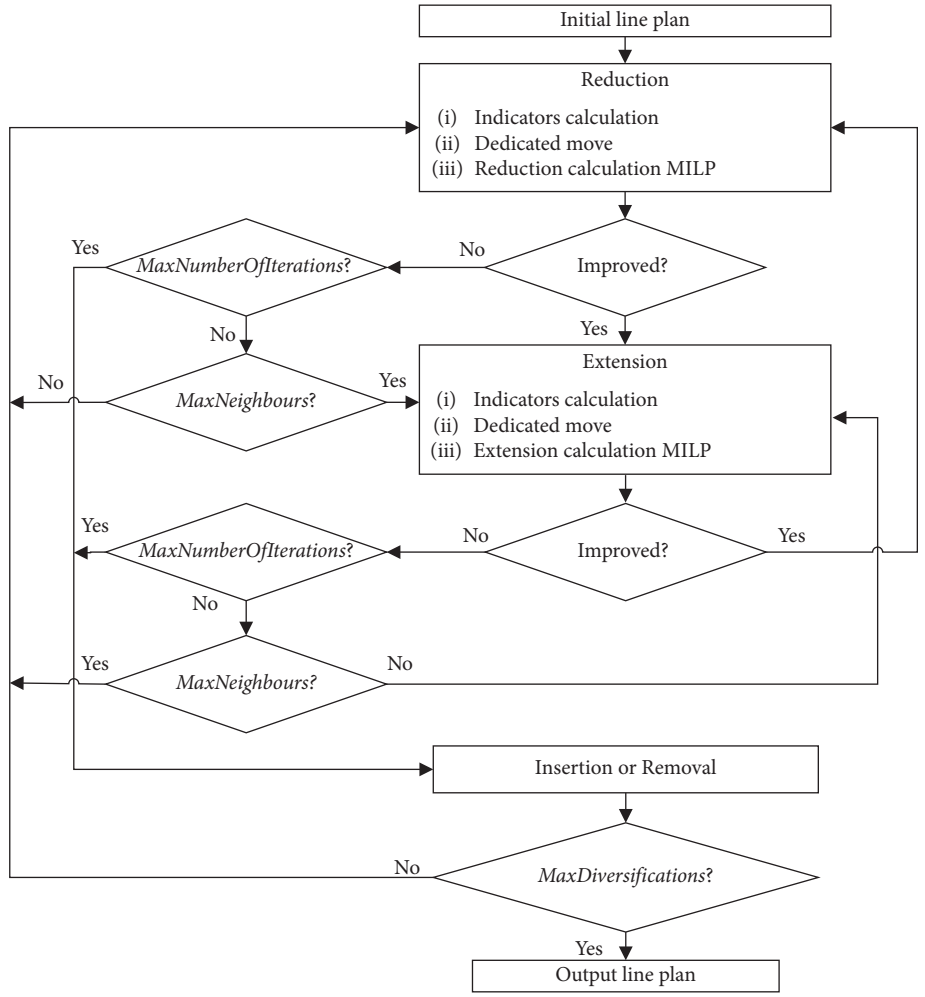


Figure 2: The process of MHIA for line planning optimization.

Table 2: The shortest paths for passenger OD pairs on the small example network.

O	D	The shortest paths
0	1	0, 2, 1
0	2	0, 2
0	3	0, 2, 3
0	4	0, 2, 1, 4/0, 2, 5, 4
0	5	0, 2, 5
0	6	0, 2, 1, 4, 6/0, 2, 5, 4, 6
1	2	1, 2
1	3	1, 2, 3
1	4	1, 4
1	5	1, 4, 5/1, 2, 5
1	6	1, 4, 6
2	3	2, 3
2	4	2, 1, 4/2, 5, 4
2	5	2, 5
2	6	2, 1, 4, 6/2, 5, 4, 6
3	4	3, 5, 4
3	5	3, 5
3	6	3, 5, 4, 6
4	5	4, 5
4	6	4, 6
5	6	5, 4, 6

removed. The *Extension* neighbourhood contains all line plans where one node, adjacent to a terminal node in the physical network, is added to one of the lines. When it comes to *Insertion*, each of the lines corresponding to OD pairs without a direct connection in the current line plan are considered to be inserted in the current line plan. For *Removal*, the neighbourhood contains all line plans where a line of the current line plan is removed. In each neighbourhood of the four operators, only feasible line plans are considered.

When considering *Reduction*, the load factors of terminal edges of each line are calculated as the evaluation indicator of the current line plan. The load factor is the actual passenger volume on the edge of a line divided by the total capacity (the number of seats) of the corresponding line. A low load factor might indicate a part of a line which is not profitable. The load factor is calculated for each terminal edge, and then, the terminal edge with the lowest load factor is removed from its line. The MILP model is then applied to optimize the frequencies and to optimally assign the passengers to the new line plan. Only when the total profit is actually increased by removing this edge, the new line plan is accepted and the algorithm continues with considering

```
(1) n_der = 0;
(2) Repeat
(3)   n_int = 0, n_red = 0, n_ext = 0;
(4)   Repeat
(5)     Reduction method
(6)       n_int = n_int + 1;
(7)       Select the L_nei and calculate the profit of L_nei by MILP;
(8)       if Z_nei > Z_cur do
(9)         L_cur ⟵ L_nei, Z_cur ⟵ Z_nei, n_red = 0;
(10)        if Z_nei > Z_best do
(11)          L_best ⟵ L_nei, Z_best ⟵ Z_nei;
(12)      else
(13)        n_red = n_red + 1;
(14)        if n_int > MaxNumberOfIterations
(15)          go to step 36;
(16)        else if n_red > MaxNeighbours
(17)          go to step 20;
(18)        else
(19)          go to step 5;
(20)    Extension method
(21)      n_int = n_int + 1;
(22)      Select the L_nei and calculate the profit of L_nei by MILP;
(23)      if Z_nei > Z_cur do
(24)        L_cur ⟵ L_nei, Z_cur ⟵ Z_nei, n_ext = 0;
(25)        if Z_nei > Z_best do
(26)          L_best ⟵ L_nei, Z_best ⟵ Z_nei;
(27)      else
(28)        n_ext = n_ext + 1;
(29)        if n_int > MaxNumberOfIterations
(30)          go to step 36
(31)        else if n_ext > MaxNeighbours
(32)        go to step 5;
(33)        else
(34)          go to step 20;
(35)  until the terminal conditions are satisfied
(36)  Disturb
(37)    Removal or Insertion
(38)      n_der = n_der + 1;
(39)      Select the L_nei and calculate the profit of L_nei by MILP;
(40)      L_cur ⟵ L_nei, Z_cur ⟵ Z_nei, go to step 3
(41) until n_der = Max Diversifications
(42) Output the best line plan and objective profit
```

Algorithm 1: Pseudocode of the MHIA of line planning.

Extension. However, when removing this edge would decrease the profit, the reduction is cancelled and the algorithm considers to remove the edge with the next lowest load factor. This continues until an edge is found that actually improves the profit or until the number of neighbours considered reaches a predefined maximum number (*MaxNeighbours*). In both cases, *Reduction* is ended, and *Extension* is considered.

Extension evaluates how many passengers that currently need a transfer could be transported directly due to extending a line with an additional edge. The edge that can provide the most additional direct connections is added. Again, the MILP model is applied to optimize the frequencies and optimally assign the passengers to the new line plan. Only when the total profit is actually increased by extending this line, the new line plan is accepted and the algorithm continues by going back to *Reduction*. However, when adding this edge would decrease the profit, the extension is cancelled and the algorithm considers to add the next most promising edge. As with *Reduction,* this continues until an edge is found that actually improves the profit or until the number of neighbours considered reaches a predefined maximum number (*MaxNeighbours*). In both cases, *Extension* is ended, and *Reduction* is considered again.

In order to limit the total computation time of *Reduction* and *Extension*, we explicitly limit the total number of times the MILP model is used. When the MILP model is used *MaxNumberOfIterations* times or when no more improvements can be found by *Reduction*

or *Extension*, the algorithm continues with the diversification phase.

In order to diversify the algorithm, two ways of disturbing the search are implemented as well: *Removal* and *Insertion*. First, however, when some lines have a frequency of zero after the frequency setting, these nonoperated lines will be removed. Then, the new line plan is selected randomly from all possible *Removal* and *Insertion* neighbourhood solutions. The solution of the disturbance is always accepted. The number of diversification iterations (*MaxDiversifications*) is fixed beforehand as a stopping criterion for the algorithm. After that, the best solution obtained during the search process will be presented as the final solution.

An alternative implementation would be to consider all the four moves at the same time. However, this would be too time-consuming and not efficient. So we choose the aforementioned operator execution order to optimize the line plan. The *Reduction* and *Extension* are used to search for the approximate local optimal solution by small changes in the input line plan. These two operators are crucial operators in the good performance of the algorithm and are considered as intensification moves. However, the *Insertion* or *Removal* is selected randomly to diversify the solution search space. These should be regarded as large changes to the input line plan.

Since the initial line plan is generated by choosing the lines that can serve as many direct passengers as possible, this typically leads to long lines. In order to avoid such negative effects of the initial line plan, the *Reduction* is applied first. After a limited number of iterations with *Reduction* and *Extension*, the random *Insertion* or *Removal* provides other search directions.

Compared to using large neighbourhood search (LNS) [35, 39], the approach proposed in this paper is different in the following aspects. In LNS, a temporary solution is given by first applying a destroy method and then a repair method. For the proposed approach in this paper, the temporary solution is given separately by *Reduction*, *Extension*, *Removal*, or *Insertion*. Specifically, the destroy method in LNS destructs randomly a part of the current solution, and then, the repair method reconstructs the destroyed part. However, our proposed approach based on four operators, respectively, modifies the current solution by a heuristic evaluation and selects the most promising one as the temporary solution.

The notation used in Algorithm 1 is as follows:

Z_{cur}: the profit of the current line plan

Z_{best}: the best profit among the calculated line plans

Z_{nei}: the profit of a neighbourhood solution of the current line plan

L_{nei}: the selected neighbourhood line plan

A_{undir}: the set of OD pairs of passengers with a longer travel time than their ideal travel time

n_{der}: the current number of *MaxDiversifications* iterations

n_{red}: the current number of *Reduction*

n_{ext}: the current number of *Extension*

n_{int}: the current number of intensification iterations

4.3. An Alternative Iterative Approach. It is very time-consuming to obtain an optimal passenger assignment and frequency setting using the mathematical model when the network becomes larger. This is mainly due to the exponentially increasing number of possible lines that should be considered. So, an alternative iterative approach (AIA), avoiding this extensive use of the mathematical model, is now presented. It will be mainly used for evaluating the performance of MHIA. The AIA framework is shown in Figure 3.

AIA uses the same modification operators and heuristic evaluation methods as MHIA. However, the calculation time of the passenger assignment is significantly reduced by using the assignment results of the former line plan. We call this "heuristic passenger assignment." If passengers have a direct connection on their shortest path, the passenger paths are assumed to remain the same as in the former line plan. The algorithm only searches for shorter paths for those who do not have a direct connection or make a detour and then it calculates the required frequency of each line. Another difference is that the *MaxNeighbours* in AIA is limited to one, while the *MaxNeighbours* in MHIA is limited to ten. Therefore, only the most promising *Reduction* (*Extension*) is evaluated (and implemented when successful) before continuing to *Extension* (*Reduction*).

5. Computational Experiments

In this section, several experiments are executed to show the performance of the proposed MHIA on solving the profit-oriented LPP. The proposed method is implemented in C# and runs on an Intel(R) Core (TM) i7-3770 CPU 3.40 GHz and 16.0 GB computer. The MILP is implemented in CPLEX version 12.6.3, using the C# application program interface of the solver with the default parameter values. The input data include the network infrastructure, passenger demand OD matrix, and other operational parameters, such as track travel time (depending on the fixed train speed), fixed cost, and variable cost elements. The parameters used throughout the experiments are given in Table 3. It should be noted that some of these values are different (more realistic) compared to the values used in Table 1. In order to make the small example network more realistic, all link lengths were multiplied by 20. As a result, the distance between node 0 and node 2, for instance, is 200 km. Due to the lack of real-world data (which is confidential for the HSR network in China), the passenger demand OD matrices are generated randomly, and different demand scenarios are considered. All instance data are made available at https://www.mech.kuleuven.be/en/cib/lp/mainpage#section-4.

After preliminary experiments with different combinations of the maximum number of *MaxNeighbours*,

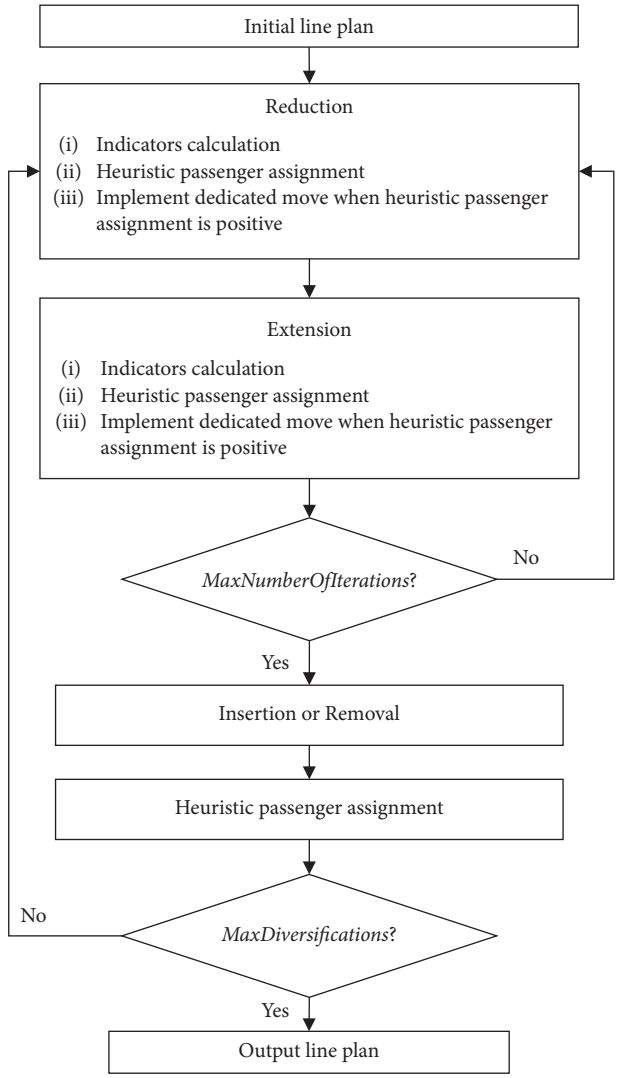


Figure 3: The alternative iterative approach framework for line planning optimization.

Table 3: Experiment parameter setting.

Name	Value	Unit
Train speed	300	km/h
Transfer time penalty	30	min
Station stopping time	3	min
Travel time value	2.5	RMB/person, min
Penalty time value	0.55	RMB/person, min
Fixed line operating cost	15,000	RMB/train
Variable line operating cost	150	RMB/train, km
Train capacity	1000	Seats/train

MaxNumberOfIterations, and *MaxDiversifications*, we set the number to 10, 20, and 30, respectively.

In the line planning problem, the transfer time is typically modelled as a penalty value [1, 7, 19–21, 29–31, 33–35]. Because the train departure times are not known during line planning, the exact transfer time is unknown and estimated by a transfer penalty. Here, we set the transfer time (penalty) to 30 minutes taking into account both the inconvenience of passengers and the practical situation in China high-speed railways.

The travel time value is used to calculate the ideal income. Here, we transform the ticket price of high-speed trains into a time-related price to link the operational revenue and the passenger travel time. However, the transfer penalty time value is used to compensate the inconvenience of making transfers.

5.1. Small Example Network Experiments. In this experiment on the small example network introduced in Section 3.3, the performance of MHIA is compared to the optimal solution obtained by using the MILP model presented in Section 3.2. In order to obtain the best possible profit-oriented line planning for this network, the MILP model starts with a pool of lines containing all 62 feasible lines that could be generated for this network. By doing this experiment, we can measure how close MHIA can get to the optimal solution for small networks.

This experiment considers three different passenger demand scenarios (PDS), based on three randomly generated OD matrices. Actually, the profit-oriented line planning problem as we propose it here has not been addressed in literature before. This makes a comparison with methods for literature [1, 7, 27, 28, 31] impossible.

The results of 10 runs of both AIA and MHIA are given in Table 4. The optimal solutions are calculated by the MILP model since all possible lines are considered. In all the tables, "PDS" refers to a passenger demand scenario. Table 4 reports the profit of the initial solution (IS) generated by AIA or MHIA, the best profit (BS) obtained by AIA and MHIA, the average profit (AveS) over all the runs, the average computation time (ACT) of one run of the approach, and the number of lines (Lines) included in the best result (BS). For MILP, BS is the optimal solution. In this experiment, the gap is the difference between the optimal solution and the best solution given by MHIA (or AIA), i.e., gap = $(BS_{MILP} - BS_{MHIA\ (or\ AIA)})/BS_{MILP}$.

From Table 4, it can be seen that the best results of MHIA are very close to the optimal solution, with an average gap of only 2.9%. Moreover, the best results of MHIA for this small example network are much better than the results given by the AIA. The computation time for all the approaches remains limited for this very small network, but, obviously, AIA is faster than MHIA and both are many times faster than MILP. We conclude here that, for a very small network, MHIA guarantees high quality solutions for the profit-oriented LPP in limited computation time. In this case, the heuristic passenger assignment used by AIA has the same initial results as the accurate passenger assignment applied in MHIA.

All detailed information about the best solutions obtained for each demand scenario (lines, frequencies, and profit calculations) and the 62 feasible lines are made available at https://www.mech.kuleuven.be/en/cib/lp/mainpage#section-4. Also the results discussed in the next sections are made available there.

Table 4: Comparison between different approaches on the small example network.

	PDS	IS	BS	AveS	ACT (s)	Lines	Gap (%)
MILP	1	—	638,670	—	26,950	4	—
	2	—	676,490	—	6680	4	—
	3	—	826,630	—	28,285	4	—
AIA	1	362,073	538,630	519,352	2	3	15.7
	2	98,773	576,724	541,066	0.9	5	14.7
	3	241,145	708,695	698,884	1	4	14.3
MHIA	1	362,073	621,244	582,370	33	5	2.7
	2	98,773	658,336	641,944	11	3	2.7
	3	241,145	799,111	759,788	16	4	3.3

5.2. Real-World HSR Network Experiments. After studying the performance of the MHIA on the small example network, we now apply it to three real-world HSR networks in China, a small network (11 nodes and 110 OD pairs), a medium network (26 nodes and 650 OD pairs), and a large network (34 nodes and 1122 OD pairs).

Given the complexity of the problem, the optimal solutions for profit-oriented line planning are not available and too time-consuming for these HSR networks, using the MILP. This makes it difficult to properly assess the performance of MHIA on these networks. Therefore, in order to test the performance of MHIA in this section, the best solutions found by MHIA will be compared with both the initial solutions found by MHIA and the best solutions found by AIA. In Section 5.3, the best solutions found by MHIA will be compared with the best results that can be obtained by solving the MILP model of Section 3.2 for these networks.

Figure 4 shows the topology of the small HSR network. We run all the experiments based on the parameters given in Table 3. The results of 10 runs of each approach are presented in Table 5. Here, the gap corresponds to gap = $(BS_{MHIA} - BS_{AIA})/BS_{AIA}$.

It should be noted that the initial solution for AIA and MHIA corresponds to the same line plan, but higher profits are calculated for MHIA due to the different passenger assignment calculations. Since AIA uses a heuristic calculation of the passenger assignment, its profits are lower than the optimal passenger assignment of MHIA on the small HSR network. Compared to AIA, MHIA can increase the operational profit by 10.7% on average while taking around ten times more computation time due to the accurate passenger assignment. The best results obtained by MHIA improve the initial solution with 40.0% on average. Taking passenger demand scenario 1 as an example, the detailed information about the best result of MHIA is shown in Figure 5 and Table 6. The numbers next to the coloured lines are the frequencies of those lines.

The high frequency on the line parts (4-9, 9-8, 8-7, and 7-6) in this solution corresponds nicely with the fact that 49% of the shortest paths include these line parts which indicate that passengers are more likely to choose these line parts. This illustrates that the best result found by MHIA is consistent with the passenger demand. The reason why line

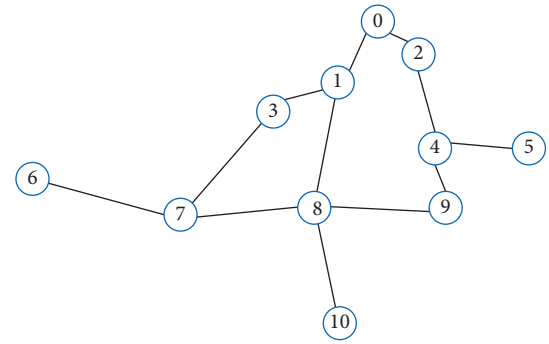


Figure 4: The topology of the small HSR network.

(4, 9, 8, and 10) has the highest frequency is that all the passengers who want to travel to destination 10 have to use that line. According to the passenger demand OD matrix, the number of passengers ending at station 10 is 2336. This indicates that 3 trains are required in order to serve all the passengers taking that train, considering the train capacity.

When the size of the considered network grows, the calculation time of looking for the optimal passenger assignment of the network will become very time-consuming. Therefore, we set the optimality gap parameter of the CPLEX solver to 5%, which means the solver will give the current best solution as a result when the gap between the current best solution and the current upper bound is lower than 5%. This significantly speeds up the passenger assignment and frequency setting. The network topology of the medium HSR network is given in Figure 6 (only the black edges), and the results of 5 runs are shown in Table 7. It should be noticed that the computation times are now expressed in minutes. The gap used here is gap = $(BS_{MHIA} - BS_{AIA})/BS_{AIA}$.

With the increasing size of the network, MHIA shows advantages in solving the profit-oriented LPP. In Table 7, the reason that the profit of the initial line plan of AIA is different from that of MHIA is again that AIA uses heuristic passenger assignment, while MHIA applies the optimal passenger assignment. On average, MHIA performs 8.3% better than AIA, and the best solution of MHIA improves the initial solution with 16.5%.

When applying MHIA to the large HSR network, corresponding to the entire 4V4H network (the black and grey edges in Figure 6), the calculation time becomes too long. Therefore, in order to make an evaluation of the calculation process, the sensitivity of algorithm parameters is tested for a single demand scenario. Since the *MaxDiversification* is one of the main parameters that determines the algorithm calculation time, different values for *MaxDiversification* are tested. In order to accelerate the algorithm calculation speed, another two values for *MaxDiversification* are considered: 10 and 20 (originally 30). The results of 5 runs for each value of *MaxDiversification* are given in Table 8. Here, gap = $(BS_{MHIA(30)} - BS_{MHIA(10\text{ or }20)})/BS_{MHIA(30)}$.

The results in Table 8 show that the solution quality remains almost the same when reducing *MaxDiversification* of MHIA. At the same time, the computation time decreases from 12.0 hours to 4.1 hours, a reduction of 63%. So, we

Table 5: Comparison between AIA and MHIA on small HSR network.

	PDS	IS ($*10^6$)	BS ($*10^6$)	AveS ($*10^6$)	ACT (s)	Lines	Gap (%)
AIA	1	2.41	3.51	3.43	12	7	—
	2	2.05	3.40	3.27	10	5	—
	3	2.43	3.23	3.18	11	5	—
MHIA	1	3.10	3.94	3.81	113	5	12.3
	2	2.08	3.74	3.70	118	4	10.0
	3	3.08	3.55	3.51	91	5	9.9

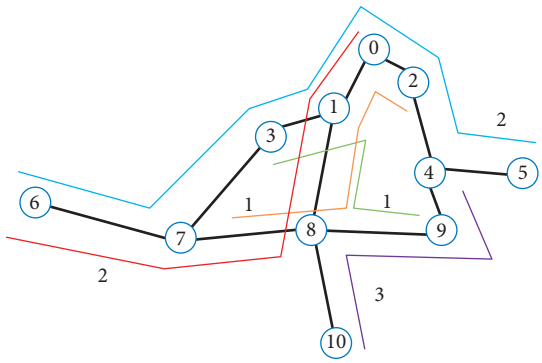


Figure 5: The best line plan of PDS 1 obtained from MHIA.

Table 6: Best line plan information of PDS 1 obtained by MHIA on the small HSR network.

Line	Distance (km)	Frequency
5, 4, 2, 0, 1, 3, 7, 6	2397	2
0, 1, 8, 7, 6	1608	2
2, 0, 1, 8, 7	1200	1
4, 9, 8, 10	1167	3
3, 1, 8, 9	891	1

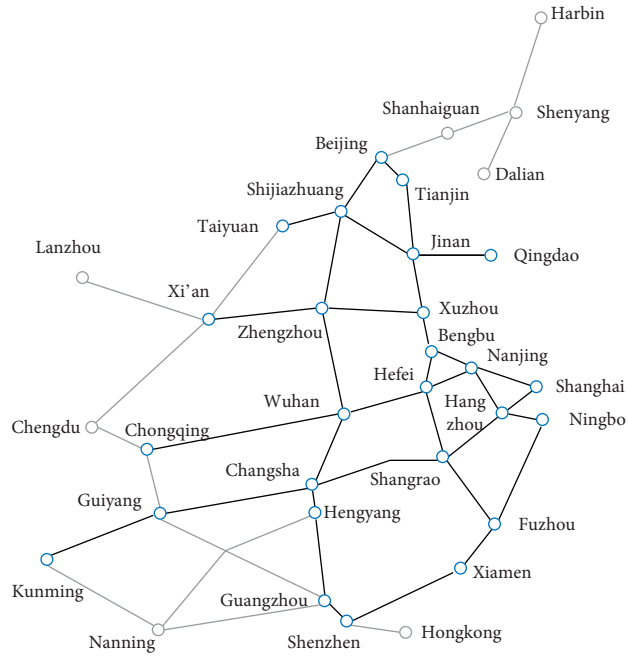


Figure 6: The topology of the medium (only black edges) and large (the black and grey edges) HSR network.

think it is appropriate to reduce the number of diversifications when solving the LPP on the large HSR network.

Finally, we run MHIA with the new parameter setting on the other two demand scenarios for the large HSR network. The parameters used in this experiment for MHIA are 5% optimality gap in CPLEX and 10 *MaxDiversification*. Also, for AIA, only 10 *MaxDiversification* are considered. The results of this setting are listed in Table 9.

For the large network case, MHIA shows an excellent ability in obtaining high quality solutions and controlling the computation time. The initial solutions of MHIA for all demand scenarios are better than the best results obtained by AIA. On average, the best results of MHIA are 13.3% better than the best results of AIA on the large HSR network. However, the computation times for AIA are clearly shorter.

These computation times mainly indicate the difficulty of the problem, not the efficiency of the algorithm. Actually, these kinds of computation times are normal for LPP of this size considering passenger assignment. Moreover, for practical applications, such computation times are acceptable, since the line planning is typically only changed in the long term.

Besides, since the problem we considered is new, the performance of the algorithm cannot be compared to other approaches from the state of the art. Therefore, we show the efficiency of the algorithm by comparing the results with optimal solutions obtained using CPLEX (only possible on smaller instances) and with AIA on all instances.

5.3. Comparison between MHIA and MILP on the HSR Network. As mentioned above, in this section, the best solutions found by MHIA will be compared with the best results that can be obtained by the MILP model of Section 3.2, for the small, medium, and large HSR network. To obtain these best results by the MILP model, all the lines obtained in the results of the different runs of MHIA (and AIA for the small HSR network) in Section 5.2 are included in the line pool for the MILP. Obviously, this does not guarantee that an optimal solution will be found, but at least a significant number of high quality lines are now considered in this line pool.

First, for the small HSR network and for each demand scenario, we use all the lines of the line plans from the result

Table 7: Comparison between AIA and MHIA on medium HSR network.

	PDS	IS ($*10^7$)	BS ($*10^7$)	AveS ($*10^7$)	ACT (min)	Lines	Gap (%)
AIA	1	1.66	2.24	2.21	13	7	—
	2	1.96	2.27	2.20	9	11	—
	3	1.76	2.14	2.11	9	10	—
MHIA	1	2.03	2.45	2.37	172	7	9.3
	2	2.18	2.40	2.37	214	7	5.7
	3	1.98	2.35	2.31	175	9	9.8

Table 8: Results for different values of *MaxDiversification* on the large HSR network.

MaxDiversification	BS ($*10^7$)	AveS ($*10^7$)	ACT (hour)	Lines	Gap (%)
10	6.79	6.64	4.1	11	−1.6
20	6.80	6.68	7.8	13	−1.8
30	6.68	6.61	12.0	11	—

Table 9: Comparison between AIA and MHIA on the large HSR network.

	PDS	IS ($*10^7$)	BS ($*10^7$)	AveS ($*10^7$)	ACT (hour)	Lines	Gap (%)
AIA	1	4.71	5.79	5.60	0.4	13	—
	2	5.16	6.11	6.02	0.3	13	—
	3	4.60	5.67	5.46	0.7	13	—
MHIA	1	6.05	6.79	6.64	4.1	11	17.3
	2	6.22	6.86	6.64	4.8	11	12.3
	3	5.71	6.25	6.09	5.2	11	10.2

of each of the 10 runs of MHIA and AIA as the line pool for the MILP model. The results are shown in Table 10.

The results show that the MHIA performs well in reaching the solutions calculated by MILP on the small HSR network for different demand scenarios. The gaps are very small. When the PDS 1 is taken as an example, there are 35 lines considered in the line pool and the detailed information about the results of both MHIA and MILP are given in Table 11. The column "Frequencies (MHIA)" presents the frequencies of the best solution found by MHIA, and the column "Frequencies (MILP)" refers to the best result the MILP model can obtain.

By comparing the frequencies in the second column and the third column, we can see the frequencies of the line in MHIA solution are typically reduced and replaced by slightly modified lines. For example, the frequency of the first line (4, 9, 8, 10) is reduced from 3 to 1, but instead, the MILP solution operates two very similar lines (4, 9, 8, 7, 6 and 4, 9, 8, 7). This could be an indication that the performance of MHIA can be improved by considering more (slightly) different lines instead of higher frequencies for fewer lines. Maybe these better solutions can also be obtained by increasing *MaxNeighbours*, *MaxNumberOfIterations,* or *MaxDiversification*. However, the quality improvement would be around 1.0%, while the computation time might significantly increase. For the current solutions, MHIA requires 113 s and MILP requires 261,445 s. We conclude that MHIA performs well in making a trade-off between the solution quality and the computation time. Moreover, having more (slightly) different lines might also make it more difficult for the passengers to understand the network. Probably, the network obtained by MHIA is even more realistic than the result obtained by MILP.

Table 10: Results of MHIA and MILP (with a given line pool) for the small HSR network.

	PDS	BS ($*10^6$)	ACT (s)	Lines	Gap (%)
MHIA	1	3.94	113	5	1.0
	2	3.74	118	4	0.8
	3	3.55	91	5	0.6
MILP	1	3.98	261,445	8	—
	2	3.77	5412	8	—
	3	3.57	6857	7	—

Now we apply the same comparison between MHIA and MILP on the medium and large HSR networks. Here, only the lines included in the line plans obtained by MHIA in different PDS are included. Otherwise, the pool of lines becomes too large to solve the MILP model. The differences between MHIA and MILP are shown in Table 12. The optimality gap of the CPLEX solver used in MILP is the same as in MHIA, i.e., 5% for both HSR networks.

The numbers between brackets in Lines column indicate the number of lines considered in the line pool of MILP, while the regular number indicates the number of lines involved in the results. In theory, the solutions obtained by MILP should be no less than the results obtained by MHIA. However, due to the 5% optimality gap of the CPLEX solver,

Table 11: Results for the given line pool for MHIA and MILP on the small HSR network.

Lines in MILP	Frequencies (MHIA)	Frequencies (MILP)
4, 9, 8, 10	3	1
5, 4, 2, 0, 1, 3, 7, 6	2	1
0, 1, 8, 7, 6	2	1
2, 0, 1, 8, 7	1	0
3, 1, 8, 9	1	0
4, 2, 0, 1, 8, 10	—	2
4, 9, 8, 7, 6	—	1
4, 9, 8, 7	—	1
5, 4, 9, 8, 7, 6	—	1
3, 1, 8	—	1
5, 4, 2, 0, 1, 3, 7	—	0
2, 0, 1, 8, 7, 6	—	0
0, 1, 8, 10	—	0
0, 1, 8, 7	—	0
0, 2, 4	—	0
1, 8, 7	—	0
5, 4, 9, 8	—	0
8, 10	—	0
3, 1	—	0
4, 2, 0, 1, 8	—	0
7, 3, 1	—	0
4, 5	—	0
5, 4, 9, 8, 10	—	0
4, 2, 0, 1, 3, 7	—	0
0, 1, 8	—	0
1, 0, 2, 4	—	0
3, 1, 0, 2, 4	—	0
1, 8, 10	—	0
1, 8, 9	—	0
1, 0, 2, 4, 5	—	0
6, 7, 8	—	0
3, 1, 8, 10	—	0
0, 1, 8, 7, 3	—	0
4, 9, 8	—	0
3, 1, 0, 2, 4, 5	—	0

Table 12: Results of the given line pool for MILP on medium and large HSR networks.

HSR networks	Approaches	PDS	BS ($*10^7$)	ACT (min)	Lines	Gap (%)
Medium	MILP	1	2.40	2.7	12 (16)	—
		2	2.50	534.7	25 (34)	—
		3	2.37	5.3	17 (25)	—
	MHIA	1	2.45	172	7	−2.0
		2	2.40	214	7	4.0
		3	2.35	175	9	0.8
Large	MILP	1	6.69	6.6	17 (17)	—
		2	6.94	53	24 (27)	—
		3	—	—	(36)	—
	MHIA	1	6.79	246	11	−1.5
		2	6.86	288	11	1.2
		3	6.25	312	11	—

it can happen that the profit of MHIA is higher than the profit of MILP, leading to a negative gap. This happens for the first demand scenario both for the medium and large networks. The calculation of PDS 3 in large network by MILP could not be obtained since the solver runs out of memory due to the size of the network and the large line pool.

For both networks, we can see in Table 12 that the optimal results given by MILP are very close to the best results obtained by MHIA for all demand scenarios. It illustrates that MHIA obtains high quality results.

6. Conclusions

In this paper, we present a matheuristic iterative approach (MHIA) for profit-oriented line planning and frequency setting, applied to a high-speed railway (HSR) network. Profit-oriented line planning considers both the operator's cost and the passenger travel time. The passenger travel time is considered by reducing the ticket price (and thus operator revenues) in case of detours or transfers. A mathematical model is discussed to define the problem in detail. MHIA integrates heuristic improvements of the line plan with an exact approach for passenger assignment and frequency setting. Two intensification and two diversification moves are considered during the algorithm.

The performance of MHIA is assessed experimentally on networks of different sizes and with different passenger demand scenarios. On the smallest network with only 7 nodes, MHIA has an average gap with the optimal solution of 2.9%. On several real-world instances based on the Chinese HSR network, MHIA improves the profit with 10.7%, 8.3%, and 13.3% on average of different size networks compared to an alternative iterative approach (AIA) which does not use the exact passenger assignment and frequency setting. Besides, MHIA increases the initial solutions of small, medium, and large networks with 40.0%, 16.5%, and 10.7%, respectively. The average computation times for the small, medium, and large network instances are 117.3 seconds, 187 minutes, and 4.7 hours, respectively.

The experiments also provide useful insights into the parameter settings when the network becomes large. With adapted parameters for larger networks, MHIA acquires high quality solutions within reasonable computation times. This confirmed by comparing the performance with the exact solution approach MILP. The best profits for all three networks are very close to the optimal solutions using MILP, when all MHIA lines are included in the line pool. For the small, medium, and large HSR networks, the gaps between the best MHIA solution and best MILP solution are 0.8%, 0.9%, and −0.2%, respectively, when taking the high-quality lines as line pools.

The large HSR network with 34 major stations where tracks split or join is considered in this paper. Comparing to all the stations of this network (more than 800 stations), the number of stations considered is still small. But all the 4V4H high-speed tracks are involved. The intermediate stations between the major stations are not considered explicitly, but stopping there could be included in the travel time between

two major stations. This provides the possibility for further study on the optimal stopping pattern and other subsequent issues.

Further work could consider different train sizes and speeds and an optimization of the stopping patterns. It would also be interesting to find out how the profit-oriented line planning can be used or modified for operator-oriented or passenger-oriented line planning. Another possibility would be to integrate AIA and MHIA in order to find an approach that can further reduce the computation time while keeping the same solution quality.

Data Availability

The data used to support the findings of this study are available online at https://www.mech.kuleuven.be/en/cib/lp/mainpage#section-4.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Key Research and Development Program of China (no. 2017YFB1200700) and the National Natural Science Foundation of China (no. U1834209). The author acknowledges the support of the China Scholarship Council (201707000074).

References

[1] H. Fu, L. Nie, L. Meng, B. R. Sperry, and Z. He, "A hierarchical line planning approach for a large-scale high speed rail network: the China case," *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 61–83, 2015.

[2] F. Liu, Z. Sun, P. Zhang, Q. Peng, and Q. Qiao, "Analyzing capacity utilization and travel patterns of Chinese high-speed trains: an exploratory data mining approach," *Journal of Advanced Transportation*, vol. 2018, Article ID 3985302, 9 pages, 2018.

[3] G. Desaulniers and M. D. Hickman, "Public transit," *Handbooks in Operations Research and Management Science*, vol. 14, pp. 69–127, North Holland, Amsterdam, Netherland, 2007.

[4] R. Borndörfer, O. Arslan, Z. Elijazyfer et al., "Line planning on path networks with application to the istanbul metrobus," in *Proceedings of the 2016 Operations Research*, Springer, Cham, Switzerland, pp. 235–241, 2018.

[5] Y. Yang, J. Li, C. Wen, P. Huang, Q. Peng, and J. Lessan, "A bi-level passenger preference-oriented line planning model for high-speed railway operations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 10, pp. 224–235, 2018.

[6] K. Steiner and S. Irnich, "Schedule-based integrated intercity bus line planning via branch-and-cut," *Transportation Science*, vol. 52, no. 4, pp. 882–897, 2018.

[7] D. Canca, A. De-Los-Santos, G. Laporte, and J. A. Mesa, "Integrated railway rapid transit network design and line planning problem with maximum profit," *Transportation Research Part E: Logistics and Transportation Review*, vol. 127, pp. 1–30, 2019.

[8] M. T. Claessens, N. M. van Dijk, and P. J. Zwaneveld, "Cost optimal allocation of rail passenger lines," *European Journal of Operational Research*, vol. 110, no. 3, pp. 474–489, 1998.

[9] J.-W. Goossens, S. Van Hoesel, and L. Kroon, "A branch-and-cut approach for solving railway line-planning problems," *Transportation Science*, vol. 38, no. 3, pp. 379–393, 2004.

[10] M. R. Bussieck, T. Lindner, and M. E. Lübbecke, "A fast algorithm for near cost optimal line plans," *Mathematical Methods of Operations Research (ZOR)*, vol. 59, no. 2, pp. 205–220, 2004.

[11] J.-W. Goossens, S. van Hoesel, and L. Kroon, "On solving multi-type railway line planning problems," *European Journal of Operational Research*, vol. 168, no. 2, pp. 403–424, 2006.

[12] M. R. Bussieck, *Optimal Lines in Public Rail Transport*, Verlag nicht ermittelbar, 1998.

[13] M. R. Bussieck, P. Kreuzer, and U. T. Zimmermann, "Optimal lines for railway systems," *European Journal of Operational Research*, vol. 96, no. 1, pp. 54–63, 1997.

[14] A. Schöbel and S. Scholl, "Line planning with minimal traveling time," in *Proceedings of the 5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS'05)*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, Mallorca, Spain, October 2006.

[15] S. Scholl, *Customer-Oriented Line Planning*, Dissertation. de, Dissertation. de, 2005.

[16] K. Nachtigall and K. Jerosch, "Simultaneous network line planning and traffic assignment," in *Proceedings of the 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2008.

[17] M. E. Pfetsch and R. Borndörfer, "Routing in line planning for public transport," in *Proceedings of the 2005 Operations Research*, Springer, Berlin, Heidelberg, Germany, pp. 405–410, 2006.

[18] R. Rosalia, "The line planning with simultaneous optimisation of operational cost and travel time," NTNU, 2017.

[19] M. Friedrich, M. Hartl, A. Schiewe, and A. Schöbel, "Integrating passengers' assignment in cost-optimal line planning," in *Proceedings of the 17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017)*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Vienna, Austria, September 2017.

[20] R. Borndörfer and M. Karbstein, "A direct connection approach to integrated line planning and passenger routing," in *Proceedings of the 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Ljubljana, Slovenia, September 2012.

[21] M. Karbstein, "Integrated line planning and passenger routing: connectivity and transfers," in *Proceedings of the 2014 Operations Research*, Springer, Cham, Switzerland, pp. 263–269, 2016.

[22] M. Schmidt and A. Schöbel, "The complexity of integrating passenger routing decisions in public transportation models," *Networks*, vol. 65, no. 3, pp. 228–243, 2015.

[23] L. A. Silman, Z. Barzily, and U. Passy, "Planning the route system for urban buses," *Computers & Operations Research*, vol. 1, no. 2, pp. 201–211, 1974.

[24] K. Kepaptsoglou and M. Karlaftis, "Transit route network design problem: review," *Journal of Transportation Engineering*, vol. 135, no. 8, pp. 491–505, 2009.

[25] G. Laporte, A. Marín, J. A. Mesa, and F. Perea, "Designing robust rapid transit networks with alternative routes," *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 54–65, 2011.
[26] Á. Marín and R. García-Ródenas, "Location of infrastructure in urban railway networks," *Computers & Operations Research*, vol. 36, no. 5, pp. 1461–1477, 2009.
[27] D. Canca, A. De-Los-Santos, G. Laporte, and J. A. Mesa, "An adaptive neighborhood search metaheuristic for the integrated railway rapid transit network design and line planning problem," *Computers & Operations Research*, vol. 78, pp. 1–14, 2017.
[28] R. Borndörfer, M. Grötschel, and M. E. Pfetsch, "A column-generation approach to line planning in public transport," *Transportation Science*, vol. 41, no. 1, pp. 123–132, 2007.
[29] L. Ahmed, C. Mumford, and A. Kheiri, "Solving urban transit route design problem using selection hyper-heuristics," *European Journal of Operational Research*, vol. 274, no. 2, pp. 545–559, 2019.
[30] V. Schmid, "Hybrid large neighborhood search for the bus rapid transit route design problem," *European Journal of Operational Research*, vol. 238, no. 2, pp. 427–437, 2014.
[31] M. Goerigk and M. Schmidt, "Line planning with user-optimal route choice," *European Journal of Operational Research*, vol. 259, no. 2, pp. 424–436, 2017.
[32] A. Schöbel, "Line planning in public transportation: models and methods," *OR Spectrum*, vol. 34, no. 3, pp. 491–510, 2012.
[33] M. E. Schmidt, *Integrating Routing Decisions in Public Transportation Problems*, Springer, New York, NY, USA, 2014.
[34] D. Liu, P. Vansteenwegen, G. Lu et al., "An iterative approach for profit-oriented railway line planning," in *Proceedings of the 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), RailNorrköping*, Linköping University Electronic Press, Linköpings universitet, Norrköping, Sweden, June 2019.
[35] D. Canca, E. Barrena, A. De-Los-Santos, and J. L. Andrade-Pineda, "Setting lines frequency and capacity in dense railway rapid transit networks with simultaneous passenger assignment," *Transportation Research Part B: Methodological*, vol. 93, pp. 251–267, 2016.
[36] Z. Gao, H. Sun, and L. L. Shan, "A continuous equilibrium network design model and algorithm for transit systems," *Transportation Research Part B: Methodological*, vol. 38, no. 3, pp. 235–250, 2004.
[37] W. Wu, R. Liu, W. Jin, and C. Ma, "Stochastic bus schedule coordination considering demand assignment and rerouting of passengers," *Transportation Research Part B: Methodological*, vol. 121, pp. 275–303, 2019.
[38] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.
[39] D. Pisinger and S. Ropke, "Large neighborhood search," *Handbook of Metaheuristics*, pp. 399–419, Springer, Boston, MA, USA, 2010.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5398298, 12 pages
https://doi.org/10.1155/2020/5398298

Research Article

Resilient Schedule Coordination for a Bus Transit Corridor

Xiongfei Lai,[1] Jing Teng ,[1] Paul Schonfeld,[2] and Lu Ling[3]

[1]*The Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai, China*
[2]*Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn L. Martin Hall, College Park, MD 20742, USA*
[3]*Lyles School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47906, USA*

Correspondence should be addressed to Jing Teng; tengjing@tongji.edu.cn

Received 7 February 2020; Revised 17 May 2020; Accepted 26 May 2020; Published 15 June 2020

Academic Editor: Yu Jiang

Providing convenient transit services at reasonable cost is important for transit agencies. Timed transfers that schedule vehicles from various routes to arrive at some transfer stations simultaneously (or nearly so) can significantly reduce wait times in transit networks, while stochastic passenger flows and complex operating environments may reduce this improvement. Although transit priority methods have been applied in some high-density cities, operating delays may cause priority failures. This paper proposes a resilient schedule coordination method for a bus transit corridor, which analyzes link travel time, passenger loading delay, and priority signal intersection delay. It maximizes resilience based on realistic passenger flow volume, whether or not transit priority is provided. The data accuracy and result validity are improved with automatically collected data from multiple bus routes in a corridor. The Yan'an Road transit corridor in Shanghai is used as a case study. The results show that the proposed method can increase the system resilience by balancing operation cost and passenger-based cost. It also provides a guideline for realistic bus schedule coordination.

1. Introduction

Providing convenient transit services for passengers at reasonable total cost is one of the main purposes of transit agencies. Bus routes operate under high pressure in China because of large passenger flow and complex road conditions, so a resilient bus system is needed.

As urban areas expand, transit networks become more complex. A rail transit network may serve as a backbone network, while bus routes connect and feed the rail transit network, serve as shuttle systems, and provide door-to-door transit services. In metropolitan areas, transfers among transit routes are needed because fixed-route buses cannot economically provide direct service among all origins and destinations. The potential value of schedule coordination is that user waiting time may be reduced at transfer stations if vehicles from different routes can arrive at transfer stations simultaneously (or nearly so). The passengers can then transfer quickly among vehicles that stop briefly near one another. A schedule coordination system can greatly reduce passenger wait times at transfer stations but cannot completely eliminate them in a system with probabilistic running times and delays.

Metropolitan areas have complex traffic conditions as well as stochastic transit demand, which may force some vehicles off their schedules and greatly increase the difficulty of achieving coordinated timed transfers. Slack time, which is a buffer parameter, should be included and optimized in schedules in order to reduce the probability of missed connections for passengers at transfer stations. Additional slack time increases operation cost but also increases transfer reliability by countering some travel time randomness [1, 2]. For each bus route, it is important to optimize the headway, which influences the vehicle operating cost and passenger waiting time. Therefore, a method is proposed here for jointly optimizing bus headways and slack times in an integrated bus system with multiple transfer stations in a transit corridor.

Schedule coordination has been well researched in different environments, while analysis and optimization of coordination performance in changing transit environments are rarely researched. Resilience reflects how a transit system is influenced by changes in its environment and how it recovers from such change. This paper proposes a method for schedule coordination considering transit system resilience. This method increases system resilience while balancing operator and user costs. This optimization is based on real-world environments, involving passenger flow fluctuations and travel times based on traffic conditions.

2. Literature Review

2.1. Schedule Coordination. Bus schedule coordination has been researched over decades. Many models have been developed for optimizing bus systems while considering transfer passengers. Among them, Rapp and Gehner [1] first developed a four-stage interactive computerized system for transfer optimization, with minimum total transfer waiting time as its objective function. Then, service quality has been considered by Andréasson [3] and Abkowitz et al. [4]. Since schedule coordination is difficult to solve analytically, a computerized tool has been developed for solving it as early as 1992 [5]. In 2001, Daganzo [6] generated a synchronized timetable for a given network by maximizing the number of simultaneous bus arrivals at the transfer nodes of the network. Since then, many extensions have been presented for solving this problem more realistically [7–10]. Relevant studies have explored whether timed transfers are appropriate for large networks with decentralized transfers [11] and how slack times can be optimized at transfer points with simple deterministic models [12] or simulation models [4]. Since passenger demand and vehicle travel times are stochastic, researchers developed probabilistic models for this problem [11]. Being limited by data sources and computational performance, research in this period mainly focused on simplifying and modeling schedule coordination in given numerical cases.

In recent years, researchers have researched this problem with more detailed and realistic methods. Ting and Schonfeld [13] proposed a schedule coordination method for a multiple hub transit network, which is more challenging than that for single hub networks. As noted in [14], the bus transit network planning problem can be presented as a sequence of four main phases, in which bus timetable preparation is the second step. Some works focus on jointly optimizing this step with vehicle scheduling [15–17]. Wu et al. presented a comprehensive review of multiobjective resynchronizing of a bus timetable, involving model and solution [18]. With real-time information, transit agencies can now predict travel time with high accuracy, thus enabling reliable and dynamic timetable optimization to be combined with bus coordination [19, 20]. Data driven modeling, as well as sufficient computational ability, makes schedule coordination more realistic and readier for application.

A rail transit network typically has a backbone role in an urban public transport system, while bus routes are usually used to connect rail transit stations with the passengers' travel origins and destinations. Multimode transit schedule coordination is well researched [21, 22]. A transport corridor is a generally linear area defined and served by one or more modes of transportation such as highways, railroads, or public transit, which share a common course. Transfers occur mainly at transfer stations where multiple routes converge. Hence, studies have been conducted to characterize different kinds of transfers and then optimize them [23–25].

Algorithms for solving this problem have been widely researched. Ibarra-Rojas and Rios-Solis have proved that the synchronized timetabling problem is NP hard. They have proposed a multistart iterated local search algorithm and a mixed integer programming method [9]. However, heuristic methods are more usually used to quickly produce timetables. Ibarra-Rojas et al. extended the heuristics from their previous work to solve a multiperiod case [26]. Other methods, including a genetic algorithm [27, 28], mixed integer programming [9], and tabu search [17], have also been used for solving this problem.

Although previous studies considered travel times when coordinating schedules, these studies usually assumed stable travel times. However, different components including the links, the stations, and even the transit priority operations may be influenced by different factors and have different impact on travel time estimation [29, 30]. For example, link travel times may be influenced by traffic, while dwell times may mainly be affected by passenger flows. With the development of advanced public transportation technologies, such as automatic vehicle location (AVL) and automatic fare collection (AFC) systems, more realistic situations with detailed travel times should be considered for schedule coordination.

2.2. Resilience. Resilience in transportation research is divided into two parts: elasticity analysis [31, 32] from the economic aspect and system resilience or vulnerability [33–35] from the physical aspect. In 1992, Wakabayashi and Iida proposed the concept of overall resilience of the system and introduced system resilience to traffic network analysis [36]. After 2000, researchers analyzed the resilience of transport system for two main cases: the recovery ability under foreseeable events (e.g., large-scale passenger flows and changes of road environment) [37, 38] or unforeseeable events (e.g., hurricanes, snowstorms, and earthquakes) [39, 40]. Compared with foreseeable system changes, unforeseeable changes are more serious and yield slower recoveries. In this field, researchers consider the change intensity and recovery ability of different traffic modes, such as rail transit compared with taxi. In addition, the study of combined resilience and vulnerability is also a hot topic. For example, the recovery cost of the system can be analyzed through the critical values [41, 42]. In transport system resilience, the main research objective is the system's restorability, which contains transportation capacity and transportation time.

3. Methodology

The proposed method analyzes a transit corridor containing a multimode bus system connecting a rail transit station. Such cases are common in real-world operations since bus routes are usually used to connect rail transit stations with the passengers' travel origins and destinations. The following assumptions are made in this study:

(1) The network of transit routes is predetermined and buses stop at every station along their routes

(2) It is assumed to be independent of transit service quality, deterministic, and uniformly distributed over time during the specified time period

(3) Preplanned transfer coordination is assumed under a no-hold policy

(4) The capacity of each bus station is unlimited

For coordinated operation, we search for optimized headways that are integer multiples of the base cycle in order to increase the possibility of immediate connections at transfer stations. To do so, schedules for all the routes have to be optimized jointly.

In the next sections total system costs with timed transfers are formulated. Then the station delay and signal priority intersection delay are checked to ensure that these optimized headways and slack times for each bus routes are reliable. Both uncoordinated and coordinated operations include nontransfer and transfer costs. Nontransfer costs include vehicle operating cost and passenger waiting cost. Transfer costs include slack-time cost, missed-connection cost, and dispatching-delay cost. Analytic solutions for uncoordinated operation are obtained. A heuristic algorithm for coordinated operation is also presented. All the results for a real-world case are verified by comparisons with the previous study of Ting and Schonfeld [13].

Symbols used in this paper are shown in Table 1. Predetermined parameters have certain baselines, which come from "TCRP 165: Transit Capacity and Quality of Service Manual" [43] and field research by the Shanghai Science and Technology Committee (unpublished reports):

3.1. Uncoordinated Operation. This paper aims to optimize bus headway in a multimode transit system. The rail transit is a part of input of bus passenger flow volume. As we can get passengers' location by both smart card data and bus's GPS data, the passenger volume can be acquired as follows:

(1) Passengers transferring from rail transit to bus can be identified from 2 consecutive smart card swiping records (rail transit to bus) in one time period (from 6:00 am to 12:00 pm)

(2) Passengers transferring from one bus line to another can be identified as 2 consecutive smart card swiping records (bus line 1 to bus line 2) in 1 time period (from 6:00 am to 12:00 pm)

(3) Other passengers can be identified as smart card swiping record

In uncoordinated operation, the objective is to minimize the total system cost by optimizing the headway independently for each route. However, many factors, such as road delay, intersection delay, and station delay, can affect bus travel time [29, 30]. With real-time information, transit agencies can now predict travel time with high accuracy [29, 30, 44]. The round-trip time, T_k, of route k is the sum of travel times on all the links of route k. The travel time can be divided into three parts: link travel time, dwell time at stations $(T_0 + T_b)$, and delay at signal priority intersections. The travel time can be expressed as

$$T_k = \sum_{i \in A(k)} E(t_i) = \sum_{r \in A(k)} E(t_r) + (T_0 + T_b) + \vartheta T_{\text{sig}}. \quad (1)$$

In reality, since some bus routes have long headways, the travel time data are insufficient for analysis. Multiple routes' data can be gathered to obtain $E(t_r)$.

Figures 1 and 2 show a transit corridor with three bus routes in Shanghai. While the planned headway of route 936 is 30 minutes, it only operates 30 times a day, so link travel time data are insufficient. However, when we combine these three routes, we collect over 300 bus round trips per day. The detailed method for combining travel times of each route in the same corridor can be found in [45].

The vehicle size used here and the linear function of vehicle operating cost proposed by Jansson [46] are as follows:

$$S_k = \frac{L_k Q_k h_k}{l_k}, \quad (2)$$

$$B_k = a_k + b_k S_k. \quad (3)$$

In (2), the demand Q_k is multiplied by a maximum load factor L_k to allow for more-than-expected passengers/bus.

The operating cost of route k is the product of the needed fleet size T_k/h_k and the unit operating cost. The total operating cost is

$$C_o = \sum_k \frac{B_k T_k}{h_k} = \sum_k \left(\frac{a_k T_k}{h_k} + \frac{L_k b_k Q_k T_k}{l_k} \right). \quad (4)$$

According to the stochastic process for passenger arrivals at stops (randomly and uniformly over time), the waiting time may be estimated as that by Welding [47] or Osuna and Newell [48]:

$$w_k = \frac{E(h_k)}{2} \left(1 + \frac{\sigma_k^2}{[E(h_k)]^2} \right). \quad (5)$$

The waiting cost is a product of waiting time w_k, unit waiting cost μ, and passenger value Q_k:

$$C_w = \sum_k \mu Q_k w_k. \quad (6)$$

When bus routes operate independently, the transfer waiting time is equal to w_k, so C_f can be formulated as

Table 1: Notation.

Symbol	Meaning	Baseline value
$A(k)$	Set of links of route k	—
a_k	Fixed vehicle operating cost of route k (RMB/min)	100
B_k	Unit vehicle operating cost (RMB/vehicle-min)	4.0
b_k	Variable vehicle operating cost of route k (RMB/seat-min)	0.2
C_o	Operating cost of route k (RMB/min)	—
C_w	Total waiting cost (RMB/min)	—
C_b	Total boarding delay cost (RMB/min)	—
C_f	Total transfer cost (RMB/min)	—
C_p	Intercycle delay cost (RMB/min)	—
C_m	Missed-connection cost (RMB/min)	—
C_d	Dispatching-delay cost (RMB/min)	—
C_s	Slack-time cost (RMB/min)	—
C_T	Total cost (RMB/min) in time period	—
C_{T0}	The lowest total cost (RMB/min) in time period	—
C_{sig}	Delay cost at signal priority intersection (RMB/min)	—
D_{mk}	Number of passengers already on board at transfer center m on route k (passengers/min)	—
$E(h_k)$	Expected value of route k headway (min)	—
$E(t_i)$	Expected travel time on link I (min)	—
$E(t_{ri})$	Expected running time on link i (min)	—
F_{mk}	Transfer demand from other routes to route k at transfer center m (passengers/min)	—
$f(t_k)$	Probability density function of arrival time on route k	—
g_{jk}	Greatest common divisor of β_j and β_k	—
h_k	Headway of route k (min)	—
L_{k}	Maximum load factor on route k	—
l_k	Load factor on route k	—
N_k	Frequency of missing signal priority (buses/hour)	—
Q_k	Demand of route k (passengers/min)	—
q_{jk}	Transfer demand from route j to k (passenger/min)	—
R_k	Total transfer demand to route k (passenger/min)	—
S_k	Vehicle size on route k	—
T_k	Round-trip time (min)	—
s_{mk}	Slack time at station m on route k (min)	—
T_0	Scheduled passenger loading time per station (min)	—
T_b	Extra passenger loading time per station (min)	—
T_i	Vehicle delay at signal priority intersection (min)	—
T_m	Passenger loading time (min)	—
T_n	Passenger alighting time (min)	—
T	Traffic signal time cycle (min)	—
T_x	Link travel time (min)	—
T_{sig}	Delay time at signal priority intersection (min)	—
T_G	Green light time in a traffic signal cycle (min)	—
T_P	Duration of time period (min)	—
v	In-vehicle value of time (RMB/passenger-min)	—
y	Base cycle (min); value is 1 min, the same as the minimum value of headway in a rail station	—
z_{jk}	The transfer waiting times between connecting routes j and k (RMB)	—
ϑ	Signal priority indicator (=1 if signal priority provided and 0 otherwise)	—
σ_k^2	Variance of headway on route k (min^2)	—
μ_1	Unit passenger waiting cost (RMB/passenger-min; baseline value is 1.5 RMB/passenger-min)	1.5
μ_2	Unit passenger waiting cost (RMB/passenger-min; baseline value is 5.0 RMB/passenger-min) from boarding delay	5.0
μ_3	Unit passenger waiting cost (RMB/passenger-min) from missing signal priority intersections	10,000
α	Unit passenger loading time (min/passenger)	0.1
β	Unit passenger alighting time (min/passenger)	0.05
β_j	Integer multiples of the base cycle	—
δ_{mk}	$\delta_{mk} = 1$ if transfer station m is on route k and 0 otherwise	—

Figure 1: Layout of the transit corridor in Shanghai.

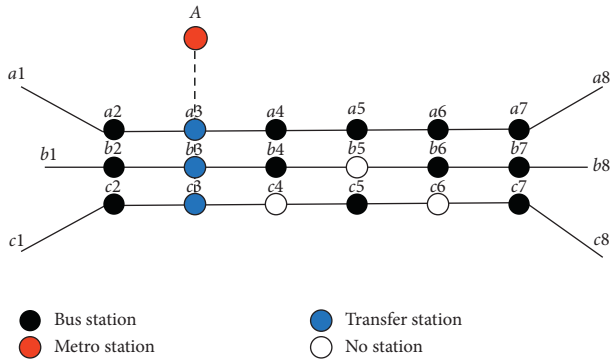


Figure 2: Initial bus routes layout.

$$C_f = \sum_j \sum_{\substack{k \\ k \neq j}} \mu q_{jk} w_k = \sum_j \sum_{\substack{k \\ k \neq j}} \mu R_k \left[\frac{E(h_k)}{2} \left(1 + \frac{\sigma_k^2}{[E(h_k)]^2} \right) \right]. \tag{7}$$

When the passenger flows per bus increase, bus dwell times at stops also increase. The stopping delay can be expressed as follows:

$$\begin{aligned} T_m &= \begin{cases} \alpha m_k - T_0, & \alpha m_k \geq T_0, \\ 0, & \text{otherwise}, \end{cases} \\ T_n &= \begin{cases} \beta n_k - T_0, & \beta n_k \geq T_0, \\ 0, & \text{otherwise}, \end{cases} \\ C_b &= \sum_k \mu_2 Q_k T_b = \mu_2 (T_m + T_n). \end{aligned} \tag{8}$$

Although the bus wait time at intersections is included in $E(t_r)$, there should be no delay at intersections with signal priority. Priorities including bus lane along with signal priority have been applied on particular bus routes to improve their operation efficiency and reliability. However, since priorities are only given in a certain period during a signal circle, when delays exceed this certain period, buses may miss the signal priority. If buses operate on exclusive lanes, these delays are mainly caused by boarding delay at stations. Thus, signal priority missing cost is introduced as follows:

$$C_{\text{sig}} = \sum_k \mu_3 Q_k T_{\text{sig}} = \begin{cases} \sum_k \mu_3 Q_k \dfrac{T - T_G}{2}, & T_b \geq T_G \\ 0, & \text{otherwise}. \end{cases} \tag{9}$$

During the optimization process, μ_3 is set extremely high to avoid missing signal priority.

To sum up, the total cost is

$$C_T = C_o + C_w + C_f. \tag{10}$$

In this paper, the goal of the optimization process is to maximize resilience in the transfer system. The resilience here is defined as the ability to resist and recover from passenger flow fluctuation. In current typical studies, when analyzing system resilience, we first define (10) as the system service efficiency function. Based on the system efficiency function, the resilience function can be defined by the system efficiency rate (C_T/C_{T0}). The system resilience function can be expressed as

$$R = \frac{\int (C_T/C_{T0}) \mathrm{d}t}{T_P}. \tag{11}$$

The optimization process is to get the maximum R.

3.2. Coordinated Operation. Slack-time cost includes passenger cost and vehicle cost. Specifically, it can be divided into three parts: The first term is the slack-time delay cost to passengers already on board. The second term is the waiting cost for passengers transferring to route k. The last term is the increased vehicle operating cost due to the slack time at each transfer station on route k. The slack-time cost can be expressed as

$$C_s = \sum_k \sum_{m \in N(c)} \left(D_{mk} v + F_{mk} \mu + \frac{B_k}{h_k} \right) s_{mk} \delta_{mk}. \tag{12}$$

The transfer waiting time should consider the difference in headways and slack times between two routes. In real cases, the passenger flows of different routes, and hence the optimized headways of those routes, may differ greatly. Therefore, this paper uses the Integer-Ratio Headways, introduced in [13]. In this case, the vehicle operating cost, passenger waiting cost, and in-vehicle cost are the same as for the uncoordinated operation. In [13] passenger waiting times are analyzed based on the condition that the headways as well as scheduled travel times between transfer stations are the same or integer multiples of the same cycle time. The waiting time is

$$z_{jk} = \frac{g_{jk} y}{h_k} \sum_{i=1}^{\left((h_k/g_{jk})-1\right)} \left(i g_{jk} y + s_{mk} \right) = g_{jk} y \left(\frac{h_k}{2 g_{jk} y} - \frac{1}{2} \right) + s_{mk}. \tag{13}$$

Thus, the transfer waiting times between connecting routes j and k are as follows:

$$C_p = \sum_j \sum_{\substack{k \\ k \neq j}} \mu_1 q_{jk} z_{jk}. \tag{14}$$

Assuming reasonably that passengers may miss at most one bus, the missed-connection cost and delay-cost formulations are simplified. Since buses from both routes meet at the transfer station once every $h_j \times h_k / (g_{jk} \times y)$ minutes, the average transfer demand is $q_{jk} \times g_{jk} \times y / h_k$ for each connecting bus. The missed-connection cost is

$$\begin{aligned} C_m = \sum_{m \in N(c)} \sum_j \sum_{\substack{k \\ k \neq j}} \mu q_{jk} \frac{g_{jk} y}{h_k} \delta_{mk} \delta_{mj} \times \Bigg[& \int_{S_{mk}}^{h_k} \int_{-h_k}^{S_{mk}} \left(h_k - S_{mk} - t_j + S_{mj} \right) f(t_k) \mathrm{d}t_k f\left(t_j\right) \mathrm{d}t_j \\ & + \int_{S_{mj}}^{h_j} \int_{S_{mk}}^{t_j - S_{mj} + S_{mk}} \left(h_k - S_{mk} - t_j + S_{mj} \right) f(t_k) \mathrm{d}t_k f\left(t_j\right) \mathrm{d}t_j \Bigg]. \end{aligned} \tag{15}$$

The dispatching-delay cost is

$$\begin{aligned} C_d = \sum_{m \in N(c)} \sum_j \sum_{\substack{k \\ k \neq j}} \mu q_{jk} \frac{g_{jk} y}{h_k} \delta_{mk} \delta_{mj} \times \Bigg[& \int_{S_{mk}}^{h_k} \int_{-h_k}^{S_{mk}} \left(t_k - S_{mk} \right) f(t_k) \mathrm{d}t_k f\left(t_j\right) \mathrm{d}t_j \\ & + \int_{S_{mj}}^{h_j} \int_{t_j - S_{mj} + S_{mk}}^{h_k} \left(t_k - S_{mk} - t_j + S_{mj} \right) f(t_k) \mathrm{d}t_k f\left(t_j\right) \mathrm{d}t_j \Bigg]. \end{aligned} \tag{16}$$

The optimization procedure is specified as follows:

We first optimize the headway for each bus route separately.

We assume $T_b = 0$ and $T_{\text{sig}} = 0$. Since these two parameters are 0, the total cost C_T in our model will be the same as in [13].

The optimal headway of route k can be obtained by setting the first derivative of the total cost with respect to headway h_k to be zero and solving for h_k. The first derivative is

$$\frac{\partial C_T}{\partial h_k} = -\frac{a_k T_k}{h_k^2} + \mu Q_k \left(\frac{1}{2} - \frac{\sigma_k^2}{2h_k^2}\right) + \mu R_k \left(\frac{1}{2} - \frac{\sigma_k^2}{2h_k^2}\right). \quad (17)$$

Thus, the optimal h_k is

$$h_k^* = \sqrt{\frac{2a_k T_k + \mu\sigma_k^2 (Q_k + R_k)}{\mu (Q_k + R_k)}}. \quad (18)$$

The second derivative of total cost with respect to h_k is

$$\frac{\partial^2 C_T}{\partial h_k^2} = \frac{2a_k T_k}{h_k^3} + \frac{\mu R_k \sigma_k^2}{h_k^3} + \frac{\mu Q_k \sigma_k^2}{h_k^3}. \quad (19)$$

Since each component in (19) is always positive, the second derivative is nonnegative, and thus h_k^* globally minimizes the total system cost; then, the resilience is maximized.

The maximal headway (policy headway) permitted between two adjacent buses in route k, $h_k^{\max} = S_k l_k / Q_k$, must be checked for each route.

As in [13], the base cycle is optimized as a round fraction of 60 min (i.e., 60 min divided by an integer). For optimizing coordinated operations, the headways of different bus routes at the same transfer station should be optimized for multiples of the base cycle y. Based on that, slack time is set to minimize the expected total costs of bus operations and passenger waiting times, including those due to missed connections, which are affected by slack times.

Compared with the heuristic algorithm provided in [13], the optimizing goal and computational basis are changed here from system cost (C_T) to system resilience (R). Besides, more parameters are checked. We check h_k^* to find if $T_b > 0$ or $T_{\text{sig}} > 0$. Since T_i precludes signal priority, T_{sig} should be 0. We rank T_b in ascending order, reduce h_k^* to reduce T_b, update T_k, and calculate h_k^*. This procedure is repeated until $N_k = 0$ and no significant further improvement is obtained.

4. Case Study

4.1. Background of This Case. This paper focuses on a bus corridor which connects a rail transit station and includes multiple bus routes. It contains a rail transit station (Jiangsu Road Rail Transit Station, which is among the busiest in Shanghai) and a bus corridor (Yan'an Road transit corridor, which is the earliest established and largest transit corridor in Shanghai). We formulate a system resilience function and find the combinations of headways and slack times which maximize that function. There are two bus network configurations in this transit corridor, as shown in Figures 2 and 3. The pre-2017 network, in which there were three bus routes operating concurrently, is shown in Figure 2. After 2017, route 71 serves as the main bus route in this corridor. It has been provided with exclusive bus lane and transit signal priority. First, analytical solutions for optimal headways in the uncoordinated transit network are found. Then, a heuristic algorithm is developed to jointly find the headways and slack times that maximize the system resilience for the coordinated transit network.

Yan'an Road bus lane starts at Zhongshan Road and ends at Gubei Road, with a total length of 10 km. It opened in about 2000 as Shanghai's first bus lane. The section considered here extends from the West Yan'an Road–Kaixuan Road (KX) bus stop to the Mid Yan'an Road–Shimen Road No. 1 (SM) bus stop. Its characteristics are illustrated in Figures 1–3.

The data acquisition process uses an automated collection system and related equipment based on AVL, to gather GPS and smart card data. The bus routes have been changed in early 2017. Thus, our case study uses two data series: February to July in 2015 and February to July in 2018. The raw data includes the following data types:

Smart card data: card id, fare collection machine id, time point

Vehicle-route data: vehicle ID, route ID, route name

Bus stop data: stop ID, stop name, stop longitude, stop latitude

Route-stop data: route ID, route name, operating direction, stop number

GPS data: route ID, vehicle ID, time point, longitude, latitude, speed, direction angle

The baseline parameter values, adopted for the present work from "TCRP 165: Transit Capacity and Quality of Service Manual" [43] and field research by the Shanghai Science and Technology Committee (unpublished reports), are $B_k = 4.0$ RMB/bus-min, $\mu_1 = 1.5$ RMB/passenger-min, $\mu_2 = 5.0$ RMB/passenger-min, $\mu_3 = 10{,}000$ RMB/passenger-min, and $v = 1.0$ RMB/passenger-min. In real networks, the travel times on each link between two hubs may differ considerably by period and travel direction. In this case, the travel direction is from suburban to urban area, and the time period is 6:00 am to 12:00 pm.

In both conditions, we calculate the total cost by five methods: The first one is planned operation, which means we use the headways provided by the transit agency. In the second method, called uncoordinated operation, the headways are optimized independently for the different routes. In the coordinated operation, the headways and slack time of different routes are optimized jointly, using the method in [13], based on average passenger volume (method 3) and maximum passenger volume (method 4). The method proposed in this paper is presented as the fifth method, called the resilience optimized operation. In this method, the headways and slack times of different routes are optimized jointly, considering boarding delay and missed signal priority delay.

4.2. General Comparison

4.2.1. Initial (Pre-2017) Bus Routes. This corridor included three bus routes: 71, 127, and 936, as shown in Figure 1. The

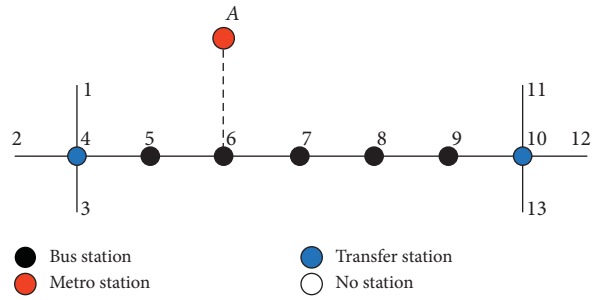


Figure 3: Modified bus routes layout.

Table 2: Comparison of different operation methods at average passenger flows.

Variables	Planned operation	Uncoordinated operation	Coordinated operation (avg. passenger volume)	Coordinated operation (max passenger volume)	Resilience optimized operation
h_a	3	3.051	4	3	4
h_b	10	9.887	12	9	8
h_c	30	28.169	28	27	28
S_{3a}	—	—	1.00	1.00	1.00
S_{3b}	—	—	2.00	1.00	2.00
S_{3c}	—	—	2.00	2.00	2.00
C_o	137.096	144.316	146.316	151.373	148.387
T_b	17.212	10.143	6.562	0	0
N_{71}	0	0	0	0	0
N_{127}	0	0	0	0	0
N_{936}	0	0	0	0	0
C_w	166.537	135.103	131.004	128.143	129.097
C_N	303.633	279.419	277.320	279.516	277.320
C_p	—	—	5.136	5.193	5.198
C_s	—	—	7.184	7.091	7.077
C_m	—	—	11.213	11.392	11.432
C_d	—	—	3.123	3.431	3.308
C_f	30.151	28.193	26.656	27.107	26.656
C_T	333.784	307.612	303.976	306.623	304.499
R	0.916	0.930	0.948	0.942	0.964

configuration of these bus routes is shown in Figure 2. Station ids $a2$ to $a7$ mean stations on route 71, station id b means stations on route 127, and station id c means stations on route 936. Note that when $i = 1$ to 7, a_i, b_i, and c_i are the same station on different routes. However, $a1$, $b1$, $c1$, $a8$, $b8$, and $c8$ are different stations. The blank points $c4$, $b5$, and $c6$ mean that there are no stations on these routes, and thus buses pass through without stopping.

The comparison of different operation methods is shown in Table 2.

Table 2 shows that all four optimizing methods increase the system resilience and reduce the total cost compared to the planned operation (method 1). The method proposed in this paper can increase the system resilience more than the other methods and does not increase the total cost (304.499 RMB compared with 306.623 RMB and 303.976 RMB). It performed well in real cases with real passenger volumes. Coordinated operation is preferable to uncoordinated operation. Its operation cost is higher than that of uncoordinated operation, but the waiting cost and transfer cost decrease more, so the total cost decreases, as shown in Table 2.

4.2.2. After Changing Bus Routes. Bus route 71 is the main route in this corridor. It operates on an exclusive lane and has signal priority. At stations 4 and 10, there are two feeder bus routes: route 1250 (marked as route a) at station 4 and route 974 (marked as route b). The configuration of these three bus routes is shown in Figure 3.

The comparative results for different operation methods after bus route restructuring are shown in Table 3.

Table 3 shows that all the four optimizing methods increase the system resilience and reduce the total cost compared to the planned operation. Coordinated operation is preferable to uncoordinated operation. Its operation cost is higher than that of uncoordinated operation, but the waiting cost and transfer cost decrease more, so the total cost decreases. The table also shows that bus route 71 misses signal priority 5 times per hour when it operates as planned. All the optimized operations reduce the missing times, and

Table 3: Comparison of different operation methods after restructuring routes.

Variables	Planned operation	Uncoordinated operation	Coordinated operation (avg. passenger volume)	Coordinated operation (max passenger volume)	Resilience optimized operation
h_a	4	3.801	4	3	4
h_b	11	11.150	12	9	8
h_c	15	14.731	16	12	16
S_{4a}	—	—	0.50	1.00	0.50
S_{4b}	—	—	1.00	0.50	1.00
S_{10a}	—	—	1.00	0.50	0.50
S_{10c}	—	—	1.00	1.00	1.00
C_o	157.116	162.210	170.310	171.310	171.145
T_b	16.211	11.312	7.143	0	0
N_{71}	5	3	2	0	0
N_{127}	0	0	0	0	0
N_{936}	0	0	0	0	0
C_w	173.342	158.332	149.433	148.707	149.067
C_N	330.458	320.542	319.743	320.017	320.212
C_p	—	—	5.973	6.854	6.104
C_s	—	—	6.985	7.013	6.847
C_m	—	—	12.211	12.610	12.146
C_d	—	—	4.016	4.123	4.073
C_f	36.131	31.631	29.185	30.600	29.170
C_T	366.589	352.173	348.928	350.617	349.382
R	0.874	0.921	0.944	0.939	0.957

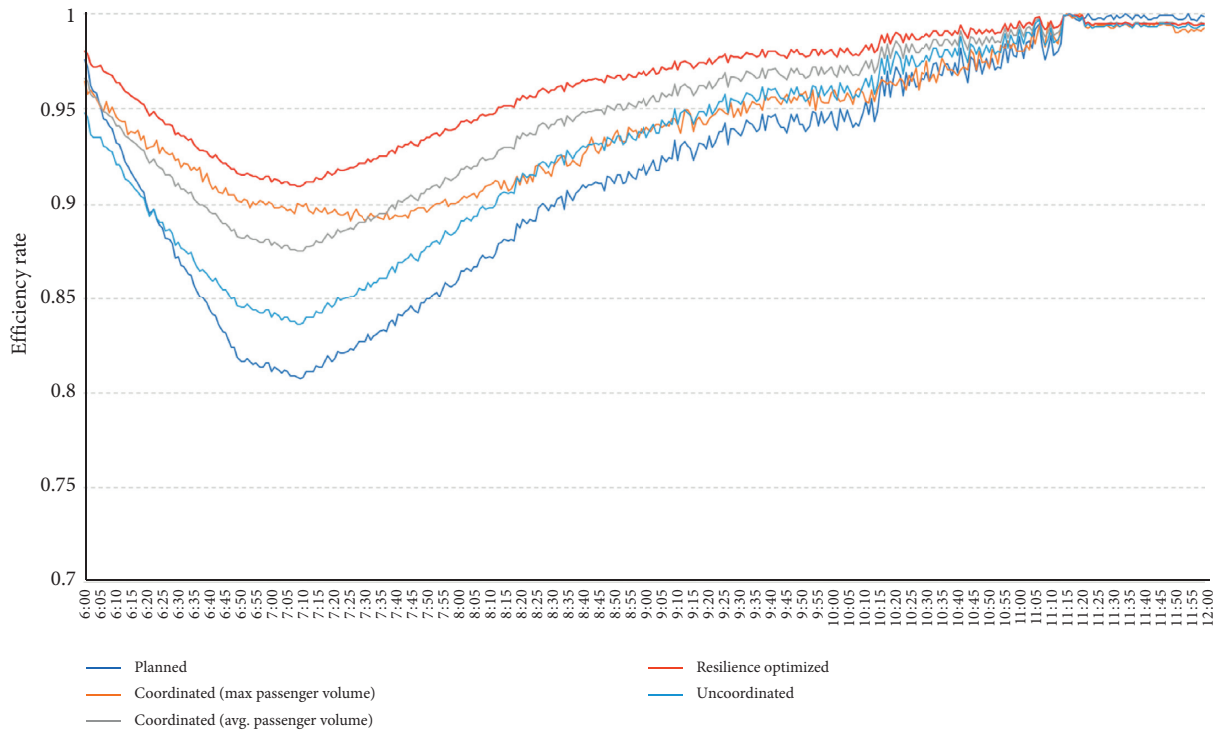


Figure 4: Detailed comparison of these optimization methods.

coordinated operations perform better: With the proposed method, the boarding delay and the hourly number of missed signal priorities become 0; the operation cost is lower than that of method four. All the optimized operations reduce the total cost; among them the coordinated operations perform better compared with the uncoordinated ones.

4.3. Detailed Comparison. To understand how these results arise, we take the results of initial bus routes to compare these five methods in detail within the whole time period. The comparison is based on the results in February to July 2018. The minute-based evaluation graph is shown below.

From Figure 4, we can understand the following:

Table 4: Levels of system efficiency rate.

Level	*A*	*B*	*C*	*D*	*E*
Efficiency rate	1–0.95	0.95–0.90	0.90–0.85	0.85–0.80	<0.80

Level A is the ideal level.

Table 5: Comparison of optimization methods based on levels of system efficiency rate.

Variables	Planned operation	Uncoordinated operation	Coordinated operation (avg. passenger volume)	Coordinated operation (max passenger volume)	Resilience optimized operation
The lowest efficiency rate	0.808 (0.074)	0.836 (0.067)	0.875 (0.076)	0.892 (0.071)	0.909 (0.056)
Time for recovering to level *A* (min)	251 (17)	207 (16)	163 (16)	198 (16)	111 (14)
Level *A* (min)	109 (11)	153 (13)	197 (13)	162 (12)	249 (10)
Level *B* (min)	84 (7)	95 (6)	99 (8)	133 (7)	111 (7)
Level *C* (min)	94 (7)	75 (7)	64 (5)	65 (5)	0 (0)
Level *D* (min)	73 (6)	37 (4)	0 (0)	0 (0)	0 (0)
Level *E* (min)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Note. The variance is shown in the brackets.

(1) The efficiency rate (C_T/C_{T0}) decreases when morning peak period starts and increases when it ends. All these 5 methods can make the system recover from passenger volume increasing, since the efficiency rates of all these 5 methods increase to 1 finally.

(2) The planned method performs the worst. When passenger volume increases, the efficiency rate decreases and reaches a very low level.

(3) The uncoordinated method performs better than the planned method but worse than the coordinated methods.

(4) When we use the methods introduced by Ting and Schonfeld [13], the results differ based on different passenger volume. When we use average passenger volume, the efficiency decreases greatly since the headway is too long. At the maximum passenger volume, the operation cost is high even when passenger volume is not very high (before 6:30 am and after 9:30 am), so the efficiency rate recovers slowly.

(5) The proposed method performs best since it uses reasonable headways with realistic passenger volume.

According to "Transit Capacity and Quality of Service Manual" (3rd edition) [43], the service quality can be divided into five levels. In this case, we can use this standard to divide the efficiency rate into five levels, which are shown in Table 4.

Then, we compare these 5 methods based on the results in Figure 4 and the results are shown in Table 5.

From Table 5, we can find that the proposed method not only resists the efficiency rate decrease but also has the shortest time for recovering to level *A*, so it has the highest resilience. Besides, among these methods, coordinated operation performs better than uncoordinated operation. Planned operation performs the worst. Coordinated operation avoids level *D*, and the proposed method avoids level *C*.

5. Conclusions and Recommendations

This paper proposed an extension of the schedule coordination method by Ting and Schonfeld [13] for jointly optimizing headways and slack times in coordinated schedules for public transportation routes within the aspect of resilience. This method is applied in a realistic case, with passenger flow fluctuations as well as travel time variability based on operating conditions. The main contributions are as follows.

Firstly, this paper applies and extends a schedule coordination model within a real multihub environment (stations 2–7 in Figure 2 and stations 4, 6, and 10 in Figure 3), considering passenger flow fluctuations and whether transit priority is provided, within the new resilience angle. This advance improves this method's practicality.

Secondly, the case study presented here shows that the coordinated schedule obtained with the proposed method can ensure high system resilience based on the verification with realistic passenger volumes, which means it performs well in resisting and recovering from passenger volume fluctuations and traffic variations. It balances the operation cost and passenger cost. The proposed method improves the resilience of the schedule.

Thirdly, this paper focuses on a transit corridor, combining data from several bus routes for analysis. It improves data volume and accuracy. This paper allows transit agencies to optimize coordinated schedules considering passenger flow fluctuations and transit priority.

Resilience is important in analyzing and optimizing transit schedules. The value of the resilient schedule coordination method proposed in this paper is demonstrated for balancing operator and user costs in real world, considering

passenger flow fluctuations and traffic variations. Further research may not only consider more parameters in different environments, but also explore more efficient algorithms for solving this problem. Besides, this method can be further extended from corridors to urban transit networks, intercity trains, buses, and airlines.

Data Availability

The GPS and smart card data used to support the findings of this study are restricted by the Shanghai Science and Technology Committee. Requests for data will be considered by the corresponding author for researchers who meet the criteria for access to confidential data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Study conception and design were conducted by Xiongfei Lai, Jing Teng, and Paul Schonfeld. Data were collected by Xiongfei Lai and Jing Teng. Analysis and interpretation of results were performed by Xiongfei Lai. Manuscript draft was prepared by Xiongfei Lai Jing Teng, Paul Schonfeld, and Lu Ling. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

This work was supported by "Public Transit System Simulation Analysis Platform and Demonstrations" (project no. 17DZ1204409), in the field of social development of "Science and Technology Innovation Action Plan," Shanghai Science and Technology Committee. It was also supported by the program of China Scholarships Council (no. 201806260146), College of Transportation Engineering, Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety.

References

[1] M. H. Rapp and C. D. Gehner, "Transfer optimization in an interactive graphic system for transit planning," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 619, pp. 27–33, 1976.

[2] K. K. T. Lee and P. Schonfeld, "Optimal slack time for timed transfers at a transit terminal," *Journal of Advanced Transportation*, vol. 25, no. 3, pp. 281–308, 1991.

[3] I. Andréasson, "Volvo approach to computer-aided transportation planning: Transportation research record," *Journal of the Transportation Research Board*, vol. 657, pp. 9–14, 1977.

[4] M. Abkowitz, R. Josef, J. Tozzi, and M. K. Driscoll, "Operational feasibility of timed transfer in transit systems," *Journal of Transportation Engineering*, vol. 113, no. 2, pp. 168–177, 1987.

[5] A. Désilets and J.-M. Rousseau, "SYNCRO: a computer-assisted tool for the synchronization of transfers in public transit networks," in *Computer-Aided Transit Scheduling*, pp. 153–166, Springer, Berlin, Germany, 1992.

[6] C. Daganzo, "On the coordination of inbound and outbound schedules at transportation terminals," in *Proceedings of the International Symposium on Transportation and Traffic Theory*, Yokohama, Japan, July 1990.

[7] A. Ceder, B. Golany, and O. Tal, "Creating bus timetables with maximal synchronization," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 10, pp. 913–928, 2001.

[8] A. Ceder and O. Tal, "Designing synchronization into bus timetables. Transportation research Record," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1760, no. 1, pp. 28–33, 2001.

[9] O. J. Ibarra-Rojas and Y. A. Rios-Solis, "Synchronization of bus timetabling," *Transportation Research Part B: Methodological*, vol. 46, no. 5, pp. 599–614, 2012.

[10] L. Zhigang, S. Jinsheng, W. Haixing, and W. Yang, "Regional bus timetabling model with synchronization," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 2, pp. 109–112, 2007.

[11] J. H. Bookbinder and A. Désilets, "Transfer optimization in a transit network," *Transportation Science*, vol. 26, no. 2, pp. 106–118, 1992.

[12] A. Ceder, *Public Transit Planning and Operation: Theory, Modeling and Practice*, Elsevier, Butterworth-Heinemann, Amsterdam, Netherlands, 2007.

[13] C.-J. Ting and P. Schonfeld, "Schedule coordination in a multiple hub transit network," *Journal of Urban Planning and Development*, vol. 131, no. 2, pp. 112–124, 2005.

[14] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, and J. C. Muñoz, "Planning, operation, and control of bus transport systems: a literature review," *Transportation Research Part B: Methodological*, vol. 77, pp. 38–75, 2015.

[15] M. Wei and B. Sun, "Bi-level programming model for multimodal regional bus timetable and vehicle dispatch with stochastic travel time," *Cluster Computing*, vol. 20, no. 1, pp. 401–411, 2017.

[16] B. Varga, T. Tettamanti, and B. Kulcsár, "Optimally combined headway and timetable reliable public transport system," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 1–26, 2018.

[17] T. Liu and A. Ceder, "Integrated public transport timetable synchronization and vehicle scheduling with demand assignment: a bi-objective bi-level model using deficit function approach," *Transportation Research Procedia*, vol. 23, pp. 341–361, 2017.

[18] Y. Wu, H. Yang, J. Tang, and Y. Yu, "Multi-objective re-synchronizing of bus timetable: model, complexity and solution," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 149–168, 2016.

[19] S. J. Berrebi, K. E. Watkins, and J. A. Laval, "A real-time bus dispatching policy to minimize passenger wait on a high frequency route," *Transportation Research Part B: Methodological*, vol. 81, pp. 377–389, 2015.

[20] Y. Xuan, J. Argote, and C. F. Daganzo, "Dynamic bus holding strategies for schedule reliability: optimal linear control and performance analysis," *Transportation Research Part B: Methodological*, vol. 45, no. 10, pp. 1831–1845, 2011.

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

[22] L. Kang, J. Wu, H. Sun, X. Zhu, and Z. Gao, "A case study on the coordination of last trains for the Beijing subway

network," *Transportation Research Part B: Methodological*, vol. 72, pp. 112–127, 2015.

[23] I. Dakic, L. Ambühl, O. Schümperlin, and M. Menendez, "On the modeling of passenger mobility for stochastic bi-modal urban corridors," *Transportation Research Part C: Emerging Technologies*, vol. 113, pp. 146–163, 2020.

[24] F. Castrillon and J. Laval, "Impact of buses on the macroscopic fundamental diagram of homogeneous arterial corridors," *Transportmetrica B: Transport Dynamics*, vol. 6, no. 4, pp. 286–301, 2018.

[25] K. Arnet, S. I. Guler, and M. Menendez, "Effects of multi-modal operations on urban roadways. Transportation research Record," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2533, no. 1, pp. 1–7, 2015.

[26] O. J. Ibarra-Rojas, F. López-Irarragorri, and Y. A. Rios-Solis, "Multiperiod Bus Timetabling," *Transportation Science*, vol. 50, no. 3, pp. 805–822, 2015.

[27] M. Kim and P. Schonfeld, "Integrating bus services with mixed fleets," *Transportation Research Part B: Methodological*, vol. 55, pp. 227–244, 2013.

[28] M. Kim and P. Schonfeld, "Integration of conventional and flexible bus services with timed transfers," *Transportation Research Part B: Methodological*, vol. 68, pp. 76–97, 2014.

[29] S. I.-J. Chien, Y. Ding, and C. Wei, "Dynamic bus arrival time prediction with artificial neural networks," *Journal of Transportation Engineering*, vol. 128, no. 5, pp. 429–438, 2002.

[30] R. Jeong and R. Rilett, "Bus arrival time prediction using artificial neural network model," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*, pp. 988–993, IEEE, Washington, WA, USA, October 2004.

[31] Homeland Security Advisory Council, "Report of the critical infrastructure task force," 2006, http://www.dhs.gov/xlibrary/assets/HSAC_CITF_Report_v2.pdf.

[32] A. Rose and S.-Y. Liao, "Modeling regional economic resilience to disasters: a computable general equilibrium analysis of water service disruptions," *Journal of Regional Science*, vol. 45, no. 1, pp. 75–112, 2005.

[33] E. D. Vugrin, D. E. Warren, M. A. Ehlen et al., "A framework for assessing the resilience of infrastructure and economic systems," in *Sustainable and Resilient Critical Infrastructure Systems*, pp. 77–116, Springer, Berlin, Germany, 2010.

[34] R. Dorbritz, *Assessing the Resilience of Transportation Systems in Case of Large-Scale Disastrous Events*, Vilnius Gediminas Technical University Press Technika, Vilnius, Lithuania, 2011.

[35] R. Li, Q. Dong, C. Jin, and R. Kang, "A new resilience measure for supply chain networks," *Sustainability*, vol. 9, no. 1, p. 144, 2017.

[36] H. Wakabayashi and Y. Iida, "Upper and lower bounds of the terminal reliability of road networks: an efficient method with Boolean algebra," *Journal of Natural Disaster Science*, vol. 14, no. 1, pp. 29–44, 1992.

[37] S. Havlin, "Robustness of a network of networks," *Physical Review Letters*, vol. 107, no. 19, Article ID 195701, 2010.

[38] J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," *Transportation Research Part E: Logistics and Transportation Review*, vol. 37, no. 2-3, pp. 191–229, 2001.

[39] Y. Asakura, "Reliability measures of an origin and destination pair in a deteriorated road network with variable flows," in *Proceedings of the Fourth Meeting of the EURO Working Group in Transportation*, Tyne, UK, September 1996.

[40] Y. Zhu, K. Ozbay, K. Xie, and H. Yang, "Using big data to study resilience of taxi and subway trips for Hurricanes Sandy and Irene," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2599, no. 1, pp. 70–80, 2016.

[41] O. Cats and E. Jenelius, "Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information," *Networks & Spatial Economics*, vol. 14, no. 3-4, pp. 435–463, 2014.

[42] D. Watling and N. C. Balijepalli, "A method to assess demand growth vulnerability of travel times on road network links," *ransportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 772–789, 2012.

[43] Transportation Research Board, *Transit Capacity and Quality of Service Manual*, Transportation Research Board, Washington, DC, USA, 3rd edition, 2013.

[44] W. Fan and Z. Gurmu, "Dynamic travel time prediction models for buses using only GPS data," *International Journal of Transportation Science and Technology*, vol. 4, no. 4, pp. 353–366, 2015.

[45] J. Teng and X. Lai, "An integrated method for urban transit evaluation and optimization," *Advances in Mechanical Engineering*, vol. 9, no. 7, 2017.

[46] J. O. Jansson, "A simple bus line model for optimisation of service frequency and bus size," *Journal of Transport Economics and Policy*, vol. 14, no. 1, pp. 53–80, 1980.

[47] P. I. Welding, "The instability of a close-interval service," *Journal of the Operational Research Society*, vol. 8, no. 3, pp. 133–142, 1957.

[48] E. E. Osuna and G. F. Newell, "Control strategies for an idealized public transportation system," *Transportation Science*, vol. 6, no. 1, pp. 52–72, 1972.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5609524, 20 pages
https://doi.org/10.1155/2020/5609524

Research Article

A Fast Approach for Reoptimization of Railway Train Platforming in Case of Train Delays

Yongxiang Zhang,[1,2,3] Qingwei Zhong,[1,2,3] Yong Yin,[1,2,3] Xu Yan,[1,2,3] and Qiyuan Peng[1,2,3]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*
[2]*National United Engineering Laboratory of Integrated and Intelligent Transportation, Southwest Jiaotong University, Chengdu 610031, China*
[3]*National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 610031, China*

Correspondence should be addressed to Yong Yin; yinyong@swjtu.edu.cn

Received 18 January 2020; Revised 24 February 2020; Accepted 20 May 2020; Published 3 June 2020

Academic Editor: Yu Jiang

Train platforming is critical for ensuring the safety and efficiency of train operations within the stations, especially when unexpected train delays occur. This paper studies the problem of reoptimization of train platforming in case of train delays, where the train station is modeled using the discretization of the platform track time-space resources. To solve the reoptimization problem, we propose a mixed integer linear programming (MILP) model, which minimizes the weighted sum of total train delays and the platform track assignment costs, subject to constraints defined by operational requirements. Moreover, we design an efficient heuristic algorithm to solve the MILP model such that it can speed up the reoptimization process with good solution precision. Furthermore, a real-world case is taken as an example to show the efficiency and effectiveness of the proposed model and algorithm. The computational results show that the MILP model established in this paper can describe the reoptimization of train platforming accurately, and it can be solved quickly by the proposed heuristic algorithm. In addition, the model and algorithm developed in this paper can provide an effective computer-aided decision-making tool for the train dispatchers in case of train delays.

1. Introduction

Railroad transportation plays an important role in providing economic and environment-friendly transport services for passengers and goods. In particular, transport demand of railroad transportation is increasing rapidly in some parts of regional markets. For instance, according to the statistical data of the National Bureau of Statistics of China, the volume of passenger and freight traffic of China railway in 2018 has increased by 9.44% and 9.15%, compared with the volume in 2017 [1]. Thus, more and more trains are required to be operated on the railway network with limited infrastructure capacity to satisfy the large demand. Therefore, the railway operators need to adopt advanced and reliable computer-aided decision systems to improve train operation efficiency. Moreover, the railway planning process includes the network design and line planning at the strategic level, the train timetabling, train platforming, rolling stock scheduling and crew scheduling at the tactical level, and the real-time traffic management at the operational level [2]. As a result, the core of the computer-aided decision systems is to develop effective optimization techniques for the different railway planning and operation stages. We refer the readers to Assad [3], Cordeau [4], and Lusby et al. [2] for excellent surveys on the related optimization methods. In general, train timetabling stage can determine the arrival and departure times of the trains at each of the visiting stations from an aggregate perspective [5–8]. Train operations at the stations, including the train routing on the arrival and departure routes and the train platform assignment decisions, are usually optimized

by solving the train platforming problem [8]. Apart from the boarding and alighting of passengers as well as the loading and unloading of goods at the station platforms, many other complicated operation tasks can be performed at the stations, such as the train turn-around movement and the shunting work. Thus, it is of critical importance to develop efficient and effective optimization techniques to improve the train operation efficiency at the stations without causing any conflicts.

Due to the complex layout of large railway stations and the associated dense train traffic, it is usually a challenging task for the train dispatchers to make a high-quality train operation plan within a station in a short period. Kroon et al. [9] have already proven that the train platforming problem is an NP-complete combinational optimization problem. As a result, many researchers have developed sophisticated and efficient models and algorithms to deal with the train platforming problem, such as the node packing model and the branch-and-cut algorithm in Zwaneveld et al. [10, 11] and the set packing model and the branch-and-price algorithm in Lusby et al. [12, 13]. In practice, however, trains may suffer from all kinds of disturbances and disruptions, such as bad weather, equipment failure, and management factors [14]. When train delays occur, the scheduled train timetable needs to be rescheduled in real time, and thus the train operations within the stations need to be reoptimized quickly to prevent any potential conflicts [15–18]. However, with regard to the train platforming problem, two managerial aspects lack in-depth study. First, most of the researchers study the train platforming problem without considering train delays, and thus they do not consider the situation where it could be impossible to make a feasible train operation plan at the stations with the given train timetable. Second, in order to generate a disposition timetable during the train rescheduling process as soon as possible, the detailed train operations at the stations are neglected, and the aggregate station capacity is adopted instead [17, 18].

In this study, we aim to bridge the research gaps lying in the train platforming and train rescheduling problems. More specifically, we reoptimize the train platforming problem in case of train delays and generate a new train operation plan within the station in real time. Our solution is to develop a mixed integer linear programming (MILP) model, where the train station is modeled using the discretized platform track time-space resources, and to propose an efficient heuristic algorithm. The goal of the proposed model and algorithm is to simultaneously minimize the deviation from the train timetable and the total train operating costs, realizing the coherence between train operations and the station management. The theoretical and practical contributions of this study mainly include the following four aspects:

(1) The train arrival and departure times and the train platform assignment are optimized simultaneously so that the negative influence of train delays can be minimized.

(2) The novel modeling method based on the discretized platform track time-space resources can describe the train conflicts accurately, where the complex binary train sequencing variables in the big-M modeling framework can be avoided.

(3) An efficient heuristic algorithm is designed to quickly obtain the near-optimal solutions for the real-time reoptimization of train platforming.

(4) A set of real-life experiments is performed to show the efficiency and effectiveness of the proposed model and algorithm.

The rest of this paper is structured as follows. Section 2 provides a comprehensive literature review on the optimization and reoptimization of train platforming problems. The method on how to describe the discretized platform track time-space resources is introduced in Section 3. In Section 4, we propose a new MILP model based on the discretized platform track time-space resources, as well as some valid equalities to strengthen the resulting MILP model. Section 5 presents the detailed algorithmic steps of a heuristic algorithm, i.e., the genetic and simulated annealing hybrid algorithm, for solving the reoptimization problem in a real-time manner. The efficiency and effectiveness of the proposed methods are verified through a set of real-life experiments in Section 6. Section 7 gives the concluding remarks and possible further research directions.

2. Literature Review

Train platforming problem is a key optimization stage in the railway hierarchical planning process, and it has attracted the attention of many researchers around the world [2]. Given the information of train arrival and departure times and the train running directions obtained at the train timetabling stage, the train dispatchers need to decide how to schedule the trains at the stations to avoid the train delays caused by the cross interference between train operations. The complexity of the train platforming problem grows quickly as the number of trains increases and the station layout becomes more and more complicated. Therefore, a lot of excellent works have been done regarding the train platforming problem, such as Carey [19], Zwaneveld et al. [10], Lusby et al. [12], Chakroborty and Vikram [20], Caprara et al. [21], and Kang et al. [22]. We next review these works mainly from two aspects, i.e., the train platforming problem at the tactical level and the train platforming problem considering the negative impact of train delays.

The studies on the train platforming problem at the tactical level were initially focused on scheduling the train movements at the stations efficiently and economically with the fixed or slightly relaxed train arrival and departure times. Zwaneveld et al. [10] developed the first node packing model for the train platforming problem, and the station routes were classified as the inbound route, outbound route, and complete. In addition, small deviations from the planned train arrival and departure times were considered to increase the possibility of finding a feasible solution, and an improved branch-and-cut algorithm was designed to find the optimal solutions with the goal of maximizing the number of trains routed through the station. Kroon et al. [9] proved that the node packing model in Zwaneveld et al. [10] was

NP-complete once each train had more than three routing possibilities. Based on the work in Zwaneveld et al. [10], Zwaneveld et al. [11] further proposed a weighted node packing model for the train platforming problem, where several hierarchical objectives can be considered with the unified weighted sum form. However, the node packing model has its own deficiencies to be put into practice. For instance, all of the conflicting train movements need to be identified in advance in the node packing model, and it usually results in a large-scale optimization model. As a result, Lusby et al. [12] built a set packing model of the train platforming problem, and a branch-and-price algorithm was designed such that the feasible train routes can be dynamically generated. Besides, the train platforming problem can be also reformulated as a graph colouring problem, where the number of colours is determined by the available tracks in the stations [23]. An integer programming model that was equivalent to graph colouring problem in Cardillo and Mione [23] was developed by Billionnet [24]. Caprara et al. [21] also allowed a slight relaxation on the ideal arrival and departure times of the trains at the platforms, and they presented a 0-1 quadratic programming model to solve the train platforming problem. In particular, the quadratic term in the objective function was designed to consider the soft incompatibility costs between train paths. Sels et al. [25] developed a new MILP model to schedule the trains in the current and future train set, and penalty costs were considered in the objective when a train was cancelled or assigned to a different platform.

Apart from the exact solution methods, some researchers also developed heuristic algorithms for the train platforming problem [26–29]. Carey and Carville [26] introduced a sequential construction heuristic algorithm for the train platforming problem at a single station, which was designed according to the train dispatchers' manual methods. In addition, trains can be delayed in the planning process to generate a feasible solution, and the assignment of trains to the undesired platforms was penalized. Carey and Crawford [27] further extended the work in Carey and Carville [26] by considering the train scheduling and platforming problems at multiple stations, where those stations were connected by several one-way lines. Similarly, trains can be also delayed when they were scheduled from one station to the corresponding next station. Kang et al. [28] proposed a simulated annealing heuristic algorithm with the goal of achieving the balanced occupation of station turnouts while minimizing the total occupation times of the station turnouts. Wu et al. [29] developed a mean-variance optimization model for the train platforming problem based on Markowitz's portfolio theory, which was aimed to minimize the total occupation time costs of groups of turnouts. Moreover, an efficient simulated annealing heuristic algorithm was designed to solve the resulting mixed integer nonlinear programming (MINLP) model.

Some researchers also tried to integrate the two stages of train timetabling and train platforming problems to further increase the railway network capacity utilization. Carey [19] proposed an integrated optimization model to simultaneously optimize the lines, platforms, and routes of the trains in a railway network, and an improved sequential searching method was adopted to solve the resulting model. In addition, the arrival and departure headways between two trains were enforced to guarantee the train running safety. Dorfman and Medanic [30] proposed a discrete event model to schedule the trains in the railway network, where the train routes were assumed to be fixed for simplicity. D'ariano et al. [31] formulated the train scheduling problem through the alternative graph modeling approach, and an efficient branch-and-bound algorithm was designed to solve the MILP model. Rodriguez [32] proposed a combined constraint programming and simulation method for scheduling trains in the junctions, where the constraint programming model representing the signalling system behaviour and the simulation model can reflect the train and driver behaviour. More recently, D'Ariano et al. [33] and Zhang et al. [8] studied the integrated train timetabling and maintenance task scheduling problems at the microscopic level, and efficient heuristic algorithms based on the commercial solvers were developed accordingly.

Some works were also performed on the optimization of a delay resistant train platforming plan and the real-time reoptimization of train platforming problem in case of unexpected train delays. Herrman [34] studied the stable train routing problem, where the goal was to keep the train schedule to be conflict-free as much as possible in case of perturbations. In addition, a local search heuristic approach was designed to solve the train schedule stabilization problem. Bešinović and Goverde [35] proposed a max-plus automata model for the station capacity assessment and a delay propagation model for the robustness evaluation of the train platforming plan. Based on these two models, the objectives of minimization of capacity occupation and maximization of robustness were then achieved with a heuristic approach. Kang et al. [22] introduced the concept of stochastic schedule with uncertain train arrival and departure times, and a simulated annealing algorithm was developed to solve the corresponding train platforming problem. D'Ariano et al. [36] designed a rolling horizon approach for the real-time train rescheduling problem, and the alternative graph formulation was adopted for solving the smaller subproblem in each rolling period. Chakroborty and Vikram [20] developed a MILP model for optimally allocating trains to the platform tracks, where the accurate train arrival times can only be available shortly before the train arrives at the station such that trains could be reassigned to different platforms. Besides, the headway between two trains was also considered while delaying the train arrival and departure times. Lusby et al. [13] proposed a set packing model for rescheduling the trains in the junctions in case of disruptions. In particular, a branch-and-price algorithm was introduced to solve the set packing model, and the objective was to minimize the weighted deviations from the trains' scheduled arrival times. Liu et al. [37] developed a MINLP model to reallocate the tracks of trains at the stations in a real-time manner, where the goal was to minimize average use time of groups of turnouts without changing the planned train times. Moreover, a genetic simulated annealing algorithm was also designed to solve the MINLP model efficiently.

Table 1 summarizes the modeling approaches, objectives, solution methods, and other characteristics of relevant publications on the optimization and reoptimization of train platforming problems, and major highlights can be drawn from Table 1. First, most of the studies focus on the train platforming problem at the tactical level, and the scheduled train arrival and departure times at the stations are usually assumed to be fixed. Therefore, the optimization objectives of those studies are generally minimizing the total costs or maximizing the total benefits. Second, exact algorithms account for the majority of the solution methods to solve the optimization or reoptimization of the train platforming problems. Therefore, efficient and flexible heuristic algorithms are required, especially for the real-time reoptimization of the train platforming problem. Third, among the studies that consider the negative influence of train delays, only one research by Chakroborty and Vikram [20] simultaneously incorporates the factors of train delay, reassignment of platform, and deviations from scheduled train arrival and departure times. However, they adopt a commercial solver to deal with the MILP model. In this study, we develop a new MILP model for the reoptimization of the train platforming problem based on the discretized time-space resources, and we also design an efficient heuristic algorithm to guarantee the achievement of real-time reoptimization.

3. Analysis of Platform Track Time-Space Resources

Let T be the planning horizon; we handle the time resources as small time units of length $\Delta\tau$. The number of time units is equal to $|T| = [T/\Delta\tau]$ in the entire planning horizon. In addition, the number of platform tracks in a station is denoted by $|I|$, i.e., the maximum spatial capacity. Hence, the platform track time-space resources of a station can be represented by a two-dimensional matrix R:

$$R = \left[r_{i,t}\right] = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,|T|} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,|T|} \\ \cdots & \cdots & \cdots & \cdots \\ r_{|I|,1} & r_{|I|,2} & \cdots & r_{|I|,|T|} \end{bmatrix}, \quad (1)$$

where i and t are the indexes of the platform track and the time unit, respectively, and a binary variable $r_{i,t}$ in matrix R denotes the occupation and vacancy state of the platform track time-space resource (i,t), where

$$r_{i,t} = \begin{cases} 1, & \text{platform track time} - \text{space resource } (i,t) \text{ is used,} \\ 0, & \text{othewise.} \end{cases} \quad (2)$$

Figure 1 provides an illustrative example of the modeling method of platform track time-space resources. Suppose that 5 trains successively arrive at or depart from station N which has 6 platform tracks within the planning horizon of 60 min. The detailed schedules of train operations in both directions are given in Figures 1(b) and 1(c), and the time unit $\Delta\tau$ is set to 5 min. Platform track time-space resources and the corresponding matrix R of a feasible usage plan are described in Figures 2(a) and 2(b), respectively. The application requirements of the time-space resources modeling method for the reoptimization of train platforming problem can be formulated as follows:

(1) *Inseparability*. A train must occupy only one platform track and cannot occupy more than one platform track simultaneously.

(2) *Exclusivity*. One platform track can only store one train at any time unit to prevent the train conflicts.

(3) *Continuity*. A train operation on one platform track with the duration equal to Δt time units cannot be interrupted. If one train starts to use track i at time t, it continues to occupy the platform track i until $t + \Delta t$, i.e., $r_{i,t} = r_{i,t+1} = r_{i,t+2} = \ldots = r_{i,t+\Delta t} = 1$.

4. Mathematical Modeling for Reoptimization of Train Platforming

4.1. Modeling Assumptions. Without loss of generality, we make the following key modeling assumptions to facilitate the modeling process.

(1) When it is impossible to eliminate the conflicts through reassignment of platforms, the train arrival and departure times can only be delayed to resolve the conflicts.

(2) One minute is applied to discretize the platform track time-space resources, and smaller time units can be adopted if necessary.

(3) Trains are modeled as single objects moving in the railway station, and thus the influence of train lengths on the clearing times of the platform tracks is not considered.

(4) By enforcing the arrival and departure headways between two trains running in the same direction as well the safety time interval for platform track operation, it can be guaranteed that there are no conflicts between any two trains.

(5) All trains will pass through the railway station, and turn-around movements of the trains are not considered for simplicity.

4.2. Definitions of Sets, Parameters, and Variables. Sets, parameters, and variables of this study are defined in Table 2. In particular, we assume that all parameter values related to time are multiples of the time unit $\Delta\tau$.

4.3. Objective Function. The objective function in equation (3) contains the weighted sum of two parts of costs. The first part is the sum of train arrival and departure delays, considering the train priority P_l and the weighting factor α. The second part is the total platform track assignment costs. In particular, the more that train l is preferred for platform track i, the smaller is the corresponding platform track assignment cost $c_{l,i}$. Note that value of $c_{l,i}$ can be determined

Table 1: Summary of relevant publications on the optimization and reoptimization of the train platforming problems.

Publication	Modeling approach	Objective	Solution method	Train delay	Reassignment of platform	Deviations from scheduled train times	Largest instance solved	Computation times (in seconds) of the largest instance
Zwaneveld et al. [10]	BILP	Max the number of routed trains	Branch-and-cut			√	A railway station with 17 platforms and 27 trains	85
Cardillo and Mione [23]	GC	Feasibility	Iterative heuristic algorithm				A railway station with 16 tracks and 242 trains	0.67
Billionnet [24]	BILP	Feasibility	Optimization software				A railway station with 14 tracks and 200 trains	1.64
Carey and Carville [26]	MILP	Min the costs of deviations from preferred times, platforms, and lines	Iterative sequential searching method		√	√	A railway station with 29 subproblems and 491 trains	—
Rodriguez [32]	CP	Min train delays	Tree search	√		√	A 15.18 km long railway junction with 24 trains	54.76
Lusby et al. [12]	BILP	Max total benefits	Branch-and-price				A railway junction with 64 block sections and 45 trains	26.4
Caprara et al. [21]	BQP	Min total costs	Branch-and-cut-and-price				A railway station with 14 platforms and 237 trains	350
Sels et al. [25]	MILP	Min total penalties	Commercial solvers		√		A railway station with 14 platforms and 160 trains	3.0
Zhang et al. [8]	MILP	Min total train journey times and maintenance tardiness costs	Iterative heuristic algorithm				A railway network with 27 stations, 55 segments, and 58 trains	7200
Liu et al. [37]	MINLP	Min average use time of groups of turnouts	Genetic simulated annealing algorithm		√		A railway station with 11 tracks and 30 trains	1800
Kang et al. [22]	MILP	Min total occupation times and balanced track occupation	Simulated annealing	√			A railway station with 11 tracks and 28 trains	726
Chakroborty and Vikram [20]	MILP	Min total costs	Commercial solver	√	√	√	A railway station with 9 platforms and 110 trains	600
Lusby et al. [13]	BILP	Min total deviations	Branch-and-price	√		√	A railway junction with 524 track sections and 66 trains	151.55

Table 1: Continued.

Publication	Modeling approach	Objective	Solution method	Train delay	Reassignment of platform	Deviations from scheduled train times	Largest instance solved	Computation times (in seconds) of the largest instance
This paper	MILP	Min total costs	Genetic and simulated annealing hybrid algorithm	√	√	√	A railway station with 11 platforms and 70 trains	32.92

Note. BILP represents 0-1 integer linear programming; GC represents graph colouring; MILP represents mixed integer linear programming; CP represents constraint programming; BQP represents 0-1 quadratic programming; "—" represents no specific information is available.

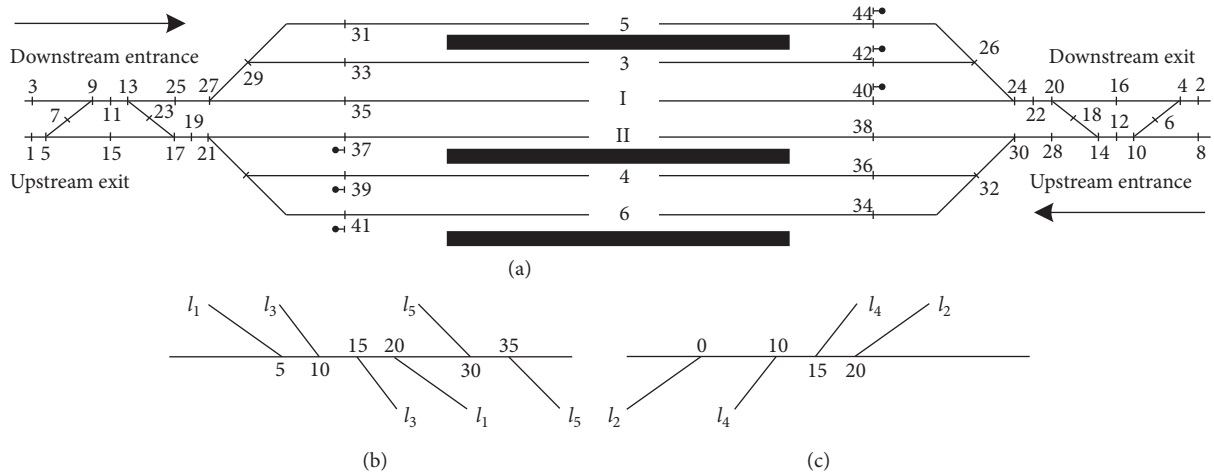


Figure 1: Layout and train timetables of station N: (a) layout of station N; (b) train diagram of station N in the downstream direction; (c) train diagram of station N in the upstream direction.

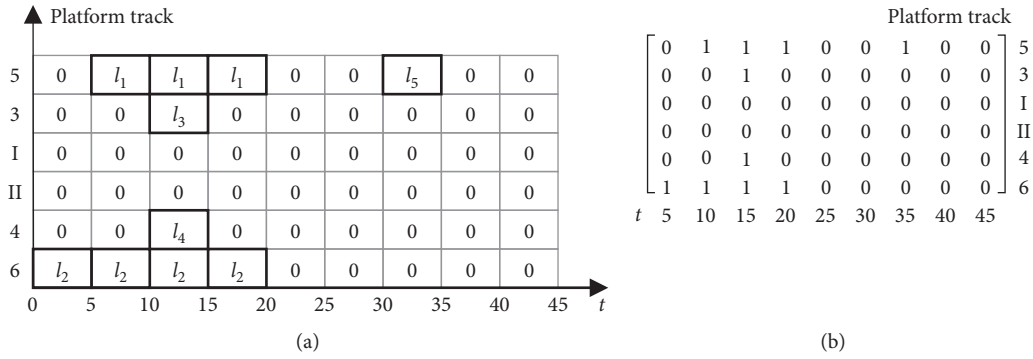


Figure 2: Platform track time-space resource usage plan and time-space resource matrix: (a) usage of platform track time-space resources; (b) platform track time-space resource matrix.

according to several influence factors, including the initial platform track assignment plan, the convenience of passenger boarding and alighting, and other special requirements for train operations. Therefore, the smaller the value of $c_{l,i}$ is, the more the train l prefers the platform track i.

$$\min z = \alpha \sum_{l \in L} P_l \left[\left(y_{l,a} - t_{l,a} \right) + \left(y_{l,d} - t_{l,d} - D \right) \right] + \sum_{l \in L} \sum_{i \in I_l} w_{l,i} c_{l,i}. \tag{3}$$

4.4. Constraints

4.4.1. Relation between Platform Track State Variables. According to definitions of platform track state variables $x_{l,i,t}$, $u_{l,i,t}$, and $v_{l,i,t}$, constraint (4) expresses the relationship among those three platform track state variables. In particular, when both the variables $u_{l,i,t}$ and $v_{l,i,t}$ are equal to 0, it can be known that train l is occupying the platform track i at time t, i.e., $x_{l,i,t} = 1$. Figure 3 shows an illustrative example on how to calculate the value of $x_{l,i,t}$ according to the values of the variables $u_{l,i,t}$ and $v_{l,i,t}$. More specifically, train l_1 arrives at the platform track 5 in Figure 1 at time 5 and it leaves from the corresponding platform track at time 20. Therefore, Figures 3(a) and 3(b) present the values of the variables $u_{l,i,t}$ and $v_{l,i,t}$ for train l_1 according to its actual arrival and departure times. In addition, according to constraint (4), the values of $v_{l,i,t}$ in Figure 3(c) can be easily obtained.

$$x_{l,i,t} = 1 - \left(u_{l,i,t} + v_{l,i,t}\right), \quad \forall l \in L, i \in I, t \in \{1, 2, \ldots, \mathrm{MT}\}. \tag{4}$$

4.4.2. Actual Train Arrival and Departure Times. Constraints (5) and (6) show that values of actual arrival time $y_{l,a}$ and actual departure time $y_{l,d}$ of train l can be inferred from the values of platform track utilization state variable $u_{l,i,t}$ and the platform track clearance state variable $v_{l,i,t}$.

$$y_{l,a} = \mathrm{MT} - \sum_{i \in I} \sum_{t=1}^{\mathrm{MT}} \left(1 - u_{l,i,t}\right), \quad \forall l \in L, \tag{5}$$

$$y_{l,d} = \mathrm{MT} - \sum_{i \in I} \sum_{t=1}^{\mathrm{MT}} v_{l,i,t}, \quad \forall l \in L. \tag{6}$$

4.4.3. Platform Track Assignment Constraint. Constraint (7) requires that each train l can only be assigned to one platform track, and the corresponding platform track has to be in the set I_l. Furthermore, constraint (8) enforces that a train l cannot be assigned to forbidden platform tracks, i.e., tracks that are not in the set I_l. Constraints (7)-(8) provide a flexible method to restrict the platform track choices of the trains.

$$\sum_{i \in I_l} w_{l,i} = 1, \quad \forall l \in L, \tag{7}$$

$$w_{l,i} = 0, \quad \forall l \in L, i \in I \backslash I_l. \tag{8}$$

4.4.4. Platform Track Capacity Constraint. Constraint (9) ensures that any platform track i can only be occupied by at most one train at any time t, such that no conflicts can occur between any two trains that are assigned to the same platform track.

$$\sum_{l \in L} x_{l,i,t} \leq 1, \quad \forall i \in I, t \in \{1, 2, \ldots, \mathrm{MT}\}. \tag{9}$$

4.4.5. Continuous Train Operation on the Platform Track. Constraints (10)–(12) guarantee that train operations on the platform tracks should be performed continuously, by enforcing the relationship between the values of variables $u_{l,i,t}$ and $v_{l,i,t}$. For instance, if $w_{l,i} = 1$ in constraint (10), which implies that train l is assigned to the platform track i, constraint (10) changes to $u_{l,i,t} \geq u_{l,i,t+1}$. According to the definition of $u_{l,i,t}$ in Section 4.2, the condition $u_{l,i,t} \geq u_{l,i,t+1}$ is obviously satisfied. On the other hand, if $w_{l,i} = 0$ in constraint (10), constraint (10) is no longer valid. Similarly, constraints (11) and (12) define the relationships between $v_{l,i,t}$ and $v_{l,i,t+1}$ as well as between $u_{l,i,t}$ and $u_{l,i,t+1}$.

$$u_{l,i,t} \geq u_{l,i,t+1} + w_{l,i} - 1, \quad \forall l \in L, i \in I, t \in \{1, 2, \ldots, \mathrm{MT} - 1\}, \tag{10}$$

$$v_{l,i,t} \leq v_{l,i,t+1} - w_{l,i} + 1, \quad \forall l \in L, i \in I, t \in \{1, 2, \ldots, \mathrm{MT} - 1\}, \tag{11}$$

$$u_{l,i,t} \leq u_{l,i,t+1} + w_{l,i}, \quad \forall l \in L, i \in I, t \in \{1, 2, \ldots, \mathrm{MT} - 1\}. \tag{12}$$

4.4.6. Safety Headway between Two Trains. Constraints (13)–(18) impose the required safety headway between two arrival or departure trains running in the same direction, and the train sequencing variables $\lambda_{l,k}$ and $\mu_{l,k}$ as well as the platform track choice variables $w_{l,i}$ and $e_{l,k}$ are also embedded in those constraints. Constraint (13) enforces the safety headway between two arrival trains in the same direction, and it is valid only if $\lambda_{l,k} = 1$, i.e., train l arrives at the station before train k. When $\lambda_{l,k} = 1$ in constraint (13), it can be shown that the safety headway between trains l and k depends on the value of the variable $e_{l,k}$. More specifically, if $e_{l,k} = 1$, trains l and k are assigned to the same platform track, and they are separated by the safety time interval D. Otherwise, trains l and k are assigned to two different platform tracks, and they are separated by the arrival headway h_a. Meanwhile, constraint (14) guarantees the safety headway between two departure trains in the same direction. Constraint (15) ensures that if both the variables $w_{l,i}$ and $w_{k,i}$ are equal to 1, the value of the variable $e_{l,k}$ can be enforced to be 1. Furthermore, constraint (16) describes that the values of the two variables $e_{l,k}$ and $e_{k,l}$ for two different trains l and k are equal. Besides, constraints (17) and (18) guarantee that only one of the train sequencing variables $\lambda_{l,k}$ and $\lambda_{k,l}$ or $\mu_{l,k}$ and $\mu_{k,l}$ between any two different trains is equal to 1.

$$y_{l,a} - y_{k,a} \geq \left(1 - e_{l,k}\right) h_a + e_{l,k} D - \lambda_{l,k} M, \quad \forall l, k \in L: l \neq k, \pi_l = \pi_k, \tag{13}$$

$$y_{l,d} - y_{k,d} \geq \left(1 - e_{l,k}\right) h_d + e_{l,k} D - \mu_{l,k} M, \quad \forall l, k \in L: l \neq k, \pi_l = \pi_k, \tag{14}$$

Table 2: Sets, parameters, and variables.

Symbol	Definition
Sets	
L	Set of trains, indexed by l
L_1	Set of delayed trains, $L_1 \subseteq L$
I	Set of platform tracks, indexed by i
I_l	Set of platform tracks that train l can utilize, $I_l \subseteq I$
T	Set of time intervals in the planning horizon, indexed by t
Parameters	
$c_{l,i}$	Cost of assigning train l to platform track i
π_l	0-1 parameter is equal to 1 if train l is running in the downstream direction; 0, otherwise
$q_{l,i}$	0-1 parameter is equal to 1 if train l was initially assigned to platform track i before a delay occurs; 0, otherwise
S	The time when the information of train delays is updated
$t_{l,a}$	Scheduled arrival time of train l
$t_{l,d}$	Scheduled departure time of train l
$t^1_{l,a}$	Estimated arrival time of train l when a delay occurs
$t^1_{l,d}$	Estimated departure time of train l when a delay occurs
Δ_l	Dwell time of train l
P_l	Priority of train l
D	Safety time interval for platform track operation
MT	Sum of the length of the planning horizon T and the safety time interval for platform track operation D
$\Delta_{\max}$	Maximum value of dwell times of all trains
h_a	Headway between two arrival trains running in the same direction
h_d	Headway between two departure trains running in the same direction
α	Objective function weighting factor
M	A sufficiently large number
Variables	
$w_{l,i}$	0-1 platform track assignment variable = 1 if train l is assigned to platform track i; = 0 otherwise
$e_{l,k}$	0-1 same platform track choice variable = 1 if train l and train k choose the same platform track; = 0 otherwise
$x_{l,i,t}$	0-1 platform track occupancy state variable = 1 if train l occupies platform track i at moment t; = 0 otherwise
$u_{l,i,t}$	0-1 platform track utilization state variable = 1 if train l has not yet arrived at platform track i at moment t; = 0 otherwise
$v_{l,i,t}$	0-1 platform track clearance state variable = 1 if train l has left platform track i at moment t; = 0 otherwise
$\lambda_{l,k}$	0-1 train arrival sequence variable = 1 if train l arrives at the station before train k; = 0 otherwise
$\mu_{l,k}$	0-1 train departure sequence variable = 1 if train l departs from the station before train k; = 0 otherwise
$y_{l,a}$	Actual arrival time of train l at the station
$y_{l,d}$	Actual departure time of train l at the station

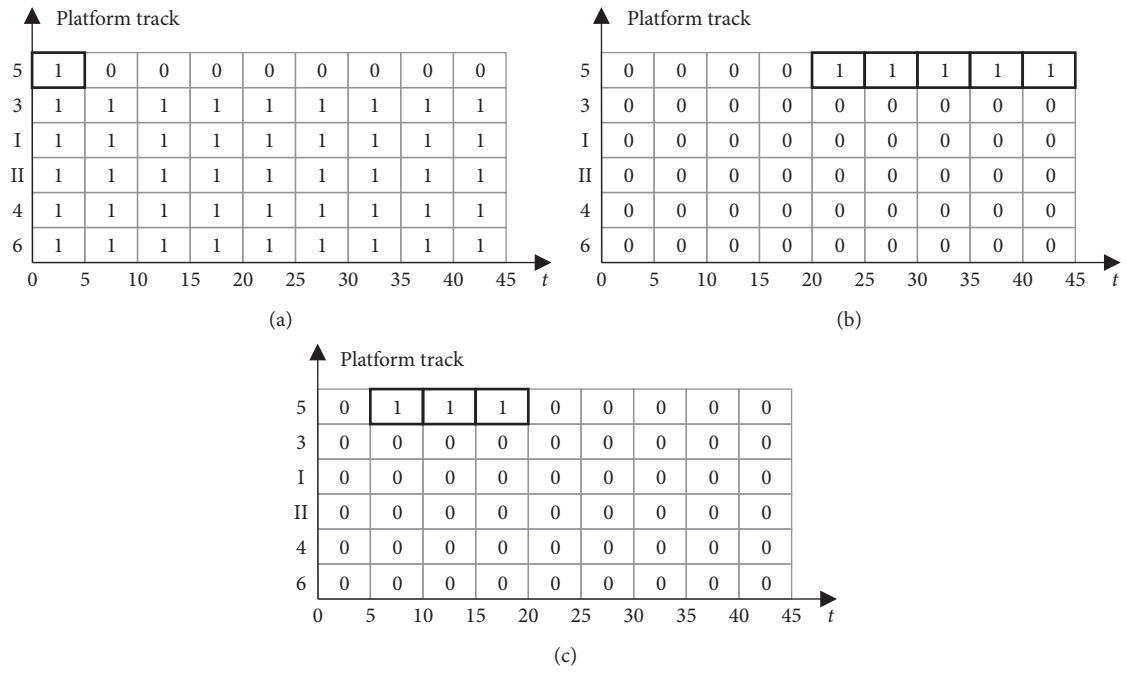


Figure 3: An illustrative example showing the values of platform track state variables $u_{l,i,t}$, $v_{l,i,t}$, and $x_{l,i,t}$: (a) values of the variable $u_{l,i,t}$; (b) values of the variable $v_{l,i,t}$; (c) values of the variable $x_{l,i,t}$, and $x_{l,i,t} = 1-(u_{l,i,t} + v_{l,i,t})$.

$$e_{l,k} \geq w_{l,i} + w_{k,i} - 1, \quad \forall l,k \in L, i \in I_l \cap I_k \colon k > l, \pi_l = \pi_k, \tag{15}$$

$$e_{l,k} = e_{k,l}, \quad \forall l,k \in L \colon k > l, \pi_l = \pi_k, \tag{16}$$

$$\lambda_{l,k} + \lambda_{k,l} = 1, \quad \forall l,k \in L \colon k > l, \pi_l = \pi_k, \tag{17}$$

$$\mu_{l,k} + \mu_{k,l} = 1, \quad \forall l,k \in L \colon k > l, \pi_l = \pi_k. \tag{18}$$

4.4.7. Minimum Train Dwell Time Constraint. Constraint (19) enforces the minimum dwell time for each train l. In particular, the right-hand side of constraint (19) can only be valid when $w_{l,i} = 1$, i.e., train l is assigned to the platform track i. Note that the safety time interval for platform track operation D is also included in the right-hand side of the constraints so that the required safety time interval for trains assigned to the same platform track is imposed. In addition, the minimum dwell time Δ_l of a train l is a deterministic value.

$$\sum_{t=1}^{\mathrm{MT}} x_{l,i,t} \geq w_{l,i}\left(\Delta_l + D\right), \quad \forall l \in L, i \in I. \tag{19}$$

4.4.8. Actual Train Arrival Time and Departure Time Constraint. Constraints (20)–(22) specify that the actual arrival and departure times of the trains should be no less than the corresponding planned arrival and departure times, respectively. Note that safety time interval D is introduced into the right-hand sides of constraints (21)-(22), such that the safety time interval D for platform track operation is implicitly satisfied between the departure time $y_{l,d}$ of train l and the arrival time of any other trains that are assigned to the same platform track.

$$y_{l,a} \geq t_{l,a}, \quad \forall l \in L, \tag{20}$$

$$y_{l,d} \geq t_{l,d} + D, \quad \forall l \in L, \tag{21}$$

$$y_{l,d} \geq y_{l,a} + \Delta_l + D, \quad \forall l \in L. \tag{22}$$

4.4.9. Domain of Variables. Constraints (23)–(25) define the domain of variables. Note that the variables $x_{l,i,t}$, $y_{l,a}$, and $y_{l,d}$ are intermediate variables to facilitate the model building process, and their values can be directly inferred from the variables $u_{l,i,t}$ and $v_{l,i,t}$.

$$w_{l,i} = \{0,1\}, \quad \forall l \in L, i \in I, \tag{23}$$

$$u_{l,i,t}, v_{l,i,t} = \{0,1\}, \quad \forall l \in L, i \in I, t \in \{1,2,\ldots,\mathrm{MT}\}, \tag{24}$$

$$e_{l,k},\ \lambda_{l,k},\ \mu_{l,k} = \{0,1\}, \quad \forall l,k \in L \colon l \neq k, \pi_l = \pi_k. \tag{25}$$

In short, formulations (3)–(25) constitute the complete MILP model for the reoptimization of train platforming problem. Due to the complicated filter conditions for the ranges of constrains (13)–(18), we simply approximate the number of constraints (13)–(18) without considering those filter conditions. As a result, the upper bound on the number of constraints of the MILP model is equal to $\mathrm{MT}|I|(2|L|+1) + 6|L| + |L|^2(5+|I|) + \sum_{l \in L}(|I| - |I_l|)$, which increases polynomial as the number of trains $|L|$, the number of platform tracks $|I|$, and the length of planning horizon T increase. Similarly, the upper bound on the number of variables is equal to $|L| \cdot |I|(1+\mathrm{MT}) + |L|^2$.

4.5. Valid Inequalities. Valid equalities are additional constraints that have been implicitly satisfied according to other necessary model constraints (4)–(25). However, they can be applied to strengthen the model formulation [17, 38]. Constraints (26)–(29) define the four valid inequalities that are related to the variables $u_{l,i,t}$, $v_{l,i,t}$, $x_{l,i,t}$, and $w_{l,i}$. Note that the valid inequalities are only implemented in a commercial solver to speed up the solving process of the commercial solver.

$$u_{l,i,t} \geq 1 - w_{l,i}, \quad \forall l \in L, i \in I, t \in \{1,2,\ldots,\mathrm{MT}\}, \tag{26}$$

$$v_{l,i,t} \leq w_{l,i}, \quad \forall l \in L, i \in I, t \in \{1,2,\ldots,\mathrm{MT}\}, \tag{27}$$

$$x_{l,i,t} \leq w_{l,i}, \quad \forall l \in L, i \in I, t \in \{1,2,\ldots,\mathrm{MT}\}, \tag{28}$$

$$x_{l,i,t} = 0, \quad \forall l \in L, i \in I, t \in \left\{1,2,\ldots,t_{l,a}-1\right\} \vee t \in \left\{t_{l,d} + \Delta_{\max} + D + 1,\ldots,\mathrm{MT}\right\}. \tag{29}$$

The principles of building valid inequalities (26)–(29) are similar. For instance, in valid inequality (26), if train l occupies platform track i, then valid inequality (26) is equivalent to $u_{l,i,t} \geq 0$ which turns out to be ineffective. However, if train l does not occupy the platform track i, then valid inequality (26) is equivalent to $u_{l,i,t} \geq 1$ which implies $u_{l,i,t} = 1$. Valid inequality (29) considers when the station capacity is not sufficient and two conflicting trains need to be assigned to the same platform track, then one of the two trains with lower priority can be delayed at most by $\Delta_{\max}$, which means $x_{l,i,t}$ can be constrained to 0 when $t \in \left\{1,2,\ldots,t_{l,a}-1\right\}$ or $t \in \left\{t_{l,d} + \Delta_{\max} + D + 1,\ldots,MT\right\}$. Overall, a total of $4\mathrm{MT}|L| \cdot |I| - \sum_{l \in L}|I|(t_{l,d} + \Delta_{\max} + D - t_{l,a} + 1)$ additional constraints are added into the model after introducing the valid inequalities (26)–(29).

5. Genetic and Simulated Annealing Hybrid Algorithm

In order to recover the train operations as soon as possible in case of train delays, a genetic and simulated annealing hybrid algorithm (GSAHA) is designed to solve the reoptimization model efficiently and effectively [37, 39, 40]. In particular, the simulated annealing algorithm (SA) has good ability in jumping out of local optimal solutions, and its convergence performance is robust against the generated initial solutions. However, the convergence speed of SA is generally slow. On

the other hand, the genetic algorithm (GA) is a parallel search algorithm that usually converges very fast, while GA is more likely to be trapped into local optimal solutions and it is not robust against the generated initial solutions. As a result, the GSAHA algorithm combines the advantages of GA and SA. Therefore, GSAHA is robust on the convergence performance while avoiding being trapped into the local optimal solutions [37, 40, 41]. Figure 4 demonstrates the solution framework of the GSAHA algorithm, including the section names corresponding to some key algorithmic steps. The implementation details for the components of GSAHA are illustrated as follows.

5.1. Chromosome Representation. Figure 5 shows the one-dimensional real-value encoding method that is used to represent chromosomes. Each chromosome denotes a platform track assignment plan. In particular, if the value of the l^{th} gene is equal to i, the l^{th} train is assigned to platform track i with its scheduled arrival and departure times. Furthermore, the length of each chromosome is equal to the number of trains $|L|$, and the genes of a chromosome are numbered in descending order according to the scheduled arrival times of trains. Moreover, the value range of each gene is located within the range $[1, |I|]$, and thus there could be $|I|^{|L|}$ possible chromosomes in total.

5.2. Generating Initial Population. Considering diversity and rationality of individuals in the initial population, the following strategies are proposed to generate the initial population.

Step 1. Denote platform tracks whose numbers are smaller than the number of platform tracks $|I|$ as the set I_1.

Step 2. Select $[|L|/(|I|-1)]$ trains that have not been selected yet and assign those trains to one of the unassigned platform tracks in set I_1.

Step 3. Repeat Step 1 until all platform tracks in set I_1 are assigned and assign the remaining $|L|-[|L|/(|I|-1)]\cdot(|I|-1)$ trains to the last platform track numbered as $|I|$.

Step 4. Repeat Step 2 and Step 3 until all individuals in the initial population are generated.

5.3. Obtaining Feasible Solutions. The chromosome designed in Section 5.1 only assigns trains to the platform tracks, i.e., to determine the platform track spatial resources that each train occupies. However, it is still possible that two trains may conflict with each other on the occupation of platform track temporal resources due to the violation of safety headway requirements, including the safety time interval between two trains assigned to the same platform track D, headway between two arrival trains running in the same direction h_a, and headway between two departure trains running in the same direction h_d. Hence, a heuristic rule is designed to resolve the temporal conflicts according to the constraints defined in Section 4.5.

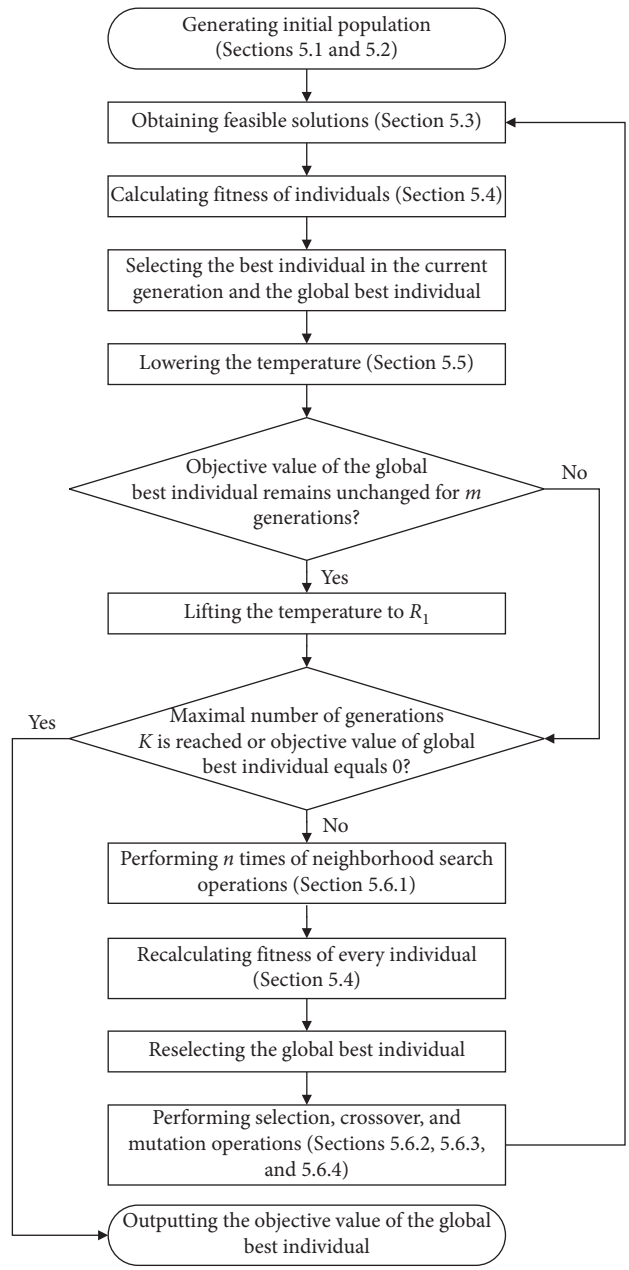


Figure 4: Solution framework of the GSAHA algorithm.

Platform track	5	2	3	6	...	4	7	1
Train	1	2	3	4	...	$\|L\|-2$	$\|L\|-1$	$\|L\|$

Figure 5: Illustration of the chromosome representation.

5.4. Fitness Function. The fitness function in equation (30) is designed to evaluate each individual such that the algorithm can achieve a better convergence performance.

$$f_i(r_k) = \exp\left\{-\frac{f(i) - f_{\min}}{r_k}\right\}, \tag{30}$$

where r_k represents the temperature at the k^{th} generation, $f(i)$ represents the objective value of the i^{th} chromosome,

For each train i $(1 \le i \le |L|)$
  For each train j $(1 \le j < i)$
    If train i conflicts with train j
      Fix the arrival and departure times of train i and record the weighted sum delay amount ϑ_i after resolving the conflicts of trains numbered before train i;
      Fix the arrival and departure times of train j and record the weighted sum delay amount ϑ_j after resolving the conflicts of trains numbered before train i;
      If $\vartheta_i \le \vartheta_i$
        Adopt the adjust method by fixing the arrival and departure times of train i;
      Else
        Adopt the adjust method by fixing the arrival and departure times of train j;
      End If $\vartheta_i \le \vartheta_i$
    End If train i conflicts with train j
  End For each train j $(1 \le j < i)$
End For each train i $(1 \le i \le |L|)$
Step 5. Sort all trains in descending order by their scheduled or estimated arrival times and number them from 1 to $|L|$.
Step 6. Use Algorithm 1 to resolve the temporal conflicts between any two trains in the order given in Step 1. Note that Algorithm 1 will not lead to a deadlock between trains where trains can always be delayed to resolve the temporal conflicts.
Step 7. Calculate the weighted sum of arrival and departure delays compared to the corresponding scheduled or estimated train arrival and departure times. This operation considers all trains in set L and the platform track assignment costs, and thus the value calculated during this step serves as the objective value of the chromosome.

Algorithm 1: A heuristic method to resolve the temporal conflicts between trains with given train order.

$f_{\min}$ represents the minimal objective value at the k^{th} generation, and $f_i(r_k)$ represents fitness value of the i^{th}chromosome when the temperature is r_k. Fitness function in equation (30) is an important feature of the SA algorithm, and it has a good capacity to accelerate the convergence of the algorithm.

5.5. Temperature Decline Function. After determining the initial temperature R, the temperature decline function in equation (31) is used to lower the temperature at each iteration.

$$r_k = R \cdot \eta^k, \tag{31}$$

where r_k represents the temperature at the k^{th} generation and the constant η represents the temperature decline rate.

5.6. Genetic Operators

5.6.1. Neighborhood Search. Neighborhood search operator is applied to every chromosome. For instance, neighborhood search operator modifies the value of one gene in chromosome i randomly to generate a new chromosome j, and the objective value $f(j)$ of chromosome j is recalculated. Chromosome j is accepted or rejected to replace chromosome i according to the probability $P_{ij}(r_k)$ in the following equation:

$$P_{ij}(r_k) = \min\left\{1,\ \exp\left(-\frac{f(j) - f(i)}{r_k}\right)\right\}. \tag{32}$$

If $P_{ij}(r_k)$ is greater than the random number δ_1 generated within the range [0, 1), then chromosome i is replaced by chromosome j. Neighborhood search operator is an important feature of the SA algorithm and it can enlarge the search space with the probability of resulting in better solutions. Moreover, neighborhood search operator is one of the main operators that can increase population diversity when the algorithm is trapped into local optimal solutions.

5.6.2. Selection. The roulette method is adopted to select parents according to the cumulative probability, as shown in the following equation:

$$C_i = \frac{\sum_{k=1}^{i} f_k}{\sum_{k=1}^{N} f_k}, \tag{33}$$

where N represents the number of individuals in the population. A random number δ_2 is generated within [0, 1); if $\delta_2 \in (C_i, C_j)$, then chromosome j is chosen as a parent. The elitism strategy is used to reduce randomness of the algorithm. Additionally, individuals are restricted to be consecutively chosen as parents to avoid the situation when the algorithm is trapped into a local optimal solution too early.

5.6.3. Crossover. Two individuals are chosen as parents each time and a random number δ_3 is generated within the range [0, 1). If δ_3 is greater than or equal to the given crossover rate λ^c, the crossover operator is not used and the two parents are reserved as children directly. Otherwise, the 2-opt crossover operator is performed. Figure 6 shows an example of the crossover operator where both the number of trains (i.e., $|L|$) and the number of platform tracks (i.e., $|I|$) are equal to 6.

5.6.4. Mutation. For each gene of a chromosome, a random number δ_4 is generated within the range [0, 1). If δ_4 is smaller than the given mutation rate λ^m, then the mutation

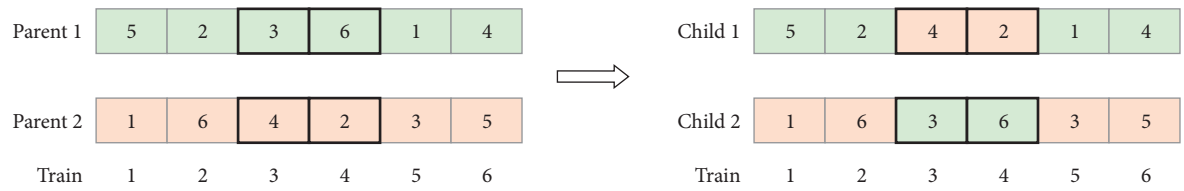


Figure 6: Illustration of the crossover operator.

operator is applied, i.e., a different platform track is randomly assigned to the gene.

6. Numerical Experiments

The proposed model is applied to a railway passenger station as shown in Figure 7, with five platform tracks (I, 3, 5, 7, 9, 11) in the downstream direction and four platform tracks (II, 4, 6, 8, 10) in the upstream direction. The time unit $\Delta\tau$ is set to 1 min, which is practical for ensuring safe train operations at the stations [22, 25, 37]. Moreover, there are 70 downstream and upstream trains, which need to conduct the necessary operations from 16:00 to 22:00. The initial scheduled arrival and departure times of the 70 trains are available at a repository on the ResearchGate website (DOI: 10.13140/RG.2.2.18253.79849), and trains are assigned with priority values ranging from 1 to 3. The initial platform track assignment plan is listed in Table 3, and the corresponding initially scheduled train operation plan is illustrated in Figure 8. Additionally, the platform track assignment costs for the downstream and upstream trains are given in Tables 4 and 5, respectively. In particular, there is a penalty of 10,000 for trains that use the platform tracks in the opposite direction.

The train delay information is available at the time of 18: 38, and it is known that 6 downstream trains and 4 upstream trains are delayed, and the estimated arrival and departure times of those delayed trains are provided in Table 6. Moreover, the 10 delayed trains in both directions are marked with the blue colour in Figure 8. Table 7 provides the parameter values used in the model. The maximum value of dwell times of all trains $\Delta_{\max}$ equals 43 min, and the length of the planning horizon T equals 360 min. The safety time interval on the platform track D equals 6 min, and the headway between two arrival trains h_a and the headway between two departure trains h_d in the same direction are both set to 5 min. The weighting factor α in the objective function is set to 200, which is determined after discussing with the train dispatchers, and the value of α can be adjusted flexibly according to the preferences of the train dispatchers. Note that the value of α can be flexibly adjusted by the train dispatchers according to their experiences. In addition, we believe that keeping trains on time with guaranteed train service quality is more important than assigning the trains to their preferred platform tracks, and thus the penalty parameters on train delays are relatively larger than the platform track occupancy costs. Moreover, when the values of input parameters are known, the initial values of the decision variables should be set in advance before running the GSAHA algorithm or commercial solvers. More specifically, $u_{l,i,1}$ is set to 1 and $v_{l,i,1}$ is to 0 for all trains $l \in L$ and platform tracks $i \in I$, such that no trains have arrived at or left the platform track at the beginning of the planning horizon. Furthermore, the variables $w_{l,i}$, $y_{l,a}$, and $y_{l,d}$ are set to their initial values before the train delay occurs, i.e., $w_{l,i} = q_{l,i}$, $y_{l,a} = t_{l,a}$, and $y_{l,d} = t_{l,d}$ for all trains $l \in L$ and platform tracks $i \in I$ with the condition of $t_{l,a} < S$.

Based on the initial scheduled arrival and departure times as well as the input data and parameters of 70 trains in Tables 3–7, a total of 21 scenarios are generated with the number of trains ranging from 10 to 70, and 3 scenarios are generated for each fixed number of trains. For instance, three scenarios (10, 2, 0), (10, 2, 1), and (10, 2, 2) are generated when the number of trains is equal to 10. Specifically, the first number "10" in the triple (10, 2, 0) denotes the number of trains, the second number "2" in the triple (10, 2, 0) represents the number of delayed trains. Besides, the third number in a triple can take the values of 0, 1, and 2 corresponding to the three scenarios, so that the estimated arrival and departure times for delayed trains in Table 6 are randomly increased with values in the ranges of (0, 0), (–1, 1), and (–2, 2), respectively.

The MILP models for the 21 scenarios in Section 4 are solved by the commercial solver CPLEX 12.7.0 with (i.e., CPLEX-WV) and without (i.e., CPLEX-WOV) including the valid inequalities in Section 4.5. Furthermore, the maximum computation times of CPLEX-WV and CPLEX-WOV are both set to 3600 s. The GSAHA is implemented in C++ programming language. Moreover, the algorithmic parameters for GSAHA are set as follows. The number of individuals in the population N is 50, the maximum number of generations K is 300, the crossover rate λ^c is 0.98, the mutation rate λ^m is 0.1, the initial temperature R is 80,000°C, the temperature decline rate η is 0.9, and the temperature is increased to $R_1 = 40{,}000$°C if the objective value of the best individual in the current generation remains unchanged for $m = 5$ iterations and $n = 3$ times of neighborhood search operations are performed each time. In addition, all of the models and algorithms are tested on a computer with Intel (R) Core (TM) i7-5600U 2.6 GHZ CPU and 12G RAM. Tables 8 and 9 list the detailed computation results of CPLEX-WOV, CPLEX-WV, and GSAHA for the 21 scenarios. More specifically, the columns "Obj 1 value (min)" and "Obj 2 value" in Table 8 represent the two components of objective values, i.e., the sum of train arrival and departure delays and total platform track assignment costs, respectively. In Table 9, the column "Obj value" denotes the overall objective values, the column "CPU time (s)" represents the computation times in sections, and the column "GAP with

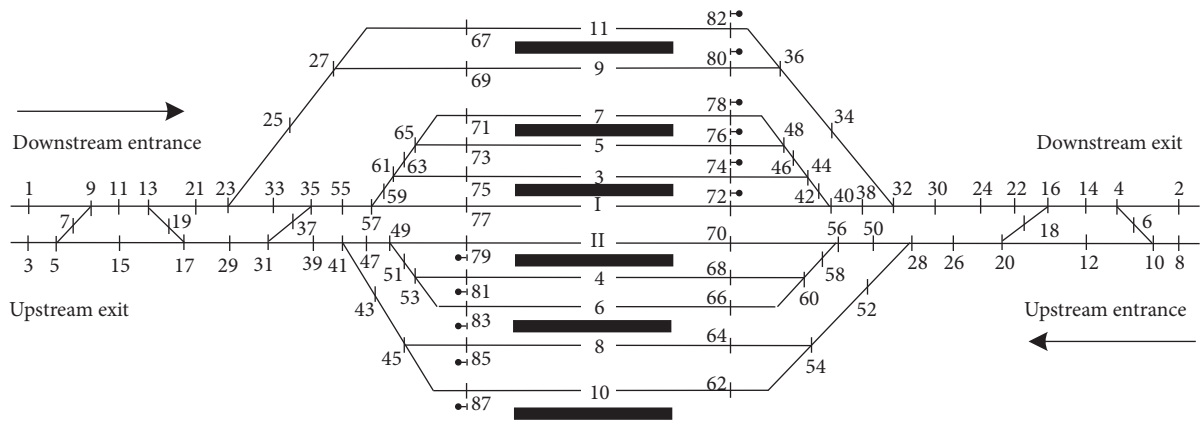


Figure 7: Layout of the railway passenger station.

Table 3: Initial platform tack assignment plan between 16:00 and 22:00.

Platform track number	Occupation trains
11	T_9, T_{29}
9	T_5, T_{19}, T_{31}, T_{41}, T_{49}
7	T_{11}, T_{21}, T_{27}, T_{33}, T_{43}, T_{47}, T_{55}, T_{63}, T_{69}
5	T_1, T_7, T_{15}, T_{25}, T_{35}, T_{39}, T_{53}, T_{61}, T_{67}, T_{73}
3	T_3, T_{13}, T_{17}, T_{23}, T_{37}, T_{45}, T_{51}, T_{57}, T_{59}, T_{65}, T_{71}, T_{75}
I	
II	
4	T_2, T_8, T_{18}, T_{22}, T_{30}, T_{34}, T_{40}, T_{42}, T_{48}, T_{54}, T_{60}, T_{64}
6	T_4, T_{12}, T_{16}, T_{24}, T_{32}, T_{38}, T_{44}, T_{50}, T_{56}, T_{62}
8	T_6, T_{14}, T_{20}, T_{28}, T_{36}, T_{46}, T_{52}, T_{58}
10	T_{10}, T_{26}

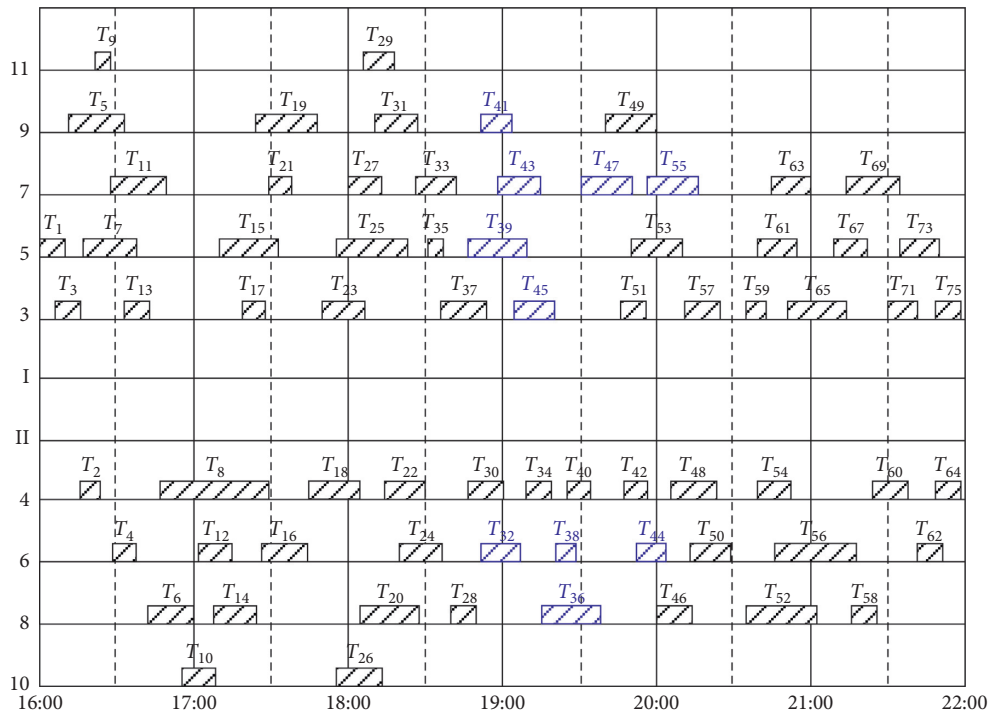


Figure 8: Initial platform track utilization scheme between 16:00 and 22:00.

Table 4: Platform track assignment costs for downstream trains with different priorities.

Train direction	Train priority	Platform track number					
		I	3	5	7	9	11
Downstream	1	600	6	12	24	48	96
	2	300	3	6	12	24	48
	3	200	2	4	8	16	32

Table 5: Platform track assignment costs for upstream trains with different priorities.

Train direction	Train priority	Platform track number					
		II	4	6	8	10	II
Upstream	1	600	6	12	24	48	600
	2	300	3	6	12	24	300
	3	200	2	4	8	16	200

Table 6: Estimated arrival and departure times for delayed trains.

Train	Arrival delay	Expected arrival time	Expected departure time	Dwell time	Priority
T_{39}	20	187	210	23	1
T_{41}	25	197	209	12	3
T_{43}	25	203	220	17	3
T_{45}	27	211	227	16	1
T_{47}	30	240	260	20	3
T_{55}	30	266	286	20	3
T_{32}	30	202	217	15	3
T_{36}	32	227	250	23	3
T_{38}	35	235	243	8	3
T_{44}	40	272	283	11	1

Table 7: Input parameters of the model.

Parameter	Value
Maximum value of dwell times of all trains Δ_{max}	43 min
Length of the planning horizon T	360 min
Safety time interval for platform track operation D	6 min
Sum of the length of the planning horizon and the safety time interval for platform track operation MT	366 min
Headway between two arrival trains h_a	5 min
Headway between two departure trains h_d	5 min
Objective function weighting factor α	200

CPLEX-WV (%)" indicates the relative gap between the objective values of GSAHA and CPLEX-WV.

Three points can be drawn from the computation results in Tables 8 and 9. First, the computation times of CPLEX-WOV, CPLEX-WV, and GSAHA increase gradually as the number of trains increases, and the maximum computation times of those three solution methods are equal to 3600 s, 679.48 s, and 28.96 s, respectively. In particular, even if CPLEX is sped up with the valid inequalities, the computation times of GSAHA are still much fewer than that of CPLEX. Therefore, it can be shown that our proposed GSAHA can solve the reoptimization problem more efficiently in a real-time manner. Second, the objective values of GSAHA are close to those of CPLEX-WV with the maximum gap value equal to 5.66%. Third, after comparing the computation times of CPLEX-WOV and CPLEX-WV, it can be seen the valid inequalities in Section 4.5 can significantly speed up the solving process of CPLEX. More specifically, CPLEX-WOV cannot obtain optimal solutions for the instances with the number of trains larger than or equal to 60 trains within 3600 s, which results in large objective values. In order to illustrate the feasibility of our proposed MILP model, the optimization solution of the scenario (70, 10, 0) using CPLEX-WV is illustrated in detail. Table 10 shows that the arrival and departure times of 11 trains are delayed after the reoptimization, and the new platform track assignment plan is provided in Table 11. Besides, the platform track utilization scheme is illustrated in Figure 9, and it can be seen from Figure 9 that all safety headway requirements are satisfied. Note that the 14 trains in Table 11 with bold fonts represent that those trains have been reassigned to different platform tracks, and those 14 trains are marked with the red colour in Figure 9. Besides, the convergence process of

Table 8: Two components of objective values of CPLEX-WOV, CPLEX-WV, and GSAHA under different scenarios.

Scenario	CPLEX-WOV		CPLEX-WV		GSAHA	
	Obj 1 value (min)	Obj 2 value	Obj 1 value (min)	Obj 2 value	Obj 1 value (min)	Obj 2 value
(10, 2, 0)	800	50	800	50	800	50
(10, 2, 1)	1200	50	1200	50	1200	50
(10, 2, 2)	800	50	800	50	800	50
(20, 3, 0)	1200	174	1200	174	1200	188
(20, 3, 1)	1800	174	1800	174	1800	174
(20, 3, 2)	1400	174	1400	174	1400	174
(30, 4, 0)	5200	278	5200	278	5200	286
(30, 4, 1)	5600	282	5600	282	5600	294
(30, 4, 2)	4400	278	4400	278	4400	283
(40, 6, 0)	800	363	800	363	800	379
(40, 6, 1)	1200	363	1200	363	1200	385
(40, 6, 2)	1200	363	1200	363	1200	395
(50, 7, 0)	8400	469	8400	469	8400	533
(50, 7, 1)	7000	469	7000	469	7000	545
(50, 7, 2)	8200	469	8200	469	8200	513
(60, 9, 0)	340400	1600	23600	583	24400	737
(60, 9, 1)	722400	2885	24200	575	25000	682
(60, 9, 2)	237800	2647	18200	579	18200	690
(70, 10, 0)	195400	1989	16400	659	16848	764
(70, 10, 1)	161600	1593	16800	685	17600	768
(70, 10, 2)	149800	2708	14600	659	15400	722

Table 9: Objective values and computation times of CPLEX-WOV, CPLEX-WV, and GSAHA under different scenarios.

Scenario	CPLEX-WOV		CPLEX-WV		GSAHA		
	Obj value	CPU time (s)	Obj value	CPU time (s)	Obj value	CPU time (s)	GAP with CPLEX-WV (%)
(10, 2, 0)	850	124.31	850	8.17	850	2.40	0
(10, 2, 1)	1250	51.73	1250	8.75	1250	2.43	0
(10, 2, 2)	850	67.49	850	8.23	850	2.34	0
(20, 3, 0)	1374	361.25	1374	24.66	1388	3.21	1.02
(20, 3, 1)	1974	377.20	1974	25.19	1974	3.10	0
(20, 3, 2)	1574	340.83	1574	24.02	1574	3.33	0
(30, 4, 0)	5478	326.84	5478	46.05	5486	4.75	0.15
(30, 4, 1)	5882	517.27	5882	53.25	5894	5.22	0.20
(30, 4, 2)	4678	208.92	4678	37.53	4683	5.01	0.11
(40, 6, 0)	1163	766.13	1163	46.30	1179	7.78	1.38
(40, 6, 1)	1563	480.83	1563	53.66	1585	7.83	1.41
(40, 6, 2)	1563	954.16	1563	55.27	1595	7.85	2.05
(50, 7, 0)	8869	1335.45	8869	105.22	8933	12.86	0.72
(50, 7, 1)	7469	1010.09	7469	87.42	7545	13.32	1.02
(50, 7, 2)	8669	1052.38	8669	121.58	8713	12.79	0.51
(60, 9, 0)	342000	3600	24183	661.22	25137	20.25	3.94
(60, 9, 1)	725285	3600	24775	678.17	25682	20.07	3.66
(60, 9, 2)	240447	3600	18779	604.38	18890	21.61	0.59
(70, 10, 0)	197389	3600	17059	679.48	17612	27.08	3.24
(70, 10, 1)	163193	3600	17485	344.52	18368	27.89	5.05
(70, 10, 2)	152508	3600	15259	529.75	16122	28.96	5.66

GSAHA is shown in Figure 10, where the algorithm can reach the near-optimal solution only after 70 iterations.

Meanwhile, for the scenario (70, 10, 0) in Tables 8 and 9, sensitivity analysis of different values of weighting factor α is performed by increasing the value of α from 0 to 10 with the step size equal to 1 for small values of α and from 40 to 440 with the step size equal to 40 for relatively larger values of α. Moreover, one special case is designed by setting $\alpha = 1$ and removing the second part of the objective function, which is aimed to simulate the case that α takes an infinity value. The optimization results of CPLEX-WV and GSAHA are listed in Table 12, and the parameter settings for the GSAHA remain unchanged. In addition, we take the objective value of GSAHA for the average optimization results of 20 runs. It can be shown that the objective values of CPLEX-WV and GSAHA increase as the value of α increases, and the solution

Table 10: Amount of secondary delay for the trains obtained by CPLEX-WV for the scenario (70, 10, 0).

Train	Priority	Secondary arrival delay (min)	Secondary departure delay (min)
T_{39}	1	0	4
T_{55}	3	0	2
T_{32}	3	0	3
T_{38}	3	2	2
T_{40}	3	2	2
T_{42}	1	5	5
T_{46}	2	2	2
T_{48}	1	2	2
T_{50}	1	0	1
T_{52}	3	2	2
T_{54}	1	2	2

Table 11: Platform tack assignment plan after the reoptimization using CPLEX-WV for the scenario (70, 10, 0).

Platform track number	Occupation trains
11	T_9, T_{29}
9	T_5, T_{19}, T_{31}, T_{41}, T_{49}, $\boldsymbol{T_{55}}$
7	T_{11}, T_{21}, T_{27}, T_{33}, T_{43}, $\boldsymbol{T_{53}}$, T_{63}, T_{69}
5	T_1, T_7, T_{15}, T_{25}, T_{35}, $\boldsymbol{T_{45}}$, $\boldsymbol{T_{47}}$, T_{61}, T_{67}, T_{73}
3	T_3, T_{13}, T_{17}, T_{23}, T_{37}, $\boldsymbol{T_{39}}$, T_{51}, T_{57}, T_{59}, T_{65}, T_{71}, T_{75}
I	
II	
4	T_2, T_8, T_{18}, T_{22}, $\boldsymbol{T_{28}}$, T_{34}, T_{40}, T_{42}, T_{48}, $\boldsymbol{T_{44}}$, T_{60}, T_{64}
6	T_4, T_{12}, T_{16}, T_{24}, $\boldsymbol{T_{30}}$, T_{32}, T_{38}, T_{50}, $\boldsymbol{T_{54}}$, $\boldsymbol{T_{58}}$, T_{62}
8	T_6, T_{14}, T_{20}, T_{46}, $\boldsymbol{T_{56}}$
10	T_{10}, T_{26}, $\boldsymbol{T_{36}}$, $\boldsymbol{T_{52}}$

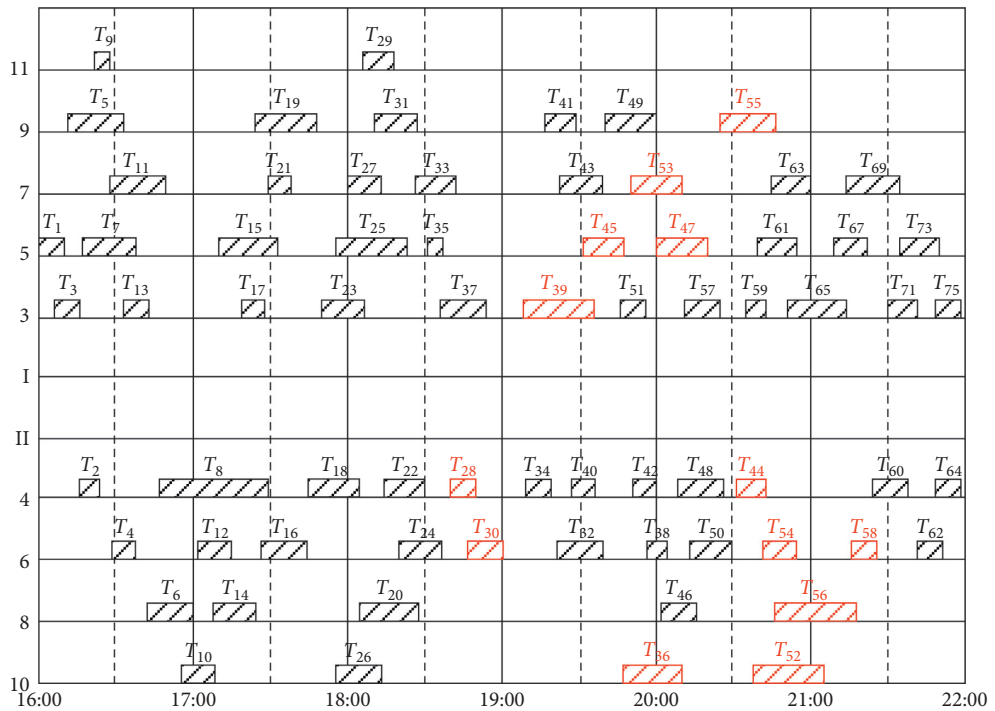


Figure 9: Platform track utilization scheme after the reoptimization using CPLEX-WV for the scenario (70, 10, 0).

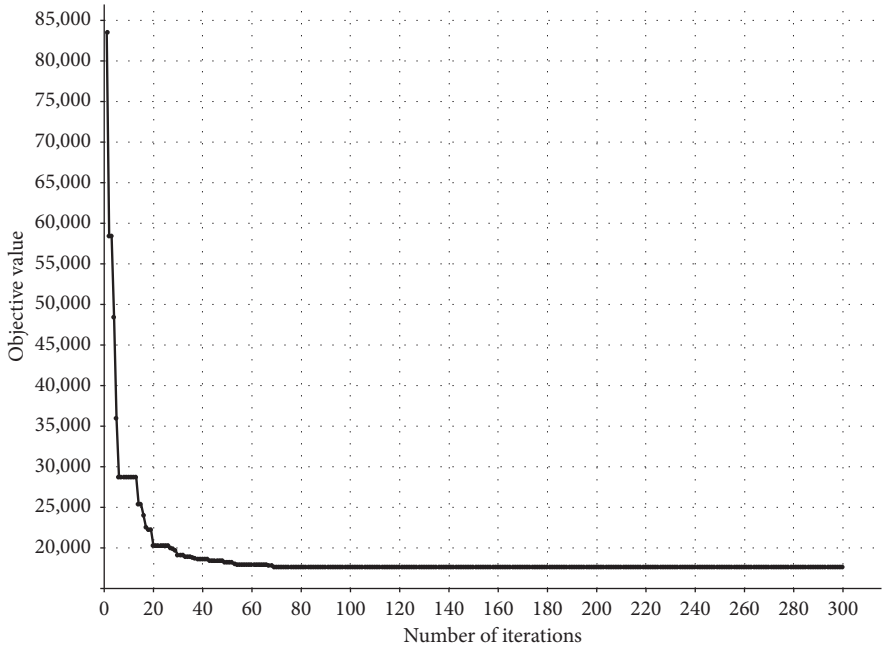


Figure 10: Convergence process of GSAHA for the scenario (70, 10, 0).

Table 12: Optimization results of CPLEX-WV and GSAHA with different values of weighting factor α.

Weighting factor α	CPLEX-WV				GSAHA				
	Obj 1 value (min)	Obj 2 value	Obj value	CPU time (s)	Obj 1 value (min)	Obj 2 value	Obj value	CPU time (s)	GAP with CPLEX-WV (%)
0	0	587	587	3600	0	477	477	32.92	−18.74
1	104	621	725	490.90	112	651	763	30.85	5.24
2	164	659	823	541	188	670	858	28.97	4.25
3	246	659	905	540.47	270	672	942	29.05	4.09
4	328	663	991	607.41	336	697	1033	29.17	4.24
5	410	663	1073	731.23	420	704	1124	29.33	4.75
6	492	659	1151	544.47	504	706	1210	29.02	5.13
7	574	659	1233	547.19	588	702	1290	28.99	4.62
8	656	663	1319	651.16	672	709	1381	29.07	4.70
9	738	669	1407	934.73	756	704	1460	29.09	3.77
10	820	663	1483	718.23	841	715	1556	28.30	4.92
40	3280	659	3939	740.22	3376	764	4140	27.60	5.10
80	6560	659	7219	589.28	6736	771	7507	28.20	3.99
120	9840	659	10499	764.37	10142	772	10914	27.52	3.95
160	13120	663	13783	447.63	13449	784	14233	27.51	3.26
200	16400	659	17059	679.48	16848	764	17612	27.08	3.24
240	19680	659	20339	388.56	20182	769	20951	27.25	3.01
280	22960	659	23619	359.88	23468	874	24342	27.49	3.06
320	26240	659	26899	595.69	26924	757	27681	27.65	2.91
360	29520	659	30179	329.28	30261	808	31069	28.23	2.95
400	32800	659	33459	340.13	33631	771	34402	28.77	2.82
440	36080	659	36739	412.67	37089	719	37808	28.22	2.91
∞	82	0	82	256.36	84	0	84	30.04	2.44

times of CPLEX-WV range from 256.36 to 3600 seconds while the solution times of GSAHA only range from 27.08 to 32.92 seconds. The last column in Table 12 shows that the objective values of GSAHA are −18.74%–5.10% larger than those of CPLEX-WV. In particular, CPLEX-WV cannot obtain the optimal solution for $\alpha = 0$ within the time limit, while GSAHA can obtain a solution that is much smaller than that of CPLEX-WV. In addition, for the case of $\alpha = \infty$, it can be shown that the minimum value of the first part of the objective function is equal to 82 min, and GSAHA can also obtain an objective value of 84 min. Note that the optimality gaps of GSAHA compared with CPLEX could be

due to two reasons in particular. First, the spatial and temporal resources are allocated separately in GSAHA. Second, a heuristic method is developed to resolve the temporal conflicts between trains. Overall, the stable performance of GSAHA regarding the solution quality and solving times shows that our proposed GSAHA is suitable to serve as an effective computer-aided decision-making tool for the train dispatchers in case of train delays.

7. Conclusions

The problem of reoptimization of the train platforming is essential in recovering the train operations within the station while minimizing the negative impact of train delays. This paper proposes a MILP model for the reoptimization problem, where the train station is represented with the discretized platform track time-space resources. In addition, a set of valid inequalities is developed to improve the performance of the commercial solver. In addition, in order to achieve the real-time reoptimization of the train platforming problem and provide a reliable decision tool for the train dispatchers, a heuristic algorithm called GSAHA that combines the advantages of SA and GA algorithms is designed. Moreover, GSAHA first allocates the spatial resources for the trains, and the temporal conflicts between trains are further resolved through a specialized heuristic rule. To test the efficiency and effectiveness of the proposed model and algorithm, a set of real-life instances is solved by a commercial solver and GSAHA. The computational results show the proposed MILP model can effectively achieve the reoptimization of the train platforming problem, and the proposed heuristic algorithm can further speed up the solving process of the commercial solver with near-optimal solutions. In addition, the computational results also show that the performance of GSAHA is stable even when the values of weighting factor α vary from 40 to 440.

The work in this paper can be extended in the following several interesting directions. First, instead of ensuring the arrival and departure safety headway between two different trains running in the same direction, the explicit consideration of train entrance and exit route conflicts can increase the station throughput capacity and reduce the train delays [10, 20]. Second, the MILP model and the heuristic algorithm GSAHA proposed in this paper can be further developed to consider different station types, such as the terminal station where trains need to perform the turnaround movement that makes the train platforming problem more complicated [42, 43]. Third, the train movements at the stations can be viewed as a set of space-time paths, and thus more efficient dual decomposition methods could be developed accordingly [7, 44, 45]. Fourth, more efficient solution algorithms can be developed for the simultaneous reoptimization of several interconnected railway stations in a railway line or even a regional railway network with more complicated structure so that the joint optimization of delay situations and train timetabling problems can be achieved [8, 19, 27, 46–49]. Finally, the integration of energy-efficient train movement with the reoptimization of train platforming can be considered for more system-wide benefits [50, 51].

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 they have no conflicts of interest.

Acknowledgments

This study was supported by the National Key R&D Program (grant no. 2017YFB1200700), National Natural Science Foundation of China (grant no. U1834209), and Open Fund Project of Chongqing Key Laboratory of Traffic & Transportation (grant no. 2018TE01).

References

[1] National Bureau of Statistics of China, 2020, http://data.stats.gov.cn/english/easyquery.htm?cn=C01.

[2] R. M. Lusby, J. Larsen, M. Ehrgott, and D. Ryan, "Railway track allocation: models and methods," *OR Spectrum*, vol. 33, no. 4, pp. 843–883, 2011.

[3] A. A. Assad, "Models for rail transportation," *Transportation Research Part A: General*, vol. 14, no. 3, pp. 205–220, 1980.

[4] J.-F. Cordeau, P. Toth, and D. Vigo, "A survey of optimization models for train routing and scheduling," *Transportation Science*, vol. 32, no. 4, pp. 380–404, 1998.

[5] U. Brännlund, P. O. Lindberg, A. Nõu, and J.-E. Nilsson, "Railway timetabling using Lagrangian relaxation," *Transportation Science*, vol. 32, no. 4, pp. 358–369, 1998.

[6] A. Caprara, M. Fischetti, and P. Toth, "Modeling and solving the train timetabling problem," *Operations Research*, vol. 50, no. 5, pp. 851–861, 2002.

[7] Y. Zhang, Q. Peng, Y. Yao, X. Zhang, and X. Zhou, "Solving cyclic train timetabling problem through model reformulation: extended time-space network construct and Alternating Direction Method of Multipliers methods," *Transportation Research Part B: Methodological*, vol. 128, pp. 344–379, 2019.

[8] Y. Zhang, A. D'Ariano, B. He, and Q. Peng, "Microscopic optimization model and algorithm for integrating train timetabling and track maintenance task scheduling," *Transportation Research Part B: Methodological*, vol. 127, pp. 237–278, 2019.

[9] L. G. Kroon, H. Edwin Romeijn, and P. J. Zwaneveld, "Routing trains through railway stations: complexity issues," *European Journal of Operational Research*, vol. 98, no. 3, pp. 485–498, 1997.

[10] P. J. Zwaneveld, L. G. Kroon, H. E. Romeijn et al., "Routing trains through railway stations: model formulation and algorithms," *Transportation Science*, vol. 30, no. 3, pp. 181–194, 1996.

[11] P. J. Zwaneveld, L. G. Kroon, and S. P. M. Van Hoesel, "Routing trains through a railway station based on a node packing model," *European Journal of Operational Research*, vol. 128, no. 1, pp. 14–33, 2001.

[12] R. Lusby, J. Larsen, D. Ryan, and M. Ehrgott, "Routing trains through railway junctions: a new set-packing approach," *Transportation Science*, vol. 45, no. 2, pp. 228–245, 2011.

[13] R. M. Lusby, J. Larsen, M. Ehrgott, and D. M. Ryan, "A set packing inspired method for real-time junction train routing,"

Computers & Operations Research, vol. 40, no. 3, pp. 713–724, 2013.

[14] V. Cacchiani, D. Huisman, M. Kidd et al., "An overview of recovery models and algorithms for real-time railway rescheduling," *Transportation Research Part B: Methodological*, vol. 63, pp. 15–37, 2014.

[15] L. Meng and X. Zhou, "Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 1080–1102, 2011.

[16] 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.

[17] S. Zhan, L. G. Kroon, L. P. Veelenturf, and J. C. Wagenaar, "Real-time high-speed train rescheduling in case of a complete blockage," *Transportation Research Part B: Methodological*, vol. 78, pp. 182–201, 2015.

[18] S. Zhan, L. G. Kroon, J. Zhao, and Q. Peng, "A rolling horizon approach to the high speed train rescheduling problem in case of a partial segment blockage," *Transportation Research Part E: Logistics and Transportation Review*, vol. 95, pp. 32–61, 2016.

[19] M. Carey, "A model and strategy for train pathing with choice of lines, platforms, and routes," *Transportation Research Part B: Methodological*, vol. 28, no. 5, pp. 333–353, 1994.

[20] P. Chakroborty and D. Vikram, "Optimum assignment of trains to platforms under partial schedule compliance," *Transportation Research Part B: Methodological*, vol. 42, no. 2, pp. 169–184, 2008.

[21] A. Caprara, L. Galli, and P. Toth, "Solution of the train platforming problem," *Transportation Science*, vol. 45, no. 2, pp. 246–257, 2011.

[22] L. Kang, Z. Lu, and Q. Meng, "Stochastic schedule–based optimization model for track allocations in large railway stations," *Journal of Transportation Engineering, Part A: Systems*, vol. 145, no. 3, Article ID 04019001, 2019.

[23] D. D. L. Cardillo and N. Mione, "k L-list λ colouring of graphs," *European Journal of Operational Research*, vol. 106, no. 1, pp. 160–164, 1998.

[24] A. Billionnet, "Using integer programming to solve the train-platforming problem," *Transportation Science*, vol. 37, no. 2, pp. 213–222, 2003.

[25] P. Sels, P. Vansteenwegen, T. Dewilde, D. Cattrysse, B. Waquet, and A. Joubert, "The train platforming problem: the infrastructure management company perspective," *Transportation Research Part B: Methodological*, vol. 61, pp. 55–72, 2014.

[26] M. Carey and S. Carville, "Scheduling and platforming trains at busy complex stations," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 3, pp. 195–224, 2003.

[27] M. Carey and I. Crawford, "Scheduling trains on a network of busy complex stations," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 159–178, 2007.

[28] L. Kang, J. Wu, and H. Sun, "Using simulated annealing in a bottleneck optimization model at railway stations," *Journal of Transportation Engineering*, vol. 138, no. 11, pp. 1396–1402, 2012.

[29] J. Wu, L. Kang, H. Sun, and X. Jia, "Track allocation optimization in railway station: mean-variance model and case study," *Journal of Transportation Engineering*, vol. 139, no. 5, pp. 540–547, 2013.

[30] M. J. Dorfman and J. Medanic, "Scheduling trains on a railway network using a discrete event model of railway traffic," *Transportation Research Part B: Methodological*, vol. 38, no. 1, pp. 81–98, 2004.

[31] A. D'ariano, D. Pacciarelli, and M. Pranzo, "A branch and bound algorithm for scheduling trains in a railway network," *European Journal of Operational Research*, vol. 183, no. 2, pp. 643–657, 2007.

[32] J. Rodriguez, "A constraint programming model for real-time train scheduling at junctions," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 231–245, 2007.

[33] A. D'Ariano, L. Meng, G. Centulio, and F. Corman, "Integrated stochastic optimization approaches for tactical scheduling of trains and railway infrastructure maintenance," *Computers & Industrial Engineering*, vol. 127, pp. 1315–1335, 2019.

[34] T. M. Herrman, *Stability of timetables and train routings through station regions*, Ph.D. thesis, Swiss Federal Insitute of Technology Zurich, Zürich, Switzerland, 2006.

[35] N. Bešinović and R. M. Goverde, "Stable and robust train routing in station areas with balanced infrastructure capacity occupation," *Public Transport*, vol. 11, no. 2, pp. 211–236, 2019.

[36] A. D'Ariano and M. Pranzo, "An advanced real-time train dispatching system for minimizing the propagation of delays in a dispatching area under severe disturbances," *Networks and Spatial Economics*, vol. 9, no. 1, pp. 63–84, 2009.

[37] W. Liu, X. Zhu, and L. Kang, "Real-time track reallocation for emergency incidents at large railway stations," *Mathematical Problems in Engineering*, vol. 2015, Article ID 296394, 11 pages, 2015.

[38] L. Kang, S. Chen, and Q. Meng, "Bus and driver scheduling with mealtime windows for a single public bus route," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 145–160, 2019.

[39] W. Xing and J. Xie, *Modern Optimization Methods*pp. 113–147, Tsinghua University Press, Beijing, China, 2nd edition, 2006.

[40] H. Yu, H. Fang, P. Yao, and Y. Yuan, "A combined genetic algorithm/simulated annealing algorithm for large scale system energy integration," *Computers & Chemical Engineering*, vol. 24, no. 8, pp. 2023–2035, 2000.

[41] H. Du, J. Fan, X. He, and M. W. Feldman, "A genetic simulated annealing algorithm to optimize the small-world network generating process," *Complexity*, vol. 2018, Article ID 1453898, 12 pages, 2018.

[42] Q. Zhong, R. M. Lusby, J. Larsen, Y. Zhang, and Q. Peng, "Rolling stock scheduling with maintenance requirements at the Chinese high-speed railway," *Transportation Research Part B: Methodological*, vol. 126, pp. 24–44, 2019.

[43] Q. Zhong, Y. Zhang, D. Wang, Q. Zhong, C. Wen, and Q. Peng, "A mixed integer linear programming model for rolling stock deadhead routing before the operation period in an urban rail transit line," *Journal of Advanced Transportation*, vol. 2020, Article ID 3809734, 18 pages, 2020.

[44] X. Zhou, L. Tong, M. Mahmoudi et al., "Open-source VRPLite package for vehicle routing with pickup and delivery: a path finding engine for scheduled transportation systems," *Urban Rail Transit*, vol. 4, no. 2, pp. 68–85, 2018.

[45] X. Chen, S. He, Y. Zhang, L. Tong, P. Shang, and X. Zhou, "Yard crane and AGV scheduling in automated container terminal: a multi-robot task allocation framework," *Transportation Research Part C: Emerging Technologies*, vol. 114, pp. 241–271, 2020.

[46] L. Kang, J. Wu, H. Sun, X. Zhu, and B. Wang, "A practical model for last train rescheduling with train delay in urban railway transit networks," *Omega*, vol. 50, pp. 29–42, 2015.

[47] L. Kang, X. Zhu, H. Sun, J. Wu, Z. Gao, and B. Hu, "Last train timetabling optimization and bus bridging service management in urban railway transit networks," *Omega*, vol. 84, pp. 31–44, 2019.

[48] L. Kang and Q. Meng, "Two-phase decomposition method for the last train departure time choice in subway networks," *Transportation Research Part B: Methodological*, vol. 104, pp. 568–582, 2017.

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

[50] W. Li, Q. Peng, Q. Li, C. Wen, Y. Zhang, and J. Lessan, "Joint operating revenue and passenger travel cost optimization in urban rail transit," *Journal of Advanced Transportation*, vol. 2018, Article ID 7805168, 15 pages, 2018.

[51] W. Li, Q. Peng, C. Wen, S. Li, X. Yan, and X. Xu, "Integrated optimization on energy saving and quality of service of urban rail transit system," *Journal of Advanced Transportation*, vol. 2020, Article ID 3474020, 22 pages, 2020.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 5740521, 13 pages
https://doi.org/10.1155/2020/5740521

Research Article

A Method for Bus OD Matrix Estimation Using Multisource Data

Di Huang,[1] **Jun Yu**,[1] **Shiyu Shen**,[2] **Zhekang Li**,[1] **Luyun Zhao**,[2] **and Cheng Gong**[2]

[1]*Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing 211189, China*
[2]*Didi Chuxing, Beijing 100085, China*

Correspondence should be addressed to Di Huang; dhuang2@seu.edu.cn

Received 14 December 2019; Revised 24 January 2020; Accepted 3 February 2020; Published 21 March 2020

Guest Editor: Tao Liu

The automated fare collection (AFC) system has gained increasing popularity among transit systems worldwide. The AFC system is usually an entry-only system that only records the serial number of the smart card and the transaction time of each use. Neither the AFC data nor the bus global positioning system (GPS) could reveal the passenger's alighting information, namely, alighting time and station. Hence, the station-to-station origin-destination (OD) trip information cannot be obtained directly from the available data sources. To address this problem, this paper proposes a methodology that estimates the OD matrix by using smart card and GPS data. In this paper, the characteristics of the basic data sources are first analyzed, based on which the bus arrival time is generated using the density-based clustering algorithm and a time correction strategy, based on which the passenger's boarding station is identified. The alighting stations are inferred based on the characteristics of bus trip chaining, which could identify over 80% of the alighting stations on average. Finally, the proposed methodology is verified by a comprehensive field survey in Suzhou, China, with 100% sample rate.

1. Introduction

The bus origin-destination (OD) matrix estimation is one of the most fundamental steps for urban transit planning, operation, and management strategies, such as passenger demand forecasting [1, 2], network design [3], transit pricing [4, 5], scheduling [6, 7], and bus operation [8, 9]. Traditionally, the acquisition of OD trip information as well as the analysis of passenger's spatial-temporal travel behavior relies largely on travel survey data, which has the drawbacks of small sample size, high costs, and being time consuming [10–14]. With the help of advanced automatic fare collection (AFC) system, the smart card data has received increasing interest as a new and reliable data source for the collection and analysis of passenger trip information [15]. The statistical performance of smart card data outperforms that of the field survey data for providing comprehensive spatial-temporal information about the transit system and capturing the dynamics of passenger trips [16, 17]. The smart card data could also detect and correct the potential sources of sampling bias [18].

There has been some work in mining the smart card data and other types of data collected during the bus operation, such as bus global positioning system (GPS) data and WiFi/Bluetooth probes data. Pelletier et al. [19] have conducted an extensive review of the cutting-edge technique of smart card data mining in public transit system. Early studies on the smart card data analysis mainly focus on the statistical analysis of urban public transport passenger flow. For example, both Kiyohiro et al. [20] and Ouyang et al. [21] use smart card data to analyze the travel pattern of urban residents. Ebadi et al. [22] utilize smart card data to construct students' activity-mobility trajectories in time-space dimension. Ji et al. [23] propose a Bayesian model to estimate trip-level OD flow matrices utilizing the data collected by Wi-Fi sensors and boarding data provided by automatic vehicle location (AVL) systems. However, compared to the smart card data, the stability of Wi-Fi sensors and the reliability of collected data are relatively low. Ma et al. [24] investigate the smart card transactions and GPS data. The results show relatively high accuracy of the passenger's origin information. Ma et al. [25] continue to mine the smart card data and

propose an efficient method to identify the passenger's travel patterns using passenger's historical travel data.

It has been widely recognized that the application of data fusion techniques which consider various correlated data sources can help to improve the accuracy of the estimation of OD trip matrix [26–29]. In such data fusion process, the following two issues have to be addressed: (1) the smart card data contains only transaction-related information, which is useful only when it can be matched with the vehicle's operational data while the boarding and alighting locations are not recorded [30]; (2) the information recorded by the smart card system might not be matched perfectly with that data of the GPS system. To identify the boarding station, two kinds of techniques are mainly adopted, namely, clustering analysis and GPS data matching. Specifically, the swiping time of ID card is first clustered and then matched with the AVL data to identify bus stops. Many efforts have been devoted for combining the smart card data with the bus dispatching information. Zhang et al. [31] identify the commuting trips of passengers by using their smart card data and use the clustering analysis method to identify commuting passengers' boarding stations according to the time recorded in the smart card data. Farzin [32] uses the GPS time matching to identify the boarding location through the correlation matching of bus IC card swiping information, GPS positioning information, and bus station locations in São Paulo, Brazil.

The identification of the alighting station is more challenging because of the absence of available alighting location information [19]. Two kinds of methods have been proposed to estimate passenger's alighting station: aggregate and disaggregate methods. In the aggregate method, passengers are assumed to alight from the bus according to a specific probability distribution with respect to the travel distance and station attractiveness [33]. To obtain more reliable estimations, a large number of studies focus on the disaggregate method based on the passenger's trip chain by combining the smart card data with other sources of data, e.g., AVL data, which provides the detailed trip and corresponding geographic information.

The trip chain is generated to describe the same individual's daily travel which is composed of several bus trips with the assumption that the passenger's boarding station of a trip is close to the alighting station of the last trip [34]. Wang et al. [35] apply the trip chaining to obtain the OD information using smart card transactions and AVL data in London. It is also the first attempt to validate the results with manual passenger survey data. However, the sample rate of some bus lines is about 60%. Barry et al. [13] infer the alighting station based on the data collected by AFC system in New York, USA. Two specific assumptions are made to define the individual's trip chain: (1) a large proportion of passengers returns to the previous station to start the next trip and (2) passengers would begin their first trip of the day at the station where they end their trip of the previous day. Zhao [36] examines the rail-to-bus trip chain by integrating the AFC and AVL data, as well as the GIS technology. It was demonstrated that the passenger's travel behavior can be obtained rigorously by using the advanced automated data collection system. Seaborn et al. [37] use the smart card fare payment data to investigate the passenger's multimodal trips in London. They illustrate that passengers between intersecting routes can be identified by the smart card data. Cui [38] first estimates the OD matrix of a single route and then extends it to the network-level OD matrix estimation. A case study of a full-size network of Chicago is conducted. The results show that the maximum likelihood estimation is suitable for the estimation of the single route OD matrix, while the proportional distribution method is recommended for the estimation of the transfer flow OD matrix in the network level. Lu et al. [39] use the AFC data from Beijing metro based on the trip chaining method and k-means clustering method to infer the actual destination, which could also be applied to the bus passengers' alighting station inference based on the smart card data. In sum, the above-mentioned studies on the passenger's trip chain are based on two assumptions: (1) the closest stop and (2) the daily symmetry rules. The identification rate of the passenger's alighting station using the trip chain method is between 67% and 71% [33, 36, 38].

In literature, bus OD matrix can be estimated in three ways, i.e., field survey, analysis of smart card data only, and the fusion of multisource data including smart card data and bus GPS data. Although existing literature has extensively studied the characteristics of different types of data, dealing with the heterogeneity and systematic error of multisource data is still an open question. Specifically, the gaps of current studies are summarized as follows: (1) the fusion of multisource data does not allow for the internal correlation in both spatial and temporal dimensions; (2) the inference of alighting stations is accurate on an aggregate level but inaccurate on an individual level; and (3) the estimated results cannot be verified comprehensively because of the absence of real data and inadequate sample rate through field experiments.

To sum up, the contribution of this study is threefold. First, this paper proposes a spatiotemporal correction method for smart card data, bus GPS data, and bus station data to address incompleteness and complexity issues of multisource data. A density-based clustering method is applied to correct the difference between the trajectories recorded by the GPS devices and the bus station data. Second, the bus arrival time is obtained through data clustering technique in the spatial dimension. An additional correction algorithm in the temporal dimension is proposed to calibrate the timestamps in smart card data to match the bus arrival time. To identify the alighting station in the individual level, this paper divides the trips into the chained dataset and unchained dataset and infers the alighting stations for individuals for the chained dataset, through which the identification rate of the alighting station can be improved by about 10%. Third, the effectiveness and feasibility of the proposed method are verified on the data collected by a large-scale field survey in Suzhou, China.

The rest of this paper is orgnized as follows.The problem statement and data utilized in this paper are introduced in Section 2. Section 3 discusses the method for identifying boarding stations by incorporating the heterogeneity in multisource data. Section 4 presents the categories of trips and the trip-chaining-based inference method for alighting

Table 1: Structure of smart card data.

Field	Field name	Description
RECORD_ID	Record number	Unique identification of a smart card record
CARD_UUID	IC card number	Unique identification of smart card
SWIPED_TYPE	Card type	General card/senior card/student card
SWIPED_TIME	Swiping time	2018-03-01 13:17:54
LINE_UUID	Line number	Bus line number
BUS_UUID	Vehicle number	Bus number
FARE	Fare	Bus fare

Table 2: Structure of bus GPS data.

Field	Field name	Description
RECORD_ID	Record number	Unique identification of a GPS equipment
BUS_UUID	Vehicle number	Unique identification of a vehicle
LINE_UUID	Line number	Unique identification of a bus line
LINE_NAME	Line name	Name of a bus line
DATA_TYPE	Type of data	Arriving/leaving a station, in-route
RECORD_TIME	GPS time	2018-03-01 13:17:54
LNG	Longitude	Longitude of a turning point
LAT	Latitude	Latitude of a turning point
ALT	Altitude	Altitude of a turning point
GPS_SPEED	GPS speed	GPS speed of the vehicle
ROTATE_ANGLE	Rotation angle	Rotation angle of a turning point

stations. Section 5 conducts a case study in Suzhou, China, to validate the performance of the proposed algorithm. Finally, we conclude the paper with some remarks and perspectives.

2. Data Description

Three crucial datasets can be obtained in the urban transit system, i.e., smart card data, bus GPS data, and bus station information data. Smart card data records the swiping time, vehicle number, and other transaction-related information of passengers, which contains mainly temporal information. In most cases, the station at which a passenger boards is not recorded. Bus GPS data records the trajectories of bus vehicles, which contain both spatial and temporal information. Bus station data records the static positions of bus stations of each bus line, which only contain the spatial information. The raw data of bus GPS and bus station location sometimes do not match because of different coordinate systems. In the rest of this section, the detailed description of these datasets is given.

2.1. Smart Card Data. The smart card data is recorded by the card swiping terminal equipped on each bus. When passengers board the vehicle, their trip information will be recorded. The recorded fields and corresponding description are listed in Table 1, including the record ID, card ID, card type, card swiping time, bus line number, and vehicle number, which are mainly temporal information.

2.2. Bus GPS Data. In practice, most buses in large- and medium-sized cities have been equipped with GPS terminals, which can accurately collect the real-time positioning information, along with the bus line number and operating direction, which contains both spatial and temporal information.

Table 3: Structure of bus station location data.

Field	Field name	Description
STOP_UUID	Number of bus line	Unique identification of a bus stop
STOP_NAME	Name of stops	Name of a bus stop
STOP_TYPE	Type of bus stop	Upward or downward
LNG	Longitude	Longitude of a bus station
LAT	Latitude	Latitude of a bus station

As shown in Table 2, the data fields include the record ID, vehicle number, bus type, bus line number, bus line name, operating status, timestamp, longitude, latitude, altitude, running speed, and running angle. The system error of the GPS data is about 10–30 meters. However, in practice, some of the GPS equipment may be obsolete and cannot be corrected in time. The systematic error would reach 50 meters.

2.3. Bus Station Location Data. The bus network information is stored in relational database in the form of the bus route information and the location information of bus stations (see Table 3). In the relational database, bus lines and bus stations are regarded as entities. A bus station belongs to at least one bus line, and a bus line contains at least one bus station. The major data fields include line number, line name, line direction, the stating and end location, and line type.

3. Passenger Boarding Station Extraction

The first step of bus OD matrix estimation is to obtain the passenger's boarding station, i.e., identifying the origin of each trip. As mentioned in the previous section, the smart card data only records the transaction time, while the spatial

information of the boarding station is absent. As for the GPS data, only the vehicle trajectories are recorded, while the spatial relationship between trajectories and the location of stations is also unclear. Hence, it is necessary to match the smart card data to the bus GPS data to extract the passenger's boarding station. In this section, we first use the clustering technique to obtain the bus arrival time, which can be further matched to the smart card's transaction time, and then extract the passenger's boarding station.

3.1. Obtaining Bus Arrival Time Based on Density Clustering Algorithm

3.1.1. Feasibility Analysis of Density Clustering Algorithm. Through the spatial matching of the bus station data and bus GPS data, the bus arrival time at each station can be obtained. However, it is difficult to directly create reference from the coordinates in bus trajectories to the location of bus stations, in that differences exist in the geographic coordinate systems, and bus GPS data can suffer from interference. These uncertainties and disturbances can be formulated as follows:

$$\begin{aligned} \mathrm{lng}_{\mathrm{LS}}^{(i,j)} &= \mathrm{lng}_{\mathrm{GPS}}^{(i,j,k)} + C_{\mathrm{lng}} + \delta_{\mathrm{lng}}^{(i,j,k)}, \\ \mathrm{lat}_{\mathrm{LS}}^{(i,j)} &= \mathrm{lat}_{\mathrm{GPS}}^{(i,j,k)} + C_{\mathrm{lat}} + \delta_{\mathrm{lat}}^{(i,j,k)}, \end{aligned} \tag{1}$$

where $\mathrm{lng}_{\mathrm{LS}}^{(i,j)}$ and $\mathrm{lat}_{\mathrm{LS}}^{(i,j)}$ represent the coordinates of bus station j on bus line i recorded in the bus station data; $\mathrm{lng}_{\mathrm{LS}}^{(i,j)}$ and $\mathrm{lat}_{\mathrm{LS}}^{(i,j)}$ represent the coordinates of the k-th sample of the bus station j on bus line i in the bus GPS data; C_{lng} and C_{lat} represent the conversion errors between the geographic coordinate systems adopted by the two different data sources; and $\delta_{\mathrm{lng}}^{(i,j,k)}$ and $\delta_{\mathrm{lat}}^{(i,j,k)}$ represent the positioning errors of the k-th sample of bus station j on bus line i in the bus GPS data. Due to the existence of these errors, it is essential to ensure the correctness of the spatial clustering and matching of data from two different sources.

The clustering algorithm is an unsupervised learning method which is capable of classifying the data into several groups by analyzing the similarity and mutuality between observed samples. The density-based clustering algorithm can further investigate the connectivity of the data from the perspective of the local density of samples. For the bus GPS data, a vehicle usually decelerates when approaching the station and then stops if this station is not skipped. This motion pattern leads to the spatial aggregation of trajectory points near bus stations, which is suitable for the use of density-based clustering algorithm. In the following section, the density-based spatial clustering of applications with noise (DBSCAN) algorithm is applied to obtain the bus station based on the GPS data [25], while the errors corresponding to the direct data fusion can therefore be avoided.

3.1.2. Obtaining the Bus Arrival Time Based on DBSCAN Algorithm. The DBSCAN algorithm can be first applied on the bus GPS data, whereby the location of bus stations in the bus GPS data can be identified. Then, the bus GPS data is further matched with the bus station data based on the connectivity between trajectory points to generate bus arrival timetable without being affected by the errors from different data sources. The detailed steps of the proposed method for bus timetable generation are as follows:

Step 1: Density-based clustering

The DBSCAN algorithm involves two steps, namely, searching for core samples and generating clusters. Before executing the algorithm, the bus station data is first matched with the bus GPS data. In the DBSCAN algorithm, two key parameters need to be defined: ε, distance, and MinPts, the minimum number of points. The ε is defined to measure the density-reachable range, within which the points are considered as the neighborhood of the same cluster. The MinPts limits the maximum number of points in the same cluster. Through DBSCAN, the samples are split into two categories, i.e., core samples and noncore samples. Instead of randomly selecting a core sample from the complete dataset, we only select the sample from the bus GPS data as a clustering seed. The cluster grows from the seed to its neighboring samples, and the seed is then labelled as the primary core sample. By repeating these steps, the clusters and their primary core samples can be generated.

Step 2: Searching for the bus stations

After finding all the clusters in the dataset, the bus station corresponding with the samples needs to be determined. Firstly, if the sample is directly density reachable from a primary core sample, it will be labelled as the core sample of the cluster that the primary core sample belongs to and connected with the same bus station of the primary core sample. Secondly, if a sample is density reachable from exactly one primary core sample, it will also be matched with the bus station of the primary core sample. Thirdly, if a sample is density reachable from multiple primary core samples, the core samples falling in its ε-neighborhood are placed in the candidate set. Then, we count the number of samples that are density reachable from each candidate core sample, and the one with most density-reachable samples will be matched with the selected sample. The bus station of that sample will be the same as the corresponding bus station of that core sample.

Step 3: Obtaining the bus arrival time

The primary core samples identified by the DBSCAN algorithm and the corresponding bus stations can be used to generate the bus timetable. The time recorded by the primary core samples is the arrival time at those bus stations. Based on the algorithm above, the matching of the bus GPS data and bus station data in the spatial dimension can be realized, and an accurate bus timetable, which is fundamental to bus OD estimation and records the exact time when each vehicle arrives at each bus station, can be generated.

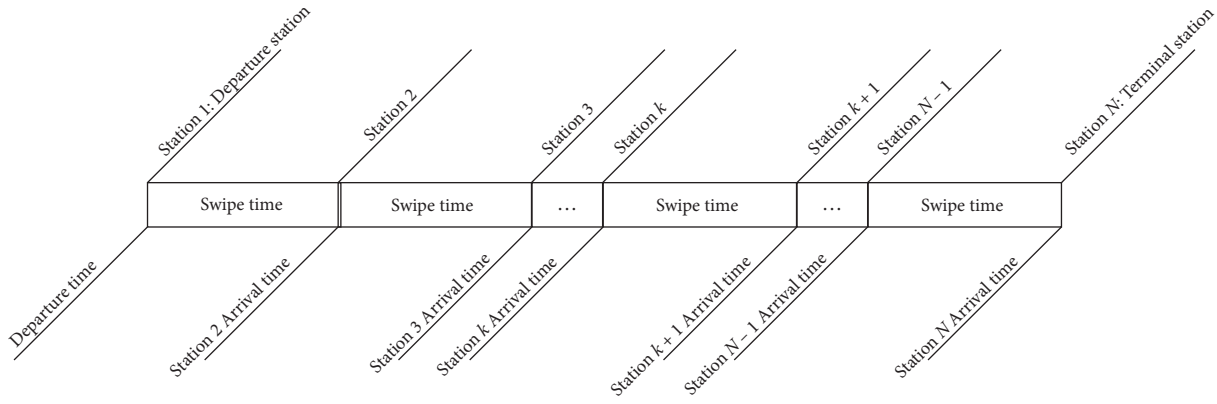


Figure 1: Relationship of card swiping time and bus arrival time.

3.2. Boarding Station Identification Based on Temporal Matching. Through the effective matching of the bus GPS data and the bus station data, the bus arrival time has been obtained. Subsequently, the boarding station can be identified by building connections between the smart card data and the bus arrival time in the temporal dimension.

3.2.1. Temporal Matching Algorithm. The premise of applying the above matching idea is to align the recorded time of the smart card data with the bus arrival time. Since the card swiping system and the GPS terminal installed on the bus operate independently, a fixed time difference between the recorded time values, referred to as the system time difference, might exist. Therefore, before identifying the boarding stations, the data from the two sources should be adjusted to ensure that they are consistent in the temporal dimension.

Generally, the time recorded in the smart card data is merely the transaction time rather than the true boarding time. As shown in Figure 1, the swiping behavior of bus passengers regularly occurs during the time the bus takes from the boarding station to the next station.

The system time difference between the smart card data and the bus GPS data will lead to the following two situations:

(1) The system time of the smart card data is earlier than that of the bus GPS data system. In such case, it is possible that the boarding station is mistakenly identified as the previous boarding station, resulting in a lower number of boardings at the current station and a higher number of boardings at the previous station.

(2) The system time of the smart card data system time is later than that of the bus GPS data system. In this case, it is possible that the boarding station is mistakenly identified as the next station, resulting in a higher number of boardings at the current station and a lower number of boardings at the previous station.

Assuming the system time difference between smart card system and bus GPS system is Δt, and this difference is constant for the two data sources. The system time difference Δt can be expressed as

$$\Delta t = \operatorname{sign}\left(\Delta t^{(i,d)}\right) \times \max_{i,d} \left|\Delta t^{(i,d)}\right|, \tag{2}$$

$$\Delta t^{(i,d)} = t_{\mathrm{IC}}^{(i,d)} - t_{\mathrm{GPS}}^{(i,d)}, \tag{3}$$

where $\Delta t^{(i,d)}$ and $\Delta t^{(j,d)}$ are subject to

$$\Delta t^{(i,d)} \times \Delta t^{(j,d)} > 0, \quad \forall \Delta t^{(i,d)}, \Delta t^{(j,d)} \in T, \tag{4}$$

$$\Delta t^{(i,d)} < \Delta T, \tag{5}$$

where $\Delta t^{(i,d)}$ represents the system time difference of i-th bus routes on day, $t_{\mathrm{IC}}^{(i,d)}$ represents the time of the first transaction of the i-th bus routes on day d, $t_{\mathrm{GPS}}^{(i,d)}$ represents the time of the first bus arrival of the i-th bus routes on day d, T represents the collection of $t_{\mathrm{IC}}^{(i,d)}$, ΔT represents the threshold of the maximum system time difference, and $\operatorname{sign}(\cdot)$ indicates whether the value is positive or negative.

Equation (2) takes the maximum time difference between the card swiping time of the first bus on that day and the first bus arrival time as the system time difference. Constraint (4) guarantees that all calculated time difference values take the same sign. Constraint (5) ensures that the first transaction recorded in the smart card data corresponds to the boarding at the first station of that bus route.

In practice, a minor lag might exist between the transaction time of the first passenger and the bus arrival time at the terminus. This can be addressed by simply adding a lag coefficient $\delta^{(i,d)}$ to equation (3), formulated as

$$\Delta t^{(i,d)} = t_{\mathrm{IC}}^{(i,d)} - t_{\mathrm{GPS}}^{(i,d)} - \delta^{(i,d)}, \tag{6}$$

where $\delta^{(i,d)}$ represents a lag coefficient satisfying a specific probability distribution, which can be calibrated according to the actual condition. Finally, using the system correction algorithm, the time of the smart card data can be revised as

$$t_{\mathrm{IC}}^{(i,d)} = t_{\mathrm{IC}}^{(i,d)} - \Delta t, \tag{7}$$

where $t_{\mathrm{IC}}^{(i,d)}$ represents the card swiping time on day d of bus line i after correction.

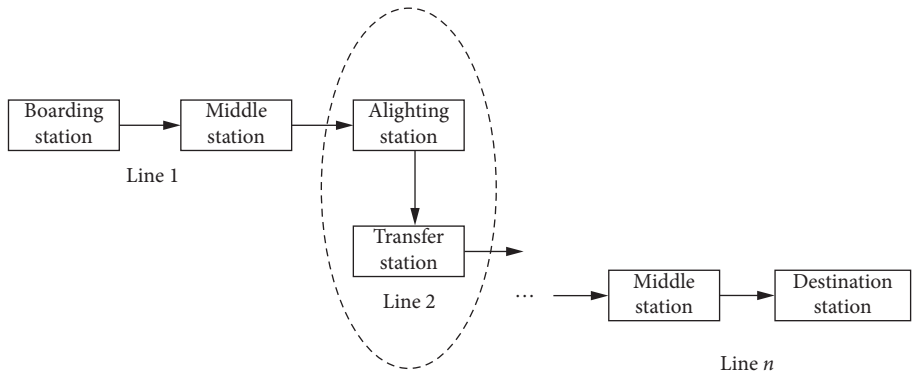


Figure 2: Typical bus trip chain.

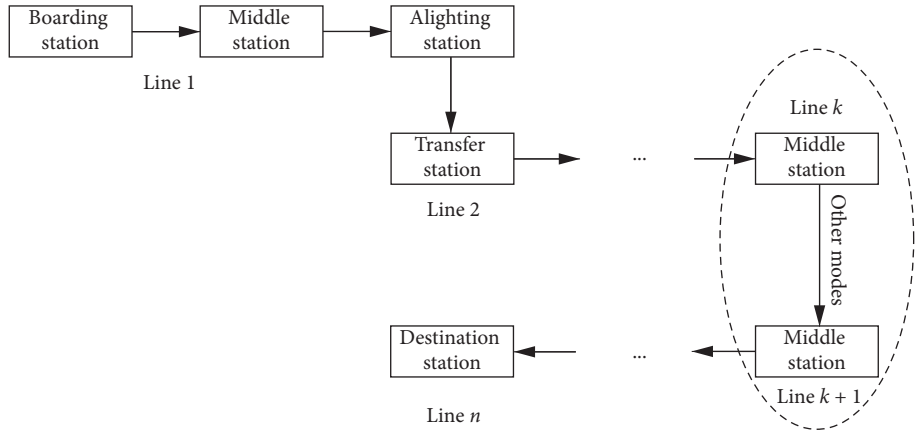


Figure 3: Unclosed bus trip chain.

3.2.2. Framework of Boarding Station Recognition. After correcting the system time difference based on the characteristics of multisource data, the overall framework of the boarding station recognition algorithm proposed in this paper can be summarized in two steps.

Step 1: System time correction

The transaction time of the first passenger recorded in the smart card data and the first bus arrival data each day for all bus lines are extracted for system time correction. According to the system time correction algorithm introduced in the previous section, the corrected smart card data can be obtained by Constraint (5).

Step 2: Boarding station identification

The boarding stations are identified based on the corrected smart card data and bus arrival timetable data according to the condition that the card swiping behavior happens in the time interval between two stations. In other words, the transaction time should be later than the bus arrival time at a station but earlier than the bus arrival time of the next station, where the first station here is regarded as the boarding station of that passenger.

4. Estimation Method of the Alighting Station Based on Trip Chaining

4.1. Definition of Bus Trip Chaining. Relevant research indicates that single trip of urban residents is the basic unit of the trip chaining. Bus trip chaining can be further defined as a process where residents take bus travel and form at least one spatial connection to neighboring travel. A bus trip chain can either be closed or unclosed. As shown in Figure 2, a typical and complete bus trip chain consists of the origin station, bus lines, middle stations, transfer stations, transfer bus lines, and the destination station. It has the following two features: (1) the passenger's alighting station of his/her non-last bus travel is spatially connected to the boarding station of his next bus travel on the same day and (2) the passenger's alighting station of his/her latest travel is spatially connected to the boarding station of his first travel.

However, in a multimodal transit system, passengers intend to use a combined travel pattern including different travel modes (e.g., bus, metro, taxi). Hence, the bus trip chain of a passenger is sometimes unclosed. As shown in Figure 3, the "alighting station" and "transfer station" in the cycle may not have clear spatial connection within the same

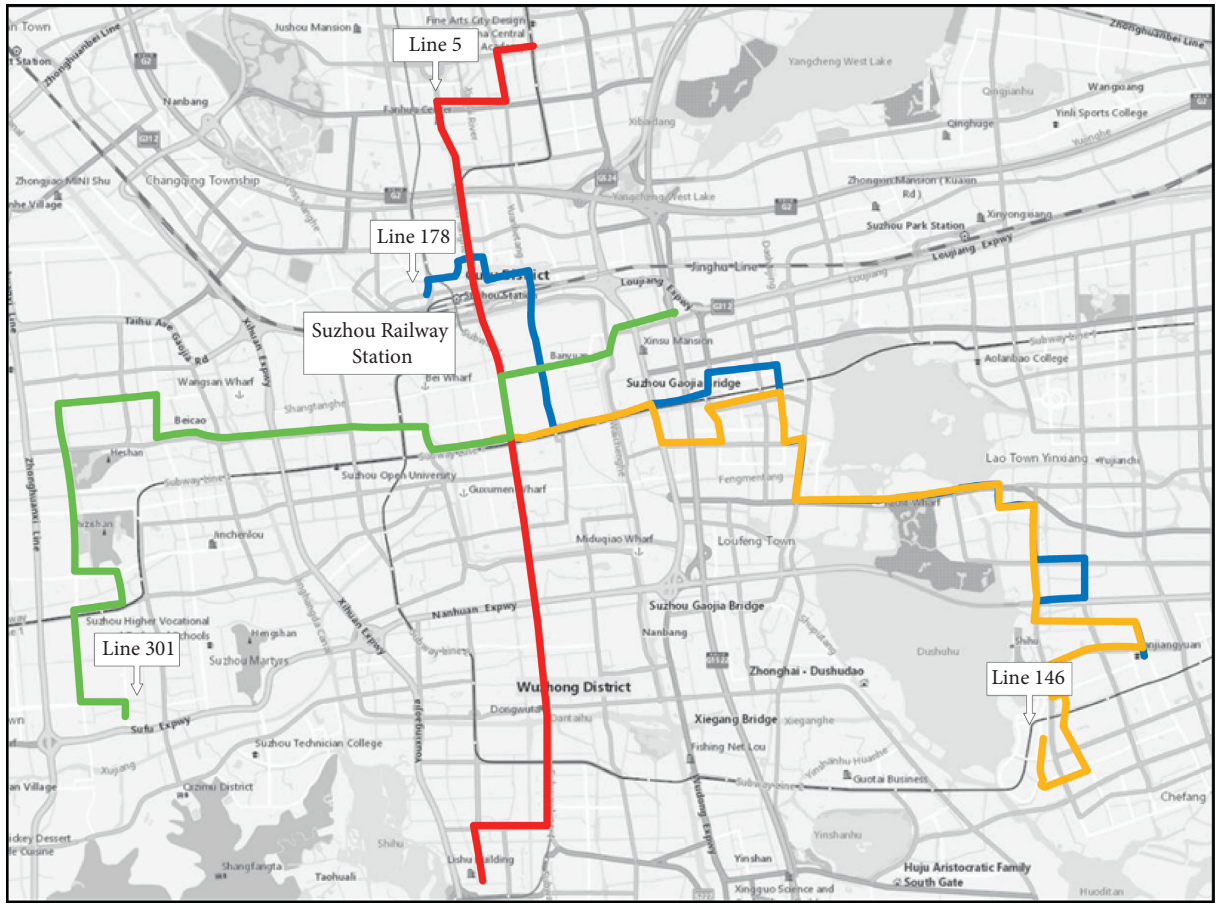


Figure 4: Layout of selected bus lines.

Table 4: Description of selected bus lines.

No. of bus line	Description	Direction
5	Main commuter line of downtown	S-N
301	Main commuter line of downtown	E-W
178	Connecting line between railway station and suburb	E-W
146	Main commuter line of suburb	S-N

day, which indicates that the passenger may choose other travel modes between these two stations. In this regard, the unclosed trip chain can be restored by the passenger's historical trip data to distinguish his/her regular or occasional trip.

In a bus trip chain, a typical passenger picks up the bus line L_1 at bus station S_1^1 and gets off at station S_1^2. Then, he/she transfers to bus line L_2 via station S_1^2 and gets off at station S_2^2. After n times of transfer, he/she arrives at the destination station S_{n+1}^2. The distance between the k-th alighting station S_k^2 and the $k+1$-th alighting station S_{k+1}^1 should be shorter than the acceptable walking distance. In summary, bus trip chaining should meet the following constraints: (1) component constraint: a bus trip chain must consist of at least two complete bus travels; (2) temporal constraint: the boarding time in last trip must be earlier than the boarding time in next trip; (3) spatial constraint: the distance between the alighting station in last trip and the boarding station in next trip should be shorter than the typically acceptable walking distance (e.g., 500 –1500 meters); and (4) transfer constraint: the transfer times in a bus trip chain should be lower than the acceptable frequency.

4.2. Alighting Stations Estimation Algorithm. On the one hand, it is difficult to estimate those unchained trips due to their irregularity and unpredictability. On the other hand, bus network planning usually focuses on the main travel behavior such as commuting. Therefore, estimating the alighting station for the chained trips can not only extract the spatiotemporal characteristics and OD matrix of most passenger's bus travel, but also meet the needs of bus network planning and design. By stripping and reforming the above dataset, the closed bus trip chain dataset and the unbroken part of unclosed bus trip chain dataset are integrated to form the dataset of chained trips. Meanwhile, the broken part of the unclosed bus trip chain data and the remaining data are added to the dataset of unchained trips.

4.2.1. Alighting Stations Estimation of Non-Last Bus Travel. According to the result of boarding station identification, it is assumed that a passenger picks up on the $j^{(L_i)}$ station of

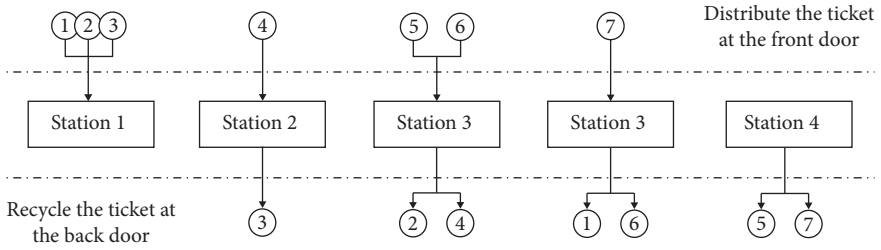


Figure 5: Illustration of the ticket recycling survey method for a single bus line.

(a) (b)

Figure 6: Distributing and recycling of the ticket at the front and back doors, respectively.

bus line L_i and his/her next boarding station is $j^{(L_i+1)}$ station of bus line L_{i+1}. Based on the spatial constraint and minimum station distance assumption, passenger would choose the alighting station which is the nearest to the next boarding station. Thus, the station which is the nearest to the $j^{(L_i+1)}$ station of line L_{i+1} from the stations in the downstream of the $j^{(L_i)}$ station online L_i is chosen as the alighting station of the last travel.

To ensure the effectiveness of data processing, the selection of the stations on bus line L_i where its distance from the $j^{(L_i+1)}$ station on bus line L_i is shorter than the acceptable walking distance, is recommended. A table storing such transfer topology information is calculated in advance for subsequent selection and processing.

4.2.2. Alighting Stations Estimation of Last Bus Travel. According to the result of boarding station identification, it is assumed that a passenger picks up on the $j^{(L_n)}$ station of bus line L_n in his/her last travel of the day, and his/her boarding station in first travel of the day is $j^{(L_1)}$ station of bus line L_1. Based on the minimum station distance assumption and commuting rules, passenger would choose the alighting station which is the nearest to the first boarding station on that day. Thus, the station which is the nearest to the $j^{(L_1)}$ station of line L_1 from the stations in the downstream of the $k^{(L_n)}$ station is chosen as the alighting station of the last travel of the day.

Table 5: Statistical description of the MAE.

Statistics	Boarding station identification Value	Alighting station identification Value
Average	1.013	1.206
Standard	0.443	0.496
Minimum	0.140	0.264
25th percentile	0.669	0.834
Median	1.007	1.144
75th percentile	1.265	1.502
Maximum	2.256	2.550

4.2.3. Alighting Stations Estimation of Unclosed Bus Trip Chain. Unclosed bus trip chain can be split into the broken and unbroken parts. If the broken part occurs in the middle of the chain, the broken parts are usually recovered by the historical data while the remaining unbroken part can be regarded as a continuous bus trip chain. If the broken chain is in the head and tail of the chain, the alighting stations of non-last bus travel can be estimated in the same way as that used in a closed bus trip chain. Instead of the recovery method based on historical data, another method named chaining extension which tries to establish a spatial connection between the neighboring bus travels in neighboring days is preferred due to its higher confidence.

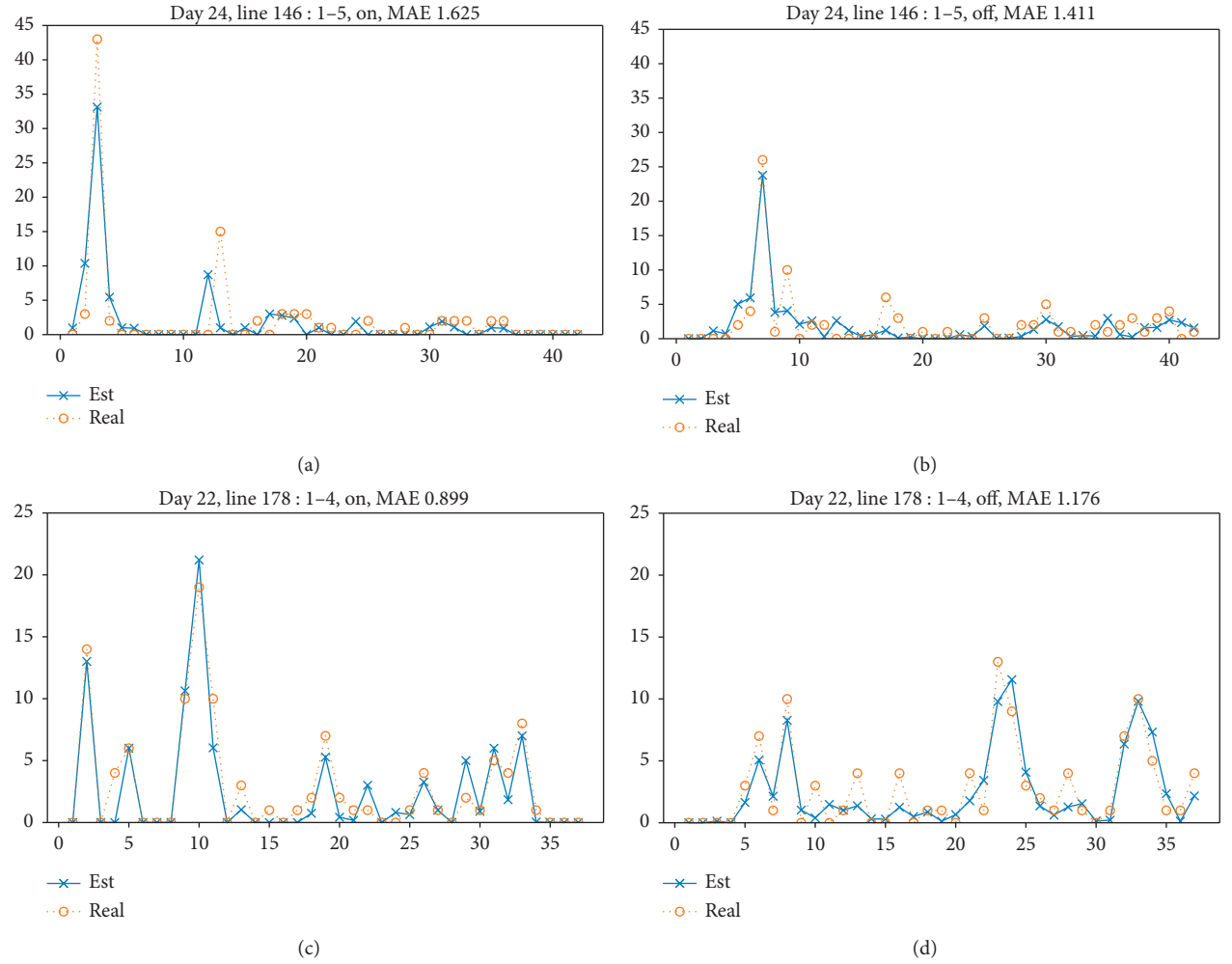


Figure 7: True values and estimated values of boarding and alighting numbers. (a) Line146: boarding passengers. (b) Line146: alighting passengers. (c) Line178: boarding passengers. (d) Line178: alighting passengers.

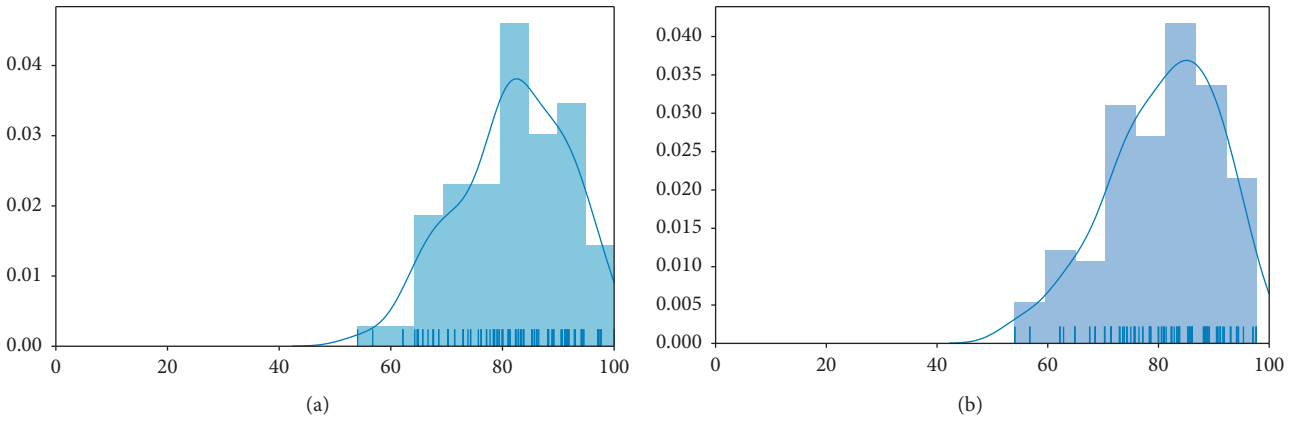


Figure 8: Percentage of stations where numbers of boarding and alighting passengers are accurately calculated. (a) Boarding station identification. (b) Alighting station estimation.

After reasonably dividing the bus datasets that have been effectively identified on the boarding station, we use the proposed algorithm based on the bus trip chain to estimate the alighting station and obtain the OD matrix.

5. Experiment and Verification

In order to verify the effectiveness of the proposed methodology, a survey was conducted in Suzhou, China, to collect the real bus travel OD information. As shown in Figure 4 four typical bus lines are selected, which have different functions (see Table 4). The ticket recycling survey method is used to collect the real number of passengers from station to station.

5.1. Ticket Recycling Survey Method. The ticket recycling survey method is a classical flow-up survey which has been widely applied to obtain the real passenger distribution along a bus route [40]. As shown in Figures 5 and 6, each passenger would receive a ticket that records the number of the station where s/he boards the bus at the front door. This ticket will be collected before the passenger alight from the back door. Though this method is time consuming and costly, it has a 100% sample rate which could obtain the real station-to-station trip information. The survey contains 72 round trips of these four lines. The total number of valid tickets is 10,551.

5.2. Accuracy Evaluation. By counting the number of board and alight passengers surveyed, the mean absolute error (MAE)of each class (i.e., from the origin station to the destination station) can be calculated. As a result, the MAE on each class is 1.326 passengers/station for numbers of boarding passengers and 1.299 passengers/station for numbers of alighting passengers. The detained statistical description of the MAE is presented in Table 5.

Figure 7 shows the verification results of bus line 146 (departure time: 8:25 am, October 24) and bus line 178 (departure time: 7:45 am, October 22). In order to find out whether the algorithm is accurate for most stations, this paper also proposes another checking indicator by calculating the percentage of the number of stations whose error is less than a certain threshold (number of passengers) to the total number of stations. It is assumed that the result is considered to be accurate when the estimated value differs from the actual value by two or less. As shown in Figure 8, the average percentage of the number of boarding stations is accurately identified as 81.8%, and this indicator is 80.9% for the alighting station estimation.

Furthermore, consider that there is a certain systematic error between the sampling time of the card reader and the POS time which is difficult to obtain. Table 6 lists the accurate inference percentage of each bus line and adds the limited dynamic time warping (LDTW). LDTW is usually used to evaluate time series similarity. In this study, the constraint of deformation distance is added on the standard dynamic time deformation. This means the value a_t in the time series can only match the adjacent time interval a_{t-1} and a_{t+1}. This indicator has better robustness than an absolute error by considering the relevance of the attraction adjacent stations to passenger flow.

Table 6: Estimation accuracy for boarding and departing passengers.

Line	Boarding station identification		Alighting station estimation	
	Accuracy (%)	LDTW	Accuracy (%)	LDTW
5	82.0	9.171	82.2	9.508
146	85.3	9.815	86.7	8.916
178	85.4	12.475	81.6	12.150
301	71.7	14.031	70.0	14.898

Another practical significance of using LDTW for evaluation is that it can prove the prediction accuracy of the algorithm for the overall trend. LDTW can be more accurate to estimate a small area instead of one station due to the fact that the adjacent stations are usually close enough to be considered as a small district. The verification results show that the accuracy of line 5 and line 146 is above 80%, and the corresponding LDTW is less than 10. The performance on line 301 is comparatively worse, but the accurate percentage is still above 70%, and the LDTW is less than 15. Tables 7 and 8 present a detailed comparison between the estimated and observed results for 10 stations with the highest passenger volume. It shows that the estimation values of the boarding passengers are almost less than 10 passengers, while the alighting passengers are less than 15 passengers.

5.3. Accuracy Evaluation of OD Matrix Estimation. By combining the results of boarding and alighting station estimation, the accuracy of OD matrix in each bus class can be calculated (the statistics do not include the OD with estimated and actual value being 0). The mean absolute error of the OD estimation is 0.581. Figure 9 shows the distribution of OD estimation accuracy. The accuracy threshold of Figure 9(a) is 2 passengers difference between the inferred value of the OD and the actual value, and the average accuracy is 94.3%. Figure 9(b) shows the accuracy threshold to 1 person, and the average accuracy is 72.8%. Table 9 lists the MAE of OD estimation in different bus lines. It can be seen from the table that the accurate percentage of each line (the error is less than 2 passengers) is above 90%, while the mean absolute error (MAE) is basically below 0.7. By comparing the estimated value with the real data of the bus travel survey, it can be found that the proposed algorithm performs well both in the identification of boarding and alighting station and in the overall OD matrix estimation. It can effectively obtain the passenger flow of each station and OD of different bus lines.

Table 7: Comparison between the estimated and observed values of the boarding passengers.

Date	No. of line	No. of bus	No. of station	Est. value	Obs. value	Error
24/Oct	146	5	3	33.125	43	9.875
23/Oct	178	2	2	28.817	39	10.183
24/Oct	178	3	2	44.720	38	6.720
23/Oct	146	1	14	32.657	34	1.343
25/Oct	5	9	4	22.418	33	10.582
23/Oct	178	3	2	35.516	32	3.516
22/Oct	146	2	14	24.355	30	5.645
22/Oct	146	0	14	19.633	24	4.367
24/Oct	146	1	14	19.281	24	4.719
23/Oct	178	2	5	18.448	23	4.552

Table 8: Comparison between the estimated and observed values of the alighting passengers.

Date	No. of line	No. of bus	No. of station	Est. value	Obs. value	Error
24/Oct	146	1	16	31.341	36	4.659
23/Oct	178	2	6	19.012	29	9.988
24/Oct	146	5	7	23.772	26	2.228
23/Oct	178	3	8	11.125	26	14.875
25/Oct	146	1	16	19.298	24	4.702
23/Oct	146	0	16	8.140	24	15.860
22/Oct	178	4	6	15.254	24	8.746
22/Oct	5	9	9	10.822	24	13.178
24/Oct	301	0	36	7.352	23	15.648
23/Oct	5	2	29	11.432	23	11.568

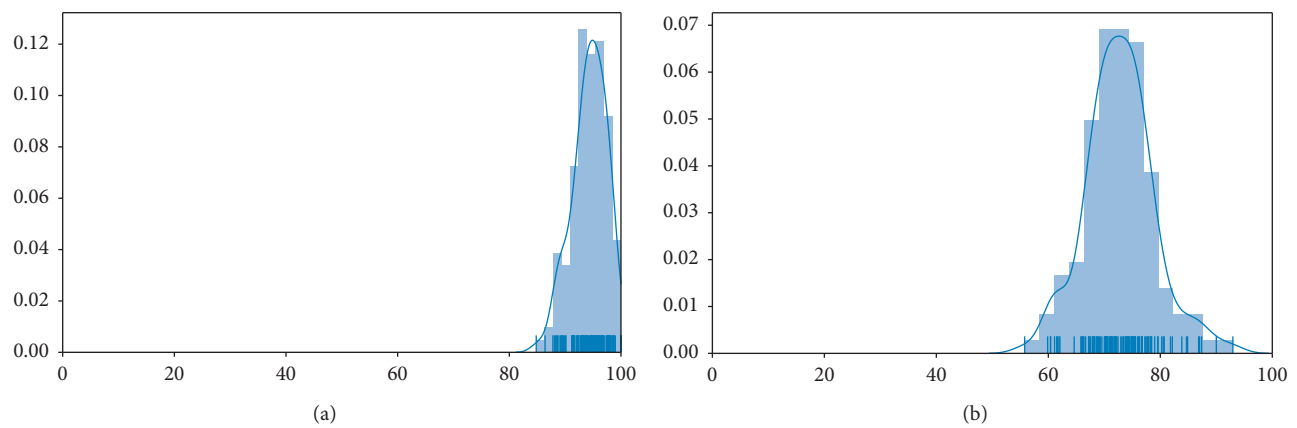


Figure 9: Percentage of accurate OD estimation. (a) Error less than 2 passengers. (b) Error less than 1 passenger.

Table 9: Descriptive statistics of OD matrix.

No. of line	Rate of accurate OD (error < 2) (%)	MAE
5	94.9	0.543
146	95.2	0.554
178	94.3	0.626
301	92.6	0.630

6. Conclusions

The accurate estimation of bus OD matrix is essential for the planning and management of urban bus system. This paper proposes a framework for estimating the OD matrix, including a boarding station identification algorithm and an alighting station inference algorithm. The boarding station identification algorithm allows for the mismatch of system time between the smart card system and the bus GPS system. These two datasets are aligned in the temporal dimension through the density-based clustering algorithm. Then, based on the identified chained and unchained trips, an alighting station inference algorithm based on trip chaining is designed. Above 80% of alighting stations could be identified by using the proposed estimation algorithm, which increases the identification rate by 10% compared with previous studies [33, 36, 38]. Finally, the accuracy of the proposed algorithm is evaluated on the data collected by the field survey in Suzhou, China. The results show that the proposed methodology could obtain an accuracy level above 90%.

Also, this algorithm can be further improved in the following aspects. First, the proposed framework works only

within the bus network. However, with the development of multimodal urban transport system, travelers' trip chains can include other travel modes such as light rail, subways, and even shared bicycles. Second, more data sources can be introduced to compensate for the incomplete market share of smart cards.

Data Availability

The data used to support the findings of this study have not been made available because of confidential issues.

Conflicts of Interest

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

Acknowledgments

This study was supported by the National Key Research and Development Program of China (No. 2018YFB1600900), the Key Project (No. 51638004) of the National Natural Science Foundation of China, the Scientific Research Foundation of Graduate School of Southeast University (No. YBPY1835), and DiDi Gaia Research Collaboration Initiative. The authors would like to thank Cheng Lyu and Yunyang Shi for their efforts in data preprocessing.

References

[1] Y. Liu, Z. Liu, and R. Jia, "DeepPF: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.

[2] E. Chen, Z. Ye, C. Wang, and M. Xu, "Subway passenger flow prediction for special events using smart card sata," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2019.

[3] D. Huang, Z. Liu, X. Fu, and P. T. Blythe, "Multimodal transit network design in a hub-and-spoke network framework," *Transportmetrica A: Transport Science*, vol. 14, no. 8, pp. 706–735, 2018.

[4] D. Huang, Z. Liu, P. Liu, and J. Chen, "Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure," *Transportation Research Part E: Logistics and Transportation Review*, vol. 96, pp. 1–19, 2016.

[5] Z. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 58–72, 2017.

[6] Y. Bie, X. Xiong, Y. Yan, and X. Qu, "Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 4–25, 2020.

[7] X. Fu and W. H. K. Lam, "Modelling joint activity-travel pattern scheduling problem in multi-modal transit networks," *Transportation*, vol. 45, no. 1, pp. 23–49, 2018.

[8] C. Wang, Z. Ye, E. Chen, M. Xu, and W. Wang, "Diffusion approximation for exploring the correlation between failure rate and bus-stop operation," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1306–1320, 2019.

[9] Y. Pan, S. Chen, F. Qiao, S. V. Ukkusuri, and K. Tang, "Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees," *Science of the Total Environment*, vol. 660, pp. 741–750, 2019.

[10] S. Tao, D. Rohde, and J. Corcoran, "Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap," *Journal of Transport Geography*, vol. 41, pp. 21–36, 2014.

[11] Y. Ji, Y. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "Public bicycle as a feeder mode to rail transit in China: the role of gender, age, income, trip purpose, and bicycle theft experience," *International Journal of Sustainable Transportation*, vol. 11, no. 4, pp. 308–317, 2017.

[12] M. Du and L. Cheng, "Better understanding the characteristics and influential factors of different travel patterns in free-floating bike sharing: evidence from Nanjing, China," *Sustainability*, vol. 10, no. 4, p. 1244, 2018.

[13] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and destination estimation in New York City with automated fare system data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1817, no. 1, pp. 183–187, 2002.

[14] Y. Yuan, M. Yang, J. Wu, S. Rasouli, and D. Lei, "Assessing bus transit service from the perspective of elderly passengers in Harbin, China," *International Journal of Sustainable Transportation*, vol. 13, no. 10, pp. 761–776, 2019.

[15] M. Utsunomiya, J. Attanucci, and N. Wilson, "Potential uses of transit smart card registration and transaction data to improve transit planning," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1971, no. 1, pp. 118–126, 2006.

[16] L. Sun, A. Tirachini, K. W. Axhausen, A. Erath, and D.-H. Lee, "Models of bus boarding and alighting dynamics," *Transportation Research Part A: Policy and Practice*, vol. 69, pp. 447–460, 2014.

[17] Y. Ji, X. Ma, M. Yang, Y. Jin, and L. Gao, "Exploring spatially varying influences on metro-bikeshare transfer: a geographically weighted Poisson regression approach," *Sustainability*, vol. 10, no. 5, p. 1526, 2018.

[18] T. Spurr, R. Chapleau, and D. Piché, "Use of subway smart card transactions for the discovery and partial correction of travel survey bias," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2405, no. 1, pp. 57–67, 2014.

[19] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: a literature review," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 557–568, 2011.

[20] A. Kiyohiro, K. Yamaguchi, H. Gao, H. Nakamura, and T. Mine, "Customer behavior analysis on after getting off the train based on usage histories of smart IC card," in *Proceedings of the 3rd International Conference on Advanced Applied Informatics*, Kitakyushu, Japan, August 2014.

[21] Q. Ouyang, Y. Lv, Y. Ren, J. Ma, and J. Li, "Passenger travel regularity analysis based on a large scale smart card data," *Journal of Advanced Transportation*, vol. 2018, Article ID 9457486, 11 pages, 2018.

[22] N. Ebadi, J. E. Kang, and S. Hasan, "Constructing activity-mobility trajectories of college students based on smart card transaction data," *International Journal of Transportation Science and Technology*, vol. 6, no. 4, pp. 316–329, 2017.

[23] Y. Ji, J. Zhao, Z. Zhang, and Y. Du, "Estimating bus loads and OD flows using location-stamped farebox and Wi-Fi signal

data," *Journal of Advanced Transportation*, vol. 2017, Article ID 6374858, 10 pages, 2017.

[24] X.-L. Ma, Y.-H. Wang, F. Chen, and J.-F. Liu, "Transit smart card data mining for passenger origin information extraction," *Journal of Zhejiang University Science C*, vol. 13, no. 10, pp. 750–760, 2012.

[25] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.

[26] T. Kusakabe and Y. Asakura, "Behavioural data mining of transit smart card data: a data fusion approach," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 179–191, 2014.

[27] D. Chen, "Research on traffic flow prediction in the big data environment based on the improved RBF neural network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2000–2008, 2017.

[28] J. Bao, P. Liu, X. Qin, and H. Zhou, "Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data," *Accident Analysis & Prevention*, vol. 120, pp. 281–294, 2018.

[29] L. Li, J. Zhang, Y. Wang, and B. Ran, "Missing value imputation for traffic-related time series data based on a multi-view learning method," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 8, pp. 2933–2943, 2019.

[30] A. A. Nunes, T. G. Dias, and J. F. e Cunha, "Passenger journey destination estimation from automated fare collection system data using spatial validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 133–142, 2016.

[31] D. Zhang, X. Zhang, and J. Wang, "Commuter travel identification based on bus IC data," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 1547–1555, 2013.

[32] J. M. Farzin, "Constructing an automated bus origin-destination matrix using farecard and global positioning system data in São Paulo, Brazil," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2072, no. 1, pp. 30–37, 2008.

[33] M. Trépanier, N. Tranchant, and R. Chapleau, "Individual trip destination estimation in a transit smart card automated fare collection system," *Journal of Intelligent Transportation Systems*, vol. 11, no. 1, pp. 1–14, 2007.

[34] N. Nassir, A. Khani, S. G. Lee, H. Noh, and M. Hickman, "Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2263, no. 1, pp. 140–150, 2011.

[35] W. Wang, J. Attanucci, and N. Wilson, "Bus passenger origin-destination estimation and related analyses using automated data collection systems," *Journal of Public Transportation*, vol. 14, no. 4, pp. 131–150, 2011.

[36] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.

[37] C. Seaborn, J. Attanucci, and N. H. M. Wilson, "Analyzing multimodal public transport journeys in London with smart card fare payment data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2121, no. 1, pp. 55–62, 2009.

[38] A. Cui, *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*, Massachusetts Institute of Technology, Cambridge, MA, USA, 2006.

[39] K. Lu, A. Khani, and B. Han, "A trip purpose-based data-driven alighting station choice model using transit smart sard data," *Complexity*, vol. 2018, Article ID 3412070, 14 pages, 2018.

[40] M. An, X. Chen, and Z. Li, "Transit OD matrix estimation based on trickey survey method," *Journal of Transportation System Engineering and Information Technology*, vol. 10, no. 1, pp. 170–176, 2009.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 6517248, 12 pages
https://doi.org/10.1155/2020/6517248

Research Article

Optimal Integrated Model for Feeder Transit Route Design and Frequency-Setting Problem with Stop Selection

Ming Wei,[1] Tao Liu,[2] Bo Sun ,[1] and Binbin Jing [3]

[1]*School of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China*
[2]*National Engineering Laboratory of Integrated Transportation Big Data Application Technology, School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China*
[3]*School of Transportation, Nantong University, Nantong 226019, China*

Correspondence should be addressed to Bo Sun; bosun@cauc.edu.cn

Received 22 December 2019; Accepted 28 January 2020; Published 9 March 2020

Academic Editor: Eneko Osaba

This study proposed a mathematical model for designing a feeder transit service for improving the service quality and accessibility of transportation hubs (such as airport and rail station). The proposed model featured an integrated framework, which simultaneously guided passengers to reach their nearest stops to get on and off the bus, designed routes to transport passengers from these selected pick-up stops to the transportation hubs, and calculated their departure frequencies. In particular, the maximum walking distance, the upper and lower limits of route frequencies, and the load factor rate of each route were fully accounted for in this study. The main objective of the proposed model was to simultaneously minimize the total walking, riding time, and waiting time of all passengers. As this study explored an NP-hard problem, a two-stage genetic algorithm combining the Dijkstra search method was further developed to yield metaoptimal solutions to the model within an acceptable time. Finally, a test instance in Chongqing City, China, demonstrated that the proposed model was an effective tool to generate a pedestrian, route, and operation plan; it reduced the total travel time, compared with the traditional model.

1. Introduction

Feeder transit system (FTS) aims at arranging access to vehicles located at different depots at all demand points and transporting residents from these selected pick-up stops to transportation hubs (rail station and airport, etc.). The FTS transit network, including a set of nodes (transportation hubs, bus stops, demand points, and depots) and links between them, is regarded as an effective tool to provide a better first/last mile service to and from the major fixed-route transit networks. Hence, a well-designed FTS transit network shifted transport demand from individual car traffic to public transport and further enhanced urban sustainability [1–3].

Similar to conventional bus operation, route design and frequency setting are two key activities in FTS operation, where the former is the input data of the latter, and they affect each other. Several studies have analyzed the impact of frequencies under different route structures to meet the same passenger demand. Frequencies were increased to improve the level of service under hub-and-spoke route structures, while they were reduced in fully connected route structures. However, the majority of FTS have neglected an integrated operation of designing transit route structures and setting their frequencies [1, 4, 5]. Therefore, it is especially important to optimize a combination of FTS route design and frequency setting to reveal an optimal relationship between in-vehicle congestion and waiting time at bus stops.

Another important observation was that most studies made an assumption of locations of demand points being pick-up stops and all stops being visited. In the real world, passengers at the demand point may walk to one of the nearby stops to get on and off the bus, and their choice behaviors are decided by routes with empty seats and shorter driving time. Selecting optimal stops in candidate routes as pick-up

locations with an assignment of demand points is also key to designing FTS routes [6, 7]. Therefore, the integrated operation of transit route design and frequency setting with stop selection is now widely regarded as an effective tool to improve service efficiency and financial status.

The main purpose of this study was to develop an integrated optimization framework for simultaneously coordinating the passenger boarding guidance, bus stop selection, feeder transit routing and frequency-setting process to trade off the total walking time from demand points to selected pick-up stops, total waiting time at selected pick-up stops, and total riding time from pick-up stops to the transportation hubs for all passengers. The study also focused on some research tasks as follows: (1) determination of the optimal design of FTS route and frequency with the assignment of all demand points to selected pick-up stops and (2) creation of a genetic algorithm- (GA-) based heuristic approach to efficiently find good solutions to such an NP-hard problem. Finally, a real-world test was performed to prove the applicability of the proposed model and algorithm.

The remainder of this study is organized as follows: Section 2 summarizes the related literature on FTS. Section 3 presents the methodology of the problem and provides a mathematical formulation for the integrated model. Section 4 presents a two-stage GA-based approach to resolve such a problem. Section 5 provides details of solutions and sensitivity analysis. Finally, Section 6 offers conclusions and future research directions.

2. Literature Review

At present, the FTS transit service has drawn great attention in the last few decades due to its convenience of first and last mile to link residential areas and transportation hubs in major fixed-route transit networks. Similar to the regular bus operation, route design and frequency setting are two key activities in improving the service level of operation cost, in-vehicle congestion, travel time, and waiting time at stops. Most previous studies optimized frequencies to generate a timetable based on an existing FTS route network. The integration of transit route structures and service frequencies could not maximize the operational efficiency of an FTS system. Hence, it is important to simultaneously coordinate the design process of these two activities in an FTS system to provide adequate feeder services from these origin places, where residents live or work, to transportation hubs (such as airport and rail station).

Most previous studies on FTS route design mainly included two categories of approaches [8]: continuous analytical and discrete optimization. In the analytical approaches, optimal relationships between route spacing, stop spacing, and frequency could be found with an idealized FTS network structure based on a uniformly distributed demand in the study area [9–11]. Further, these approaches also optimized the stop locations and the route structures, and their headways by considering the time-dependent [12], space-dependent [13], and time-and-space-dependent demands [14–17]. However, these analytical methods have significant limitations in practical applications of FTS route design [8], because the two key model inputs were idealized FTS network structures and the demand was uniformly distributed.

To overcome passenger demand endogeneity in a continuous analytical FTS, several discrete optimization approaches were considered to handle the FTS route design. First, this problem was decomposed into two sequential subproblems, which assigned stops to a route in the first stage and then determined the order of route by visiting the selected ones [18]. Second, a bilevel optimization model was built to achieve a user–operator equilibrium, where route structures were determined at first and the transit assignment was obtained [19]. Third, a nonlinear optimization model was built for passenger assignments and solved through heuristic algorithms [20]. However, some exact algorithms could only solve small-scale instances fast [18]. Advanced metaheuristics (e.g., GA) are commonly used to solve large-scale instances efficiently [2–4, 6, 7].

In the FTS frequency setting, the number of trips for each route was determined to analyze the effect of frequency changes on the level of service. Their research ideas and methods were the same as those of regular bus operations. Early studies used analytical models [12, 21, 22] to set frequency based on a fixed demand-line assignment. These studies were later extended with consideration of uncertainty factors such as demand and riding/boarding/alighting times [3, 23]. For example, Verbas and Mahmassani [24, 25] extended the uncertain frequency-setting model with space–time changed demand. Besides, the design of frequency-setting decisions in short-turn and limited-stop routes was also presented to meet the uneven demand [16, 26, 27].

In recent years, the integration of the regular bus route design and frequency setting has gained an ever-increasing interest. In early research studies, sequential approaches, bus route design considerations into the frequency setting, or vice versa did not guarantee a global optimal solution [28]. To avoid such disadvantages, the following approaches were used to handle the integration operation: (1) partial integrations, that is, the bus route design with consideration of frequency setting [29] and (2) complete integrations, that is, coordinated decisions of the two problems [30]. Complete integrated operations handle decisions of each subproblem and hence are much more complex than the subproblem [31]. Further, some exact and heuristic approaches were proposed to deal with large-scale problems [4, 5, 28–33].

A review of literature on FTS revealed the following critical issues, which deserve further investigations:

(1) Although several studies proposed various integration operations of regular bus route design and frequency setting, few of them took them into account in FTS. The difference between them was the route structures and demand patterns (i.e., the

transportation hub) [2, 4, 5, 28–31]. Therefore, this study was also very meaningful.

(2) Traditional FTS assumed that demand points were stops. People at demand points, where passengers live or work, can walk to one of the nearest stops to get on and off the bus. In such a case, an operation of integrated pedestrian guidance (from demand points to selected stops) and transit routing (from pick-up places to transportation hubs) can find an optimal relation between the walking and riding time of residents and the operation efficiency of vehicles [1, 6, 7, 34].

(3) Such integration of stop selection, transit routing, and frequency setting was a new research problem extending from an NP-hard vehicle route problem, and a novel heuristic algorithm was needed to efficiently resolve large-scale instances [8].

3. Methodology

3.1. Research Framework. In the FTS model, demand points, where residents live or work, are not bus stops. Some stops are selected in candidate routes as pick-up locations visited by transit routes starting at depots and ending at transportation hubs (such as airport and rail station). Passengers at different demand points can get on and off at the same stop by walking some distance. Passengers at a demand point can choose one of the several nearby bus stops to get on and off the bus. The assignment of each demand point to one of the nearest stops is determined by the route capacity and the total travel time of passengers. The main purpose of this study was to propose an integrated optimization model for coordinating bus stop selection, feeder transit routing, and frequency-setting process, where some stops with their assigned demand points are first chosen as pick-up locations in candidate routes by considering some constraints such as maximum walking distance in the process of stop selection. Then, the feeder transit route design aims at determining the sequence of these vehicle nodes (i.e., pick-up locations, bus depots, and transportation hubs) visited by buses by satisfying some route constraints such as capacity, travel time, and mileage range limits. In the frequency-setting process, it finally optimizes the number of departure trips per hour to balance passenger waiting time and vehicle load rate. The key inputs of the proposed model were locations of demand points, stop candidate routes, and transportation hubs and depots. This was supplemented by data about walking distance matrix between demand points and stops as well as travel distance and time matrix between these vehicle nodes in the study area using an open GIS tool (Google or Baidu map) to reflect the real traffic status. A novel mixed-integer linear programming model was developed to find the optimal coupling between them to guide passengers of demand points to be quickly and comfortably transported from selected pick-up stops to their associated transportation hubs. The organization chart of the proposed framework is shown in Figure 1.

A small FTS network consisting of one transportation hub (M), eight stop candidate routes (S1–S8), five demand points (C1–C5), and two bus depots (D1 and D2) was considered to better interpret the underlying principles of the proposed model and its scope, illustrated in Figure 2. In the figure, the number around the circle denotes the number of people at the demand point and the number near the square is the number of people getting on and off the bus at the bus stop. The optimization process yielded an assignment of five demand points to four selected stops, described as S4 (C5); similarly, for other selected stops, they were described as S3 (C4), S2 (C3), and S1 (C1, C2). Considering S1 (C1, C2) for instance, the number of people getting on and off the bus at S1 was the sum of that of C1 and C2. Two feeder routes were then yielded as follows: route 1 was illustrated by [D3–S4–S3–M] and route 2 by [D2–S2–S1–M]. Once the route design was completed, the number of departure trips per hour was determined using the maximum and minimum departure frequency and the full-load rate. Considering route 1 as an example, its departure frequency was obtained by dividing its passenger capacity (i.e., 90 per/h) by the rated passenger load of the vehicle (i.e., 15 per/veh), that is, $90/15 = 6$ veh/h.

The objective was to find the optimal design of the FTS route and frequency that simultaneously minimized the total walking time from demand points to selected stops, total waiting time at selected pick-up stops, and total travel time in vehicles of passengers from pick-up stops to transportation hubs for all passengers. The proposed methodology fitted well with real-life situations using the following assumptions:

(1) Each demand point must only be assigned to a selected stop, and each selected stop can only be covered once by one route.

(2) The real walking or travel distance/time matrix between these nodes in the traffic network is obtained from an Open GIS tool.

(3) By analyzing the signaling data of mobile phones, the number of passengers at each demand point in each period can be obtained. Also, it is assumed that the passengers arrive at the station evenly to get on and off the bus.

3.2. Model Formulation

3.2.1. Notation. To facilitate the problem presentation of the proposed model, mathematical symbols of related parameters and variables are summarized in Table 1.

3.2.2. Formulation.

$$\max f = \sum_{\forall i \in I} \sum_{\forall j \in N} h_{ij} \cdot \frac{d_{ij}}{v} \cdot q_i + \sum_{\forall j \in M} \sum_{\forall k \in K} y_j^k \cdot \sum_{\forall i \in I} h_{ij} \cdot q_i \cdot \left[\left(T_k - t_j^k \right) + \frac{60}{(f_k/2)} \right], \tag{1}$$

subjected to

$$h_{ij} \le z_j, \quad \forall i \in I, \forall j \in N, \tag{2}$$

$$\sum_{\forall j \in N} h_{ij} = 1, \quad \forall i \in I, \tag{3}$$

$$h_{ij} \cdot d_{ij} \le W, \quad \forall i \in I, \forall j \in N, \tag{4}$$

$$y_j^k \le z_j, \quad \forall k \in K, \forall j \in N, \tag{5}$$

$$\sum_{\forall j \in N} y_j^k \ge 1, \quad \forall k \in K, \tag{6}$$

$$\sum_{\forall j \in I} x_{jm}^k = 1, \quad \forall k \in K, \forall m \in M, \tag{7}$$

$$\sum_{\forall j \in N} x_{mj}^k = 0, \quad \forall k \in K, \forall m \in M, \tag{8}$$

$$\sum_{\forall j \in N} x_{jm}^k = 0, \quad \forall k \in K, \forall m \in D, \tag{9}$$

$$\sum_{\forall j \in N} x_{mj}^k = 1, \quad \forall k \in K, \forall m \in D, \tag{10}$$

$$\sum_{\forall m \in N \cup D \cup M} x_{jm}^k = \sum_{\forall m \in N \cup D \cup M} x_{mj}^k = y_j^k, \quad \forall k \in K, \forall j \in N, \tag{11}$$

$$U_{jk} - U_{mk} + |N \cup D \cup M| \cdot x_{jm}^k \ge |N \cup D \cup M| - 1, \forall k \in K, \forall j, m \in N \cup D \cup M, \tag{12}$$

$$t_j^k + t_{jm} - \left(1 - x_{jm}^k\right) H \le t_m^k, \quad \forall k \in K, \forall j, m \in N \cup D \cup M, \tag{13}$$

$$t_j^k + t_{jm} + \left(1 - x_{jm}^k\right) H \ge t_m^k, \quad \forall k \in K, \forall j, m \in N \cup D \cup M, \tag{14}$$

$$q_j^k + \sum_{\forall i \in I} h_{ij} \cdot q_i - \left(1 - x_{jm}^k\right) H \le q_m^k, \quad \forall k \in K, \forall j, m \in N \cup D \cup M, \tag{15}$$

$$q_j^k + \sum_{\forall i \in I} h_{ij} \cdot q_i + \left(1 - x_{jm}^k\right) H \ge q_m^k, \quad \forall k \in K, \forall j, m \in N \cup D \cup M, \tag{16}$$

$$D_k = \sum_{\forall j, m \in N \cup D \cup M} x_{jm}^k d_{jm} \le D_{\max}, \quad \forall k \in K, \tag{17}$$

$$T_k = \sum_{\forall j, m \in N \cup D \cup M} x_{jm}^k t_{jm} \ge T_{\min}, \quad \forall k \in K, \tag{18}$$

$$F_{Min} \le f_k \le F_{Max}, \quad \forall k \in K, \tag{19}$$

$$R_{Min} \le r_k = \frac{\sum_{\forall j \in N} q_j^k}{Q} \le R_{Max}, \quad \forall k \in K. \tag{20}$$

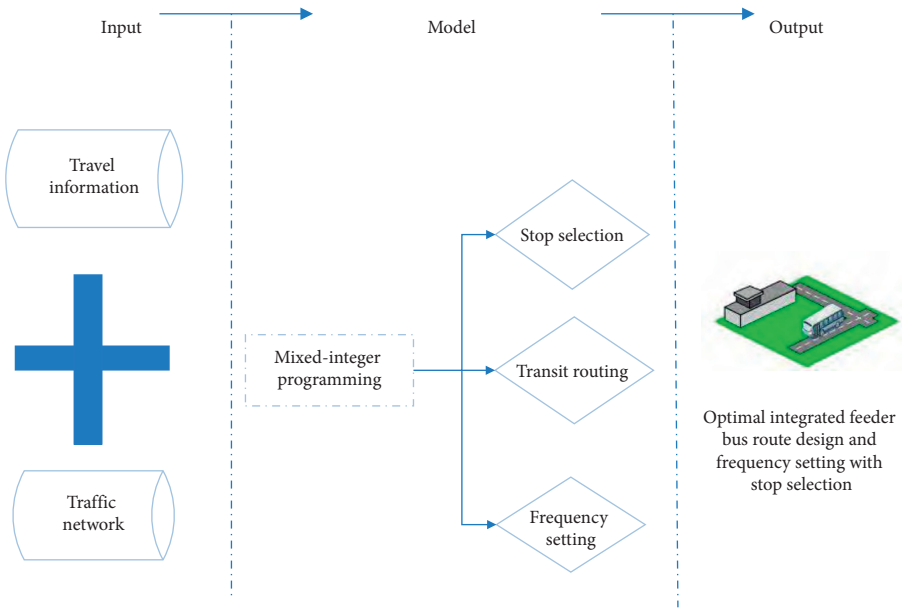


Figure 1: Organization chart of the proposed framework.

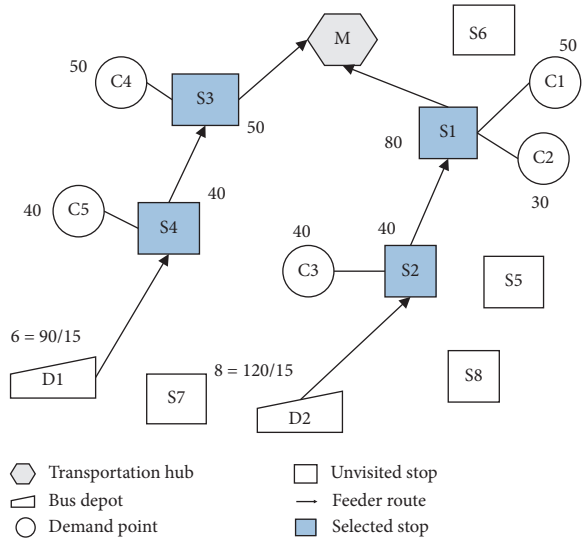


Figure 2: Typical example of the proposed model.

Objective (1) aims at minimizing the total walking, riding, and waiting time of all passengers. Constraint (2) ensures that a demand point must be allocated to a selected stop. Constraint (3) ensures that each demand point must be assigned to a pick-up location. Constraint (4) ensures that the walking distance from each demand point to the bus stop is not less than the maximum allowable value. Constraint (5) ensures that each selected bus stop is covered by feeder buses. Constraint (6) ensures that each vehicle visits at least one selected stop. Constraints (7)–(10) are used for each route eventually starting from the depot and ending at the transportation hub. Constraint (11) indicates that each bus stop cannot be visited by two routes at the same time. Constraint (12) indicates that subtour elimination is met. Constraints (13) and (14) calculate the arrival time of the vehicle k visiting the current stop. Constraints (15) and (16) calculate the load capacity of adjacent stops visited by each route. Constraints (17) and (18) guarantee that travel mileage and time of each route should meet their upper and lower limits. Constraints (19) and (20) indicate that the frequency setting and load factor of each route must meet their minimum and maximum values.

4. Solution Method

This model involves four core variables, including z_j, h_{ij}, x_{jm}^k, and f_k, where h_{ij} decides z_j. Obviously, demand point i can be assigned to route k according to h_{ij} and x_{jm}^k. Hence, a two-stage GA-based heuristic approach [7, 35, 36] was developed to solve this problem.

In the first stage, GA was used to allocate all demand points to different routes and calculate their departure frequencies according to the passenger flow of each route, where the total waiting time of passengers at stops in the objective function was determined.

Table 1: Mathematical symbols in the FTS model.

Index:	
i	Demand point
j, m	Vehicular node (depot, stop, and transportation hub) index
k	Route
Sets:	
I	Set of demand points
K	Set of routes
N	Set of candidate stops
D	Set of depots
M	Set of transportation hubs
Parameters:	
q_i	Number of persons at the demand points i during peak hours; $\forall i \in I$
Q	Rated vehicle capacity
$F_{\max}$	Maximum departure frequency
$F_{\min}$	Minimum departure frequency
r_k	Load rate of route k
$R_{\max}$	Maximum load rate
$R_{\min}$	Minimum load rate
$D_{\max}$	Maximum route length
$T_{\min}$	Minimum travel time of feeder bus route
D_k	The total mileage of the route k visiting all vehicular nodes; $\forall k \in K$
T_k	The total travel time of the route k visiting all vehicular nodes; $\forall k \in K$
W	Maximum walking distance
v	Average walking speed
d_{ij}	Map-based distance from rail stations, bus depots, and candidate route stop to demand point i, j; $\forall i$, $j \in I \cup M \cup D \cup N$
t_{ij}	Map-based distance from rail stations, bus depots, and candidate route stop to demand points i, j; $\forall i$, $j \in I \cup M \cup D \cup N$
H	Relatively large fixed value
Decision variables:	
z_j	Whether the candidate j is selected as a feeder bus stop; $\forall j \in N$
h_{ij}	Whether the demand point i is assigned to a selected stop j; $\forall i \in I$, $\forall j \in N$
x_{jm}^k	Whether a vehicular node j precedes vehicular node m on the route k; $\forall k \in K$, $\forall j \in N$
f_k	Frequency of route k during peak hour
y_j^k	Whether the node j is covered by the route k; $\forall j \in I \cup D \cup M \cup N$, $\forall k \in K$
t_j^k	The travel time of the route k visiting vehicular node j; $\forall k \in K$, $\forall j \in I \cup M \cup D \cup N$
q_j^k	Number of persons at the vehicular node j assigned to route k; $\forall k \in K$, $\forall j \in I \cup M \cup D \cup N$
U_{jk}	Auxiliary (real) variable for subtour elimination constraint in the route of bus k; $\forall k \in K$

In the second stage, the greedy algorithm and the Dijkstra algorithm were embedded in GA, where demand points were assigned to selected stops, and feeder routes were designed to visit these selected stops. Through these two operations, the total walking time and riding time of all passengers at demand points in the objective function were also determined.

4.1. GA in the First Stage. The chromosome of solution to the problem was represented by a two-dimensional vector $U = (U_1, U_2)$, where each element u_i in $U_1 = (u_1, u_2, \ldots, u_I)$, being a natural number in 1, 2, ..., K, denotes a route u_i visiting the demand point i ($i = 1, 2, \ldots, I$); each element u_k in $U_2 = (u_{I+1}, u_{I+2}, \ldots, u_{I+K})$, being a real number, denotes departure frequency of the route k ($k = 1, 2, \ldots, K$). For example, a chromosome vector $U = \{11221245\}$ of two routes and six demand points could be coded as follows: route 1 visits demand points 1, 2, and 5; route 2 visits demand point 3, 4, and 6; and the departure frequencies of the two routes are 4 and 5, respectively.

A randomly generated chromosome may not comply with the partial constraint. To evaluate the quality of chromosomes, the fitness function is designed using penalty functions to eliminate infeasible solutions from the population in the process of evolution.

$$\begin{aligned} F = f &+ H \cdot \sum_{\forall k \in K} \left[\max\left\{ \sum_{\forall i,\, j \in N \cup D \cup M} x_{ij}^k d_{ij} - D_{\max},\, 0 \right\} + \max\left\{ \sum_{\forall i,\, j \in N \cup D \cup M} x_{ij}^k t_{ij} - T_{\min},\, 0 \right\} \right] \\ &+ H \cdot \left[\max\{ f_k - F_{\max},\, 0 \} + \max\{ F_{\min} - f_k,\, 0 \} \right] \\ &+ H \cdot \left[\max\left\{ \frac{\sum_{\forall j \in N} q_j^k}{Q} - R_{\max},\, 0 \right\} + \max\left\{ R_{\min} - \frac{\sum_{\forall j \in N} q_j^k}{Q},\, 0 \right\} \right]. \end{aligned} \tag{21}$$

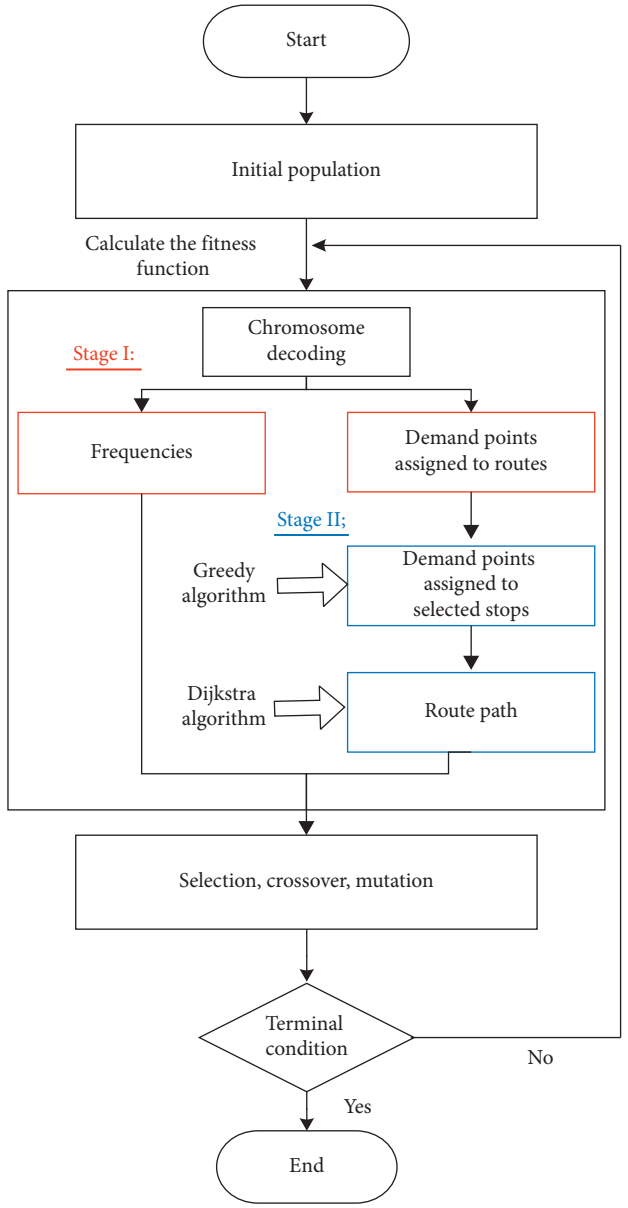


Figure 3: Flowchart of the GA-based two-stage algorithm.

Figure 3 details the evolutionary process of the GA. A given number of solutions were randomly generated in the initial population. Note that the departure frequency of the route could be obtained to determine the part of the evaluation fitness function. A specific heuristic at the second stage was embedded in this first-stage algorithm to obtain the assignment of each demand point to one of the nearest stops and route visiting stops to determine the rest of the evaluation fitness function. After evaluating the fitness values of these solutions, a certain number of offspring were generated from parents using the roulette-wheel selection method. To generate a better solution, single-point crossover and mutation operators were then adopted to exchange genes of any two individuals, or the genes of the individuals themselves to prevent the same genes in individual chromosome from appearing on next population. The algorithm was stopped until a preset maximal number was met.

Table 2: Information of all demand points.

No.	q_i
P1	21
P2	36
P3	40
P4	63
P5	18
P6	32
P7	17
P8	24
P9	24
P10	16
P11	23
P12	27
P13	13
P14	32
P15	40
P16	36
P17	17
P18	35
P19	23
P20	48
P21	42
P22	35
P23	42
P24	40
P25	35

4.2. Greedy and Dijkstra Algorithm in the Second Stage. In the first stage, all demand points were assigned to different routes. If the assignment of each demand point to one of the nearest stops and a sequence of stops along with designed routes were determined, a complete solution could be obtained.

In this section, a stop would be selected for each demand point based on the principle of minimization of walking distance, that is, $h_{ij} = \{j \mid \min[p_i \cdot d_{ij}/v], \forall i \in I\}, \forall j \in N$. In such a case, the Dijkstra algorithm (seen from references 6 and 7) was implemented to search the shortest route dispatching from a depot to the transportation hub so as to order the sequence of selected stops for the route, that is, x_{jm}^k.

4.3. Numerical Example

4.3.1. Example Description and Data Preparation. An FTS network, consisting of 5 depots (D1–D1), 25 demand points (P1–P25), 42 candidate stops (S1–S42), and 1 transportation hub (M), was used in the case study to prove the effectiveness of models and algorithms. The travel information of all passengers is shown in Table 2. The main inputs of this study were as follows:

(i) Number of routes: 3
(ii) Capacity of a vehicle: 35 per
(iii) Maximum travel time of each route: 30 min
(iv) Minimum length of each route: 4 km
(v) Minimum and maximum hourly departure frequencies: 5 and 10

Table 3: Assignment of all demand points to selected pick-up stops visited by routes.

Demand point	Selected stops	Number of persons	Assigned route	Walk distance (m)
P7	S3	44	R1	303.9
P12				244.9
P11	S6	23		111.3
P24	S34	40		73.7
P25	S22	35		106.9
P1	S11	21		219.8
P2	S10	36		215.6
P3	S13	40		54.8
P4	S14	63		140.9
P5	S15	18		192
P6	S16	32	R2	116.6
P8	S8	24		38.
P9	S7	24		175.5
P10	S5	16		145.3
P13	S4	45		439.7
P14				235.8
P15	S17	40		219.6
P16	S30	36		188.4
P17	S24	17		78.9
P18	S26	35		66.2
P19	S41	71	R3	106.7
P20				36.7
P21	S40	42		13.3
P22	S39	35		20
P23	S32	42		147.3

Table 4: Optimal scheme for route design and frequency setting.

Route	Sequence of stops visited by routes	Frequencies	Load rate	Length (km)	Travel time (min)
R1	D1-S3-S6-S22-S34-M	5	0.81	12.0	25.6
R2	D2-S4-S5-S7-S8-S10-S11-S13-S14-S15-S16-S17-M	9	1.14	10.5	19.3
R3	D3-S24-S26-S30-S32-S39-S40-S41-M	9	0.88	9.9	24.5

(vi) Minimum and maximum load factor: 0.7 and 1.2

(vii) Walking speed $v = 110$ m/min

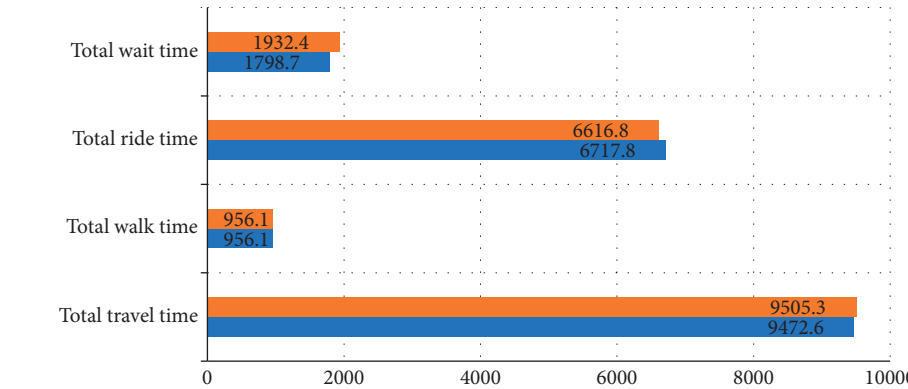
(vii) Parameters of the hybrid algorithm: iteration times, 100; chromosome number, 500; crossover rate, 0.9; and mutation rate, 0.1

5. Results

The solution involves stop selection, pedestrian guidance, transit routing, and scheduling. Table 3 details the assignment of all demand points to selected stops visited by three routes, which include the number of people getting on and off at a station and the walking distance between demand points and selected stops. Taking a selected stop S3 visited by the route R1 as an example, passengers at P3 and P12 get on at this place by walking 303.9 m and 244.9 m, respectively. Table 4 also details the routing and scheduling plans for these routes, which include the route length, running time, departure frequency, and load factor. Taking the route R1 as an example, the total number of passengers was 142 per/h, so the departure frequency was 5 veh/h, and the load rate was 0.81. A map-based graphical illustration of bus routing plans and passenger guidance is shown in Figure 4, where a yellow circle denotes the transportation hub, a red square denotes a depot, a blue dot denotes a stop, and a black dot denotes a demand point. The green dash line represents route 1, the red dash line represents route 2, and the blue dash line represents route 3. In addition to the bus route plan, the red solid lines between selected blue dots and all-black dots indicate the walking paths of passengers.

Furthermore, Figure 5 analyzes the difference in solution performance between the traditional and the proposed model under three routes. The walking distance of these two models remained unchanged due to the reasonable layout of demand points and stops. Compared with the traditional model, the total travel time in this study increased while the total waiting time decreased. As the increase in the former was less than the decrease in the latter, the objective value (i.e., total travel time) in this study was less than that of the traditional model. This was because this study considered the route design and departure frequency to pursue global optimization, avoiding local optimum of the traditional model, where first the route design is determined and then the departure frequency is calculated.

Figure 4: Roadmap of the optimal solution.



	Total travel time	Total walk time	Total ride time	Total wait time
Traditional model	9505.3	956.1	6616.8	1932.4
Proposed model	9472.6	956.1	6717.8	1798.7

Figure 5: Comparative analysis of the traditional and the proposed models under three routes.

Table 5: Comparison of model and algorithm performance under different routes.

Scenario	Objective (min)		Computation time (min)		Total riding time (min)	Total waiting time (min)	Total walking time (min)	Average load factor	Total frequency	Total route length (km)	Total route time (min)
	Cplex	GA	Cplex	GA							
3 routes	10230.3	9472.5	20.1	1.2	6717.8	1798.7	956.1	0.94	23	32.4	69.4
4 routes	9398.1	8621.8	126.3	1.6	5580.5	20851	956.1	0.74	30	33.1	73.8
5 routes	9557.2	8900.2	1514.6	1.9	4798.8	3145.3	956.1	0.74	30	35.7	76.0

5.1. Sensitivity Analysis. Table 5 also analyzes the influence of different routes on the model performance. The results are shown as follows:

(1) As each route started from the depot and returned to the transportation hub, the increase in the number of routes led to some invalid mileage and time, which increased the total travel distance and time of all routes. However, the total riding time of passengers gradually reduced.

(2) As the number of routes increased, the number of passengers in demand points visited by each route decreased, which led to a reduction in the departure frequency. In such a case, the waiting time of all passengers was gradually increased.

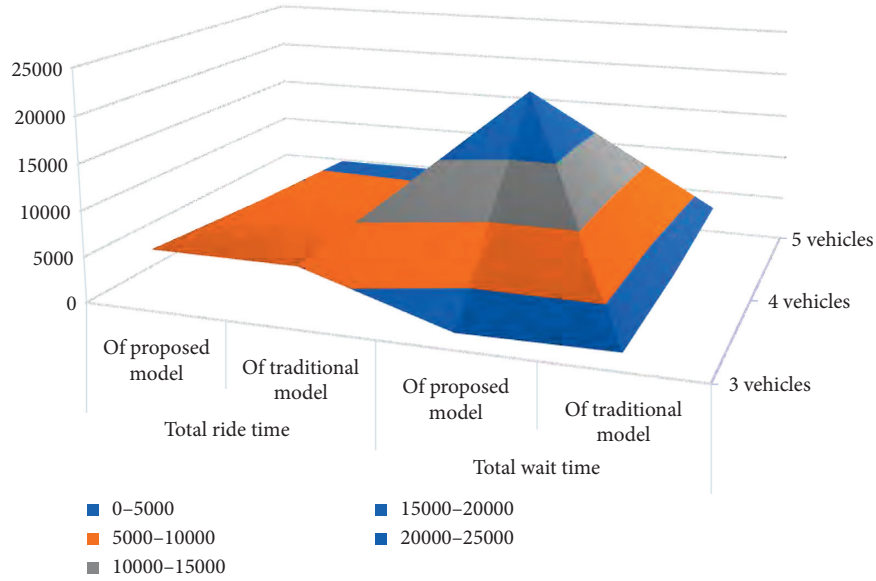


Figure 6: Comparative analysis of the traditional and the proposed models under different routes.

(3) With the increase in the number of routes, the total walking time of passengers remained unchanged because passengers tended to choose the nearest stops to get on and off the bus. Hence, the changing trend of the objective function was related to the decrease in the travel time of the passengers and the increase in the waiting time.

Besides, the difference in the solution quality between the proposed algorithm and the Cplex was within 10%, but the solution time greatly reduced, which proved the effectiveness of the algorithm.

Finally, Figure 6 further analyzes the difference in model performance between the traditional and the proposed models under different routes. Their waiting time reduced, and riding time increased. The results were consistent with those shown in Figure 5.

6. Conclusions

This study presented an integrated model for FTS to simultaneously coordinate stop selection, feeder transit routing, and frequency-setting process to trade off the operation efficiency. To solve such an NP-hard large-scale problem, a two-stage heuristic algorithm-based GA was further designed, where the assignment of all demand points to selected pick-up locations visited by different routes, as well as the determination of departure frequency, was obtained using GA in the first stage. In the second stage, the Dijkstra algorithm was embedded in the GA to calculate the shortest path for the vehicle visiting selected stops. A case study showed that the proposed model could save 0.34% of the total travel time for all residents, compared with the traditional model. The main reason was that the traditional model separated the line design from the departure frequency, which led to the departure frequency range and full-load level not affecting the design process of the route layout. In such a case, the passenger flow of each route might be unbalanced and the waiting time of some routes might increase. In order to validate the performance of developed heuristic algorithm, a comparison of CPLEX solutions and heuristic algorithm solutions in case of designing 3 routes, 4 routes, and 5 routes, respectively, the results suggested that the proposed algorithm can yield effective solutions to the proposed problem in an acceptable computation time. Furthermore, the increase in the number of routes led to the decrease in the travel time of the passengers and the increase in the waiting time.

A reduction in the departure frequency of each route was also done.

However, the premise of this study was that each demand point must be assigned to only a selected stop, which neglected the notion that travel needs were split and each group of passengers could choose a different stop to get on and off the bus when the number of passengers at the demand points exceeded the route capacity. Further, inferred OD table and travel time are static in this paper. In this sense, the problem considered here is that model uses a static representation of the pedestrian and vehicular network. Hence, it is worthwhile to extend the proposed model with a split travel demand and time-varying demand and travel times in future studies.

Data Availability

Some or all data, models, or codes generated or used during the study are available from the corresponding author by request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was jointly supported by the Central College Basic Scientific Research Operating Expenses in Civil Aviation University of China (no. 3122019126), the National Engineering Laboratory of Integrated Transportation Big Data Application Technology (no. CTBDAT201907), Nantong Science and Technology Innovation Program (MS22018012), the Six Talent Peaks Project of Jiangsu Province, China (SZCY-009), Key Science and Technology Projects in the Transportation Industry in China (2018-MS3-083), Chengdu Science and Technology Bureau (no. 2019-RK00-00029-ZF), and Science and Technology Think Tank of Nantong City (no. CXZK001, no. CXZK002).

References

[1] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, and J. C. Muñoz, "Planning, operation, and control of bus transport systems: a literature review," *Transportation Research Part B: Methodological*, vol. 77, pp. 38–75, 2015.

[2] M. Wei, B. Sun, and S. L. Zhu, "A hybrid-type indicator set pairs analysis model for evaluating transit operational efficiency," *Journal of Nonlinear and Convex Analysis*, vol. 30, no. 5, pp. 895–904, 2019.

[3] M. Wei, B. Sun, and R. Sun, "Expected value model of bus gas station site layout problem with fuzzy demand in supplementary fuel using genetic algorithm," *Cluster Computing*, vol. 22, no. 1, pp. 809–818, 2019.

[4] W. Y. Szeto and Y. Jiang, "Transit route and frequency design: bi-level modeling and hybrid artificial bee colony algorithm approach," *Transportation Research Part B: Methodological*, vol. 67, pp. 235–263, 2014.

[5] W. Y. Szeto and Y. Wu, "A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong," *European Journal of Operational Research*, vol. 209, no. 2, pp. 141–155, 2011.

[6] B. Sun, M. Wei, and W. Wu, "An optimization model for demand-responsive feeder transit services based on ride-sharing car," *Information*, vol. 10, no. 12, pp. 370–386, 2019.

[7] X. Li, M. Wei, J. Hu, Y. Yuan, and H. Jiang, "An agent-based model for dispatching real-time demand-responsive feeder bus," *Mathematical Problems in Engineering*, vol. 2018, Article ID 6925764, 11 pages, 2018.

[8] L.-B. Deng, W. Gao, W.-L. Zhou, and T.-Z. Lai, "Optimal design of feeder-bus network related to urban rail line based on transfer system," *Procedia—Social and Behavioral Sciences*, vol. 96, pp. 2383–2394, 2013.

[9] E. Holryod, "Optimum bus service: a theoretical model for a large uniform urban area," in *Proceedings of the Third International Symposium on the Theory of Traffic Flow*, pp. 308–328, Elsevier, New York, NY, USA, June 1965.

[10] B. F. Byrne, "Public transportation line positions and headways for minimum user and system cost in a radial case," *Transportation Research*, vol. 9, no. 2-3, pp. 97–102, 1975.

[11] A. Ceder, *Public Transit Planning and Operation: Theory Modeling and Practice*, Taylor & Francis, Didcot, UK, 2007.

[12] F.J. M. Salzborn, "Optimum bus scheduling," *Transportation Science*, vol. 6, pp. 137–148, 1972.

[13] J. H. Bookbinder and A. Désilets, "Transfer optimization in a transit network," *Transportation Science*, vol. 26, no. 2, pp. 106–118, 1992.

[14] V. F. Hurdle, "Minimum cost locations for parallel public transit lines," *Transportation Science*, vol. 7, no. 4, pp. 340–350, 1973.

[15] S. Zhu and F. Zhu, "Multi-objective bike-way network design problem with space–time accessibility constraint," *Transportation*, vol. 45, pp. 345–361, 2019.

[16] H. Zhang, S. Zhao, H. Liu, and S. Liang, "Design of limited-stop service based on the degree of unbalance of passenger demand," *PLoS One*, vol. 13, no. 3, Article ID e0193855, 2018.

[17] W. Fang, T. Wang, and R. Huang, "Network design of multi-objective continuous equilibrium with stochastic demand," *Journal of Central South University (Science and Technology)*, vol. 49, no. 9, pp. 2350–2355, 2018.

[18] Q. K. Wan and H. K. Lo, "A mixed integer formulation for multiple-route transit network design," *Journal of Mathematical Modelling and Algorithms*, vol. 2, no. 4, pp. 299–308, 2003.

[19] S. Chiou, "Bilevel programming for the continuous transport network design problem," *Transportation Research Part B: Methodological*, vol. 39, no. 4, pp. 361–383, 2005.

[20] F. Zhao and X. Zeng, "Optimization of transit route network, vehicle headways and timetables for large-scale transit networks," *European Journal of Operational Research*, vol. 186, no. 2, pp. 841–855, 2008.

[21] S. Schéele, "A supply model for public transit services," *Transportation Research Part B: Methodological*, vol. 14, no. 1-2, pp. 133–146, 1980.

[22] A. F. Han and N. H. M. Wilson, "The allocation of buses in heavily utilized networks with overlapping routes," *Transportation Research Part B: Methodological*, vol. 16, no. 3, pp. 221–232, 1982.

[23] M. Wei and B. Sun, "Bi-level programming model for multimodal regional bus timetable and vehicle dispatch with stochastic travel time," *Cluster Computing*, vol. 20, no. 1, pp. 401–411, 2017.

[24] İ. Ö. Verbas and H. S. Mahmassani, "Optimal allocation of service frequencies over transit network routes and time periods," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2334, no. 1, pp. 50–59, 2013.

[25] I. O. Verbas and H. S. Mahmassani, "Integrated frequency allocation and user assignment in multi-modal transit networks: methodology and application to large-scale urban systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2498, no. 1, pp. 37–45, 2015.

[26] N. Nassir, M. Hickman, and Z.-L. Ma, "A strategy-based recursive path choice model for public transit smart card data," *Transportation Research Part B: Methodological*, vol. 126, pp. 528–548, 2019.

[27] V. Chiraphadhanakul and C. Barnhart, "Incremental bus service design: combining limited-stop and local bus services," *Public Transport*, vol. 5, no. 1-2, pp. 53–78, 2013.

[28] E. Ruano-Daza, C. Cobos, T. J. José, M. Mendoza, and A. Paz, "A multi-objective bilevel approach based on global-best harmony search for defining optimal routes and frequencies for bus rapid transit systems," *Applied Soft Computing*, vol. 67, pp. 567–583, 2018.

[29] M. T. Brugal, G. Barrio, L. D. L. Fuente, E. Regidor, L. Royuela, and J. M. Suelves, "Factors associated with non-fatal heroin overdose: assessing the effect of frequency and route of heroin administration," *Addiction*, vol. 97, no. 3, pp. 319–327, 2002.

[30] A. T. Buba and L. S. Lee, "A differential evolution for simultaneous transit network design and frequency setting problem," *Expert Systems with Applications*, vol. 106, pp. 277–289, 2018.

[31] M. Nikoli and D. Teodorovi, "A simultaneous transit network design and frequency setting: computing with bees," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7200–7209, 2014.

[32] B. Sun, M. Wei, C. F. Yang, and A. Ceder, "Solving demand-responsive feeder transit service design with fuzzy travel demand: a collaborative ant colony algorithm approach," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 1, pp. 3555–3563, 2019.

[33] B. Sun, M. Wei, and S. Zhu, "Optimal design of demand-responsive feeder transit services with passengers' multiple time windows and satisfaction," *Future Internet*, vol. 10, no. 3, pp. 30–46, 2018.

[34] S. Chien, J. Byun, and A. Bladikas, "Optimal stop spacing and headway of congested transit system considering realistic wait times," *Transportation Planning and Technology*, vol. 33, no. 6, pp. 495–513, 2010.

[35] J. Wang, T. Weng, and Q. Zhang, "A two-stage multiobjective evolutionary algorithm for multiobjective multidepot vehicle routing problem with time windows," *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp. 2467–2478, 2019.

[36] D. Manjarres, L. Mabe, X. Oregi, and I. Landa-Torres, "Two-stage multi-objective meta-heuristics for environmental and cost-optimal energy refurbishment at district level," *Sustainability*, vol. 11, no. 5, p. 1495, 2019.

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7645153, 10 pages
https://doi.org/10.1155/2020/7645153

Research Article

Evaluation of Public Transport-Based Accessibility to Health Facilities considering Spatial Heterogeneity

Yuliang Zhang,[1] **Wenxiang Li**,[2] **Haopeng Deng**,[3] **and Ye Li**[1]

[1]*The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China*
[2]*Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*
[3]*State High-Tech Industrial Innovation Center, Shenzhen 518063, China*

Correspondence should be addressed to Wenxiang Li; lwxffff@gmail.com

Received 5 November 2019; Revised 3 January 2020; Accepted 8 January 2020; Published 1 February 2020

Guest Editor: Tao Liu

Ensuring adequate public transport-based accessibility to health facilities in different regions is a major concern of social equity and public health for government. However, the imbalanced spatial distribution of health facilities may lead to an inaccurate evaluation of the accessibility, which is shaped by both land use and transportation. To address this problem, this study proposed a new approach to evaluate the adequacy of public transport-based accessibility to health facilities considering the spatial heterogeneity. First, we obtained the spatial distribution of health facilities based on POI data, calculated the population centroids of census tract-based mobile phone positioning data, and estimated travel times from population centroids to every health facility based on web map services. Second, the public transport-based accessibility to health facilities was measured by the isochrone approach. Then, the spatial heterogeneity of the health facilities was quantified by a spatial proximity index based on the gravity model. At last, a benchmark curve of accessibility vs. spatial proximity was established to evaluate the public transport-based accessibility to health facilities in different areas with spatial heterogeneity. A case study of 218 census tracts in Shanghai was conducted to verify this method. Consequently, we successfully identified the census tracts where the public transport-based accessibility to health facilities is insufficient. It shows that even some census tracts within the central city areas are still short of public transport-based accessibility to health facilities, whereas some tracts in the urban periphery may have adequate public transport-based accessibility even though there are limited health facilities nearby.

1. Introduction

Because health and treatment are critical to our daily lives, sufficient accessibility to health facilities should be ensured in different regions and for different groups of people. In addition, as a large portion of the patients are the elderly, the poor and the disabled who do not own a car have to depend on public transport service to get to the hospital. Thus, how to evaluate the public transport-based accessibility to health facilities is an important problem that needs to be addressed.

In fact, the accessibility is shaped by two factors, namely, the land use and transportation [1]. Specifically, land use factor can be regarded as the spatial proximity of the health facilities, whereas the transportation factor is the level of public transport services. Because of the aggregation effects of urban populations and resources, spatial differentiation between central urban areas and suburban areas of the city is an inevitable economic trend, which is also referred to as spatial heterogeneity. Therefore, the land use factor resulting in the imbalanced distribution of health facilities may have a more significant impact on the accessibility than the development of public transport. For example, the regions surrounded by many health facilities may have high public transport-based accessibility to health facilities even though the level of public transport service is poor. On the contrary, the public transport-based accessibility in regions with few health facilities around can never reach a high level even though the public transport system is well developed.

Although the land use factor is not easy to be changed because the construction of health facilities is costly and time

consuming, improving the public transport system is more realistic. Then how to evaluate and improve the public transport-based accessibility in different regions with spatial heterogeneity? Most previous studies just directly compared the accessibility values of the different regions without consideration of the spatial heterogeneity. However, because the land use factor determines the availability and density of health facilities, it is unreasonable to require the regions with few health facilities around to achieve the same level of accessibility as the regions surrounded by many health facilities.

To address the problem above, this article aims to propose a new approach to evaluate the public transport-based accessibility to health facilities considering the spatial heterogeneity. By exploring the relationship between accessibility and spatial proximity of health facilities, we can establish a benchmark curve for evaluation. As a result, there is a benchmark accessibility corresponding to each value of spatial proximity. It is also regarded as the evaluation standard that the public transport-based accessibility should reach. In this way, we can estimate the gap between the actual accessibility and benchmark accessibility and recognize the geographical areas where the public transport-based accessibility to health facilities is insufficient.

The remainder of this article is organized as follows: Section 2 reviews the literature related to this study. The data including points of interests, mobile phone positioning, and travel time are analyzed in Section 3. Afterward, the methods of public transport-based accessibility measurements, spatial proximity analysis of health facilities, and evaluation model considering spatial heterogeneity are described in Section 4. Section 5 discusses the results and policy implications for the improvement of the accessibility in Shanghai. The final section concludes the highlights and limitations of this study and suggests future directions.

2. Literature Review

2.1. Measurements of Accessibility. Accessibility can be defined and measured with different methods [1]. These include well-known definitions such as the ease of reaching any activity area using a specific transport system [2], potential opportunities for interaction [3], the freedom of individuals to decide whether to participate in different activities [4], and the overall benefits provided by a given transport system [5]. According to studies of Baradaran and Ramjerdi [6] and Handy and Niemeier [7], accessibility measures have often been classified into four categories: (1) travel-cost approach, (2) isochrone approach, (3) gravity-based approach, and (4) utility-based approach [8]. A comparison of the four accessibility measures is presented in Table 1. Different measures have their own advantages and disadvantages. Researchers need to select the appropriate approach according to the research interests, objectives, and data acquisition.

Overall, existing research has provided a solid foundation for the measurement of accessibility. Different measures have their own advantages and disadvantages. Therefore, researchers need to choose the appropriate approach according to the research interests, objectives, and data acquisition. Because this article focuses on the evaluation of the public transport-based accessibility to health facilities, this article will only apply one of them rather than propose a new method for the accessibility measurement.

2.2. Accessibility to Health Facilities. In terms of accessibility to health facilities, there are many empirical studies in the international literature. Luo and Wang [10] used both the floating catchment area (FCA) method and gravity-based method to examine the spatial accessibility to primary health care in the Chicago ten-county region. They further proposed an enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians in northern Illinois [11]. These methods can identify persons with inadequate access to primary care physicians. Davy et al. [12] built a framework synthesis to analyze the accessibility to primary health-care services for indigenous people. They found that issues relating to the cultural and social determinants of health such as unemployment and low levels of education can influence whether indigenous patients, their families, and communities were able to access health care. Agbenyo et al. [13] presented an overview of geographic accessibility to health care services in rural Ghana using a mixed approach. They found that poor conditions of roads were the major barriers for household's accessibility to the district hospital.

These studies above mainly focus on the car-based accessibility to health facilities. However, many people, such as the elderly, the poor and the disabled, are typically more dependent on public transport to access the healthcare service. Higgs et al. [14] investigated the impact of different modes of travel (car versus bus) on associations between different measures of General Practitioner (GP) supply and the percentage of elderly patients. Hou and Jiang [15] analyzed the public transport-based accessibility of residential districts to the hospitals during the peak and off-peak hours in Changchun and examined the problems of public transport systems. LaMondia et al. [16] undertook a statistical comparison among four commonly applied measures of transit accessibility to healthcare facilities. Their results indicate that different categories of accessibility measures provide drastically different interpretations of accessibility which are not comparable and interchangeable. However, none of the previous studies have ever discussed the relationship between accessibility to health facilities and the public transport services.

2.3. Evaluation of Public Transport-Based Accessibility. Accessibility can be evaluated from various perspectives, including a particular group, mode, location or activity [17]. How accessibility is evaluated affects many planning decisions [18]. Meanwhile, four types of components can be identified from different evaluations: land use, transportation, temporal and individual [1]. In this study, we focus on the public transport-based accessibility, which describes the accessibility of locations to specific destinations by public transport mode. This is different from the public

Table 1: Comparisons of the four accessibility measures [9].

Accessibility measures	Description	Formula	Advantages	Disadvantages
Travel-cost approach	The ease with which any land-use activity can be reached from a location using a particular transport system	$A_i = \sum_{j=1}^{n} (1/f(c_{ij}))$	(i) Quite easy to understand and to calculate (ii) Less demanding on data	(i) Neglects variations across individuals (ii) Highly sensitive to the choice of demarcation area (iii) Neglects variations in the locations
Isochrone approach	The number of opportunities that can be reached within a given travel time, distance, or generalized cost	$A_i = \sum_{j \in \{c_{ij} < c^*\}} N_j$	(i) Relatively easy to understand and to calculate (ii) Easy to visualize	(i) Neglects variations across individuals (ii) Highly sensitive to the size of the range and the representation of opportunities (iii) Neglects variations in the locations
Gravity-based approach	Potential of opportunities for interaction	$A_i = \sum_{j=1}^{n} N_j / f(c_{ij})$	(i) Able to differentiate between locations (ii) Represents the joint effect of transport systems and land use patterns on accessibility	(i) Neglects variations across individuals (ii) Ambiguity regarding the magnitude of indicators
Utility-based approach	Expected maximum utility from a random utility considering individual characteristics	$A_i^n = \max(U_{ij}^n) = (1/\mu)\ln\sum_{j=1}^{n} e^{\mu(v_j^n - c_{ij}^n)}$	(i) Captures individual differences (ii) Captures the impact of all modes including auto, transit, and nonmotorized options	(i) Demands extensive data on locations and individuals' travel behavior

Note: A_i represents the measure of accessibility at location i, n represents the number of included locations, $f(c_{ij})$ represents the deterrence function and c_{ij} represents a variable that represents travel cost between locations i and j, $c^* c*$ represents the predetermined threshold within which the activity opportunities are counted, N_j represents the opportunities in a zone j, U_{ij}^n represents the utility of group n selecting the alternative, v_j represents some measure reflecting the attraction of the alternative j, and μ represents a positive scale parameter.

transport accessibility which was defined as "the quality of transit serving a particular location and the ease with which people can access that service" [19]. In short, the former is accessibility by public transport mode, while the latter is accessibility to public transport services.

On the one hand, there is a lot of research assessing the accessibility to public transport with consideration of the access to bus stops, duration of public transport journey, and access to destinations via public transport [20]. Many evaluation indexes have been proposed. One of the most recent developments and now seemingly one of the most widely used indexes is the Public Transport Accessibility Level (PTAL) developed by London Borough of Hammersmith and Fullham [21]. It reflects the walking time from a point of interest to the public transport access points, the reliability of the service nodes available, the number of services available within the catchment, and the average waiting time [19].

On the other hand, there are few studies that pay attention to the evaluation of public transport-based accessibility. Majority of them are limited to the simple comparison of accessibility measures. For example, public transport-based accessibility has been compared with car-based accessibility to employment and other land uses in some researches [22–26]. Yan-yan et al. [27] compared the public transport-based accessibility of different traffic analysis zones in Beijing by calculating the accessibility measures based on GIS. In their conclusions, the accessibility in dense-resource regions is always higher than that in sparse resource area [28]. However, there is a problem that the accessibility evaluation criterion for the regions with different spatial attributes should never be consistent because the amounts of potential opportunities are not comparable. In fact, the regions with high accessibility measures not always have sufficient accessibility because the evaluation standard should be higher with more resources to be connected. Therefore, high accessibility measure does not mean sufficient accessibility. These facts may have a great impact on the evaluation result of public transport-based accessibility to health facilities.

In summary, existing studies mainly focus on the measurements of accessibility but rarely look into the adequacy of public transport-based accessibility. Most of them fail to take the spatial heterogeneity into account when directly comparing the accessibility measures in different geographical regions. As a result, it is possible that high accessibility regions still lack public transport services, which cannot be recognized by previous methods. Therefore, it is more reasonable to compare the level of public transport services among the regions with the same spatial proximity

of the health facilities. In this context, this study expects to improve the evaluation method of public transport-based accessibility to health facilities by incorporating the spatial heterogeneity into the evaluation criterion. In this way, the problem of consistent evaluation criterion for regions with different spatial attributes can be addressed.

3. Data Preparation

Shanghai is a megacity covering an area of 6,340.5 square kilometers in China. This study takes Shanghai as a case to evaluate the public transport-based accessibility to health facilities in 218 census tracts, which were defined by the government of Shanghai in 2016 for the purpose of taking a census. We obtained the geography files of the census tracts from the Open Street Map (OSM) to display the boundaries of the census tracts on the map. Besides, the POI data and mobile phone positioning data were collected and analyzed to support this study. Both of them contain the information of the longitudes and latitudes, with which we can match them to the census tracts on the map using the Spatial Analyst Tool "Intersection" of ArcGIS.

3.1. Health Facility Analysis based on POI Data. A point of interest (POI) is a specific point location that someone may find useful or interesting, including businesses, hospitals, hotels, residences, educational buildings, and shopping malls. The POI data are usually described by a name, address, category, and a set of geospatial coordinates [29]. Because this study focuses on the health facilities, the POI data of all hospitals and healthcare centers in Shanghai were collected using the API of Baidu Map which is a web-based tool for interactions between users and enterprises. For example, one of POI data record ("a5097bfab4a5a97af13cbaa3," "Shanghai Changzheng Hospital," "Shanghai," "Huangpu," "415 Fengyang Road," "Medical Facility," "Class 3," "121.473726," "31.23890086") represents the item ID, name, province, region, address, category tag, class tag, longitude, and latitude, respectively.

As of July 2016, there are 679 regular and authoritative health facilities in Shanghai, including 78 Class-3(best) hospitals (or healthcare centers), 184 Class-2 hospitals, and 417 Class-1 or other hospitals. Given the addresses and longitudes and latitudes of these health facilities, the spatial distribution is displayed in Figure 1. Because health facilities with different classes may have different attractions for people, it is necessary to consider them separately. In addition, different groups of people may tend to choose different classes of hospitals, which usually depends on the patient's condition, income, age, education, etc. Thus, we give different attraction weights to the three classes of health facilities based on a web-based questionnaire survey where the importance of the three classes of health facilities was rated by 598 respondents sampled randomly. The sample distribution is close to the population distribution. The results show that 56% respondents tend to choose the Class-3 hospital, 29% respondents tend to choose Class-2 hospitals, whereas the rest of respondents tend to choose Class-1 or other hospitals. Then, the weights of different classes are normalized as shown in Table 2.

3.2. Population Distribution Analysis Based on Mobile Phone Positioning Data. As the population is seldom distributed homogeneously within a census tract, the population-weighted centroid rather than the simple geographic centroid of a census tract can represent the location of the population more accurately [10]. The population centroid of a tract may be distant from its geographic centroid, particularly in rural or peripheral suburban areas where tracts are large, and population tends to concentrate in limited space. The geographic coordinate (X_i, Y_i) of the population centroid for each census tract (i) is defined as,

$$X_i = \frac{\sum_{k=1}^{n} P_k X_k}{\sum_{k=1}^{n} P_k},$$
$$Y_i = \frac{\sum_{k=1}^{n} P_k Y_k}{\sum_{k=1}^{n} P_k}, \quad (1)$$

where (X_k, Y_k) is the geographic coordinate of home location k, P_k is the resident population of home location k, in the census tract (i), and n is the number of distinct home location.

However, the available census data in China can only provide the total population within the census tract, without the exact home locations. Fortunately, lots of studies have proved that mobile phone positioning data can provide good potential to identify the home and work anchor points of the population, since it tracks the positions of every mobile phone whenever physical moving, calling, texting, or surfing the Internet [9, 30]. Therefore, we use the mobile phone positioning data from China Unicom Communications Corporation of Shanghai collected between March 1, 2016, and March 31, 2016, to infer the home locations of residents.

The dataset contains much useful information such as the mobile identification number of the user, the time when the data record was collected, and the longitude and latitude of the user's estimated location. Then, an algorithm is developed to identify the home locations of residents based on the following assumptions: (1) most people stay at home between 10 pm–6 am every day, (2) the most frequent location appeared during this period is the aggregated home location, and (3) users whose home location remains consistent for more than 20 days during a month are residents of this area. More details about this algorithm could be found in reference [9, 29]. As a result, the resident population (P_i) of each distinct position (X_k, Y_k) can be estimated. Then, the distributions of population densities and centroid of 218 census tracts can be obtained, as shown in Figure 2.

3.3. Travel Time Estimation Based on Web Map Service. There is a tool of Baidu web map services called "Route Matrix API," which can calculate the travel times and distances between origins and destinations in the real-world road network [9]. The input parameters of this tool are

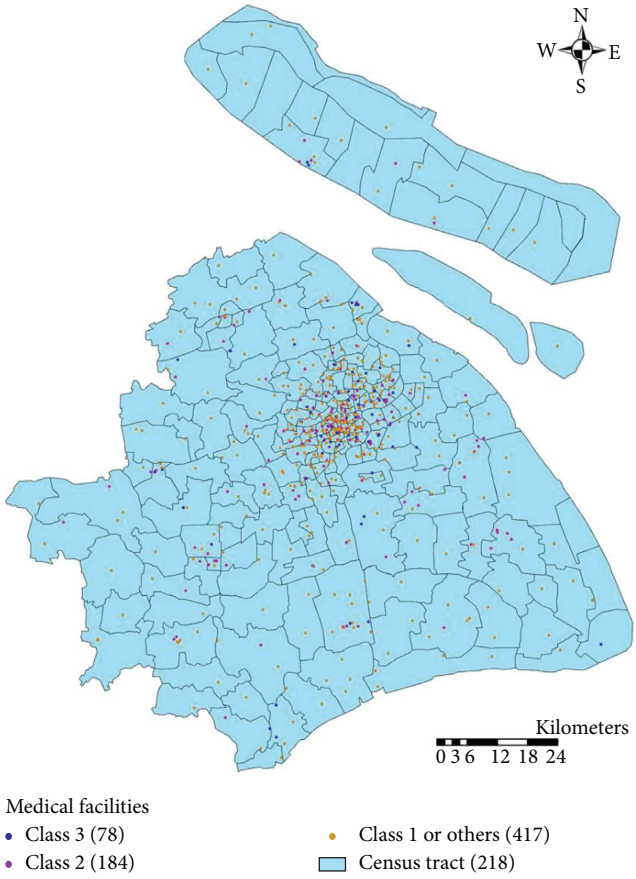


Figure 1: Spatial distribution of health facilities in Shanghai.

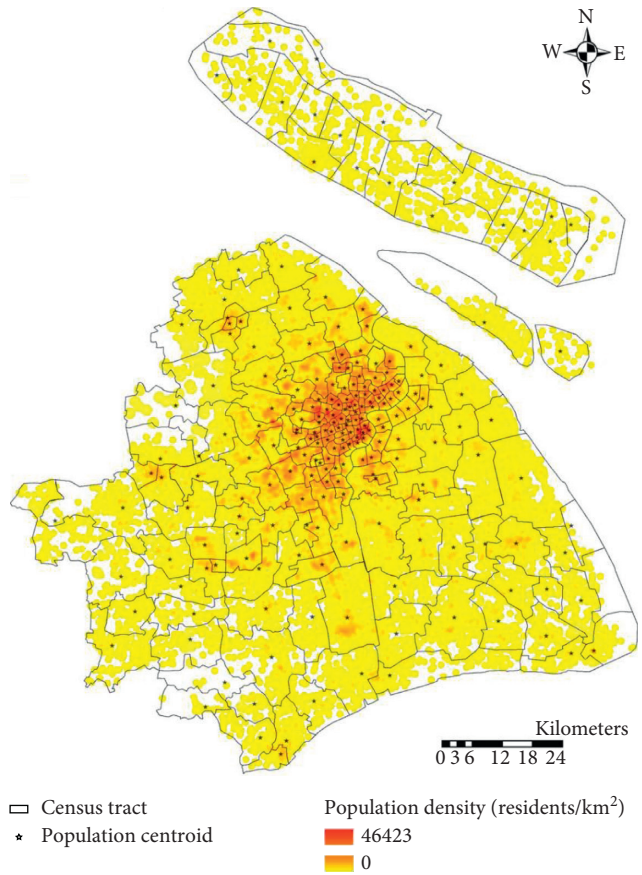


Figure 2: Distributions of population densities and centroids of 218 census tracts.

Table 2: Attraction weights of hospitals with different classes.

Hospital class	Attraction weight
Class 3	1
Class 2	0.52
Class 1 or others	0.27

geographic coordinates of the origins and the destinations (health facilities), and the mode of transportation (public transport). It uses the shortest path algorithm to estimate the travel time from origins to destinations by public transport mode considering the real-time and historical traffic status [31]. The estimated travel time also includes walking time from the origin to the public transport station and waiting time at the station. In addition, the public transport data of the web map service are almost complete and up to date, containing all operated routes (of subways, buses, trolley, etc.) in the city. Therefore, the estimated travel times based on this tool are more accurate than the travel times from survey data or planning models [32].

Based on this tool, we can estimate the travel time (t_{ij}) from each population centroid of the census tract (i) to each health facility (j) by public transport, given the locations of health facilities and population centroid of every census tract which are obtained above. To eliminate the influence of the commuting traffic, all the travel time was measured during 9 : 00 am–11 : 00 am at the off-peak hours.

4. Methodology

4.1. Measurement of Public Transport-Based Accessibility. This study adopts the isochrone approach which is also referred to as "cumulative opportunity" method to measure the public transport-based accessibility to health facilities. This measurement counts the number of health facilities that can be reached within a given threshold of travel time, distance, or generalized cost by public transport. Specifically, for each census tract (i), search every health facility (j) that are within the threshold travel time (t) from the population centroid of the census tract (i), and sum up the weighted attractions of these health facilities. Mathematically, the public transport-based accessibility to health facilities can be measured as

$$A_i = \sum_{j \in \{t_{ij} < t_0\}} w_j N_j, \tag{2}$$

where A_i represents the public transport-based accessibility for census tract (i), N_j represents the total number of attractions at each health facility (j), w_j represents the attraction weight of the health facility (j), and t_{ij} represents the travel time from population centroid of census tract (i) to the health facility (j) by public transport. According to the 5th Shanghai Comprehensive Transport Survey, the average travel time for residents in Shanghai is 29 minutes, whereas

32.3 minutes is the average travel time in the center area of the city. As a result, we take 30 minutes as the travel time threshold (t_0).

4.2. Spatial Proximity Analysis of Health Facilities. Because the health facilities are distributed unevenly throughout the city, there is inevitably spatial heterogeneity among these census tracts. Meanwhile, the spatial interaction between two geographical entities usually declines as the distance between them increases, which is generally called distance decay [33]. Therefore, these factors may have a great impact on the evaluation of public transport-based accessibility to health facilities. For example, if there are few health facilities available around this census tract, the public transport-based accessibility to health facilities for this tract can never reach a high level, even though the public transport system is good enough. To measure the spatial heterogeneity of the health facilities in different tracts, this study introduces a spatial proximity index derived from the gravity model [34], representing the overall spatial proximity to all available health facilities in the city:

$$R_i = \sum_{j=1}^{m} \frac{w_j N_j}{d_{ij}^{\eta}}, \tag{3}$$

where R_i is the spatial proximity index of health facilities for census tract (i), m is the total number of all health facilities, d_{ij} represents the geodetic distance between the population centroid of census tract (i) and the health facility (j), and η represents the distance sensitivity parameter which reflects the degree of unwillingness people feel toward traveling for an extended distance. In this article, we set $\eta = 1.08$ according to previous studies [35]. Therefore, we can use the spatial proximity index to represent the spatial heterogeneity, which can also measure the spatial aggregative state of all available health facilities. The larger the R_i is, the more concentratedly the health facilities are distributed around the population centroid of the census tract.

4.3. Evaluation of Public Transport-Based Accessibility to Health Facilities. As mentioned above, accessibility is generally determined by two factors: transportation and land use. In this study, these two factors are represented by the level of public transport services and the spatial proximity to all available health facilities respectively. To evaluate the adequacy of public transport services for health facilities independently, it is necessary to eliminate the effects of the spatial proximity. Therefore, we need to construct the relationship between the public transport-based accessibility and the spatial proximity.

Obviously, the accessibility increases with the spatial proximity index if the public transport service is sufficient. However, the growth of the accessibility will not be endless because of the limits of transportation capacity and space (as shown in Figure 3). It is very similar to the growth of population. Consequently, we employ the logistic function which is widely used for population growth [36] to construct the relationship between accessibility and spatial proximity with the following equation:

$$\widehat{A}_i = \frac{\alpha}{1 + e^{(\beta - R_i)/\chi}}, \tag{4}$$

where $\widehat{A}_i$ is defined as the benchmark accessibility, given a specific value of R_i; α is a parameter representing the growth limit; β is a parameter representing R's value at the inflection point of the curve where the value of A will be $\alpha/2$; and χ is a scale parameter.

In general, regression analysis estimates the conditional expectation of the dependent variable, given the independent variables. In other words, it reflects the average value of the dependent variable when the independent variables are fixed [37]. Therefore, the benchmark value ($\widehat{A}_i$) represents the expected (average) level of accessibility in the census tracts with the same level of spatial proximity. Based on this, we can establish a benchmark curve of accessibility vs. spatial proximity index to identify the census tracts where the public transport services should be improved to match the spatial proximity. If the actual accessibility (A_i) in a census tract is lower than $\widehat{A}_i$, then the public transport-based accessibility in that tract is considered to be insufficient, indicating that the public transport services are poor, and vice versa. With the calibrated benchmark model, we can calculate the difference value (ΔA_i) between the actual accessibility and benchmark accessibility as follows:

$$\Delta A_i = A_i - \widehat{A}_i. \tag{5}$$

However, the absolute value of gap may not suit for comparison. Therefore, we propose an index based on the relative value to evaluate the adequacy of public transport-based accessibility as follows.

$$\mathrm{EI}_i = \frac{\Delta A_i}{A_i} \times 100\%. \tag{6}$$

This evaluation index represents the percentage by which the actual accessibility is better or worse than the benchmark accessibility [38, 39]. Thus, we can use it to compare and grade the adequacy of public transport-based accessibility of different census tracts with spatial heterogeneity.

5. Results and Discussions

With the data and method presented above, we can respectively calculate the public transport-based accessibility to health facilities of 218 census tracts in Shanghai. The spatial distribution of accessibility is presented in Figure 4(a). Overall, the city of Shanghai is divided into 5 regions by 4 main ring roads. The regions within the Outer Ring Road are regarded as the central city areas of Shanghai, while the regions outside the Outer Ring Road are suburban areas. It can be found that the accessibility measures in central city areas are much higher than those of suburban areas. In the central city areas, the closer to the Inner Ring Road, the higher the public transport-based accessibility will be. Besides, the accessibility measures in the west of Huangpu River are usually higher than those in the east of Huangpu River. However, in the suburban areas, the

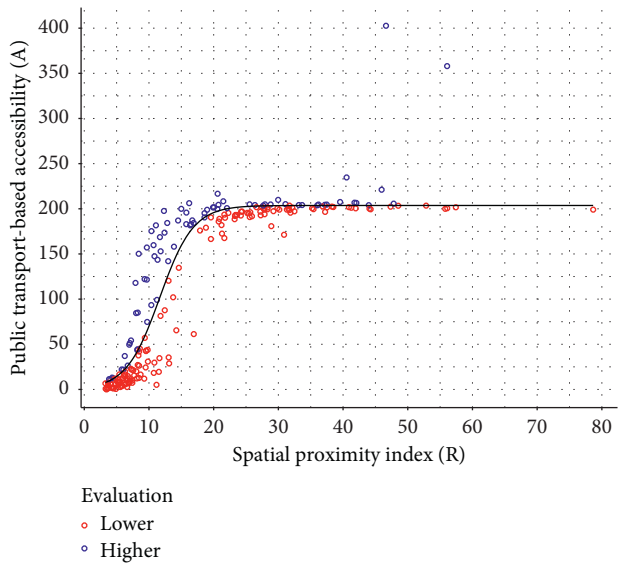


Figure 3: Benchmark curve of accessibility vs. spatial proximity.

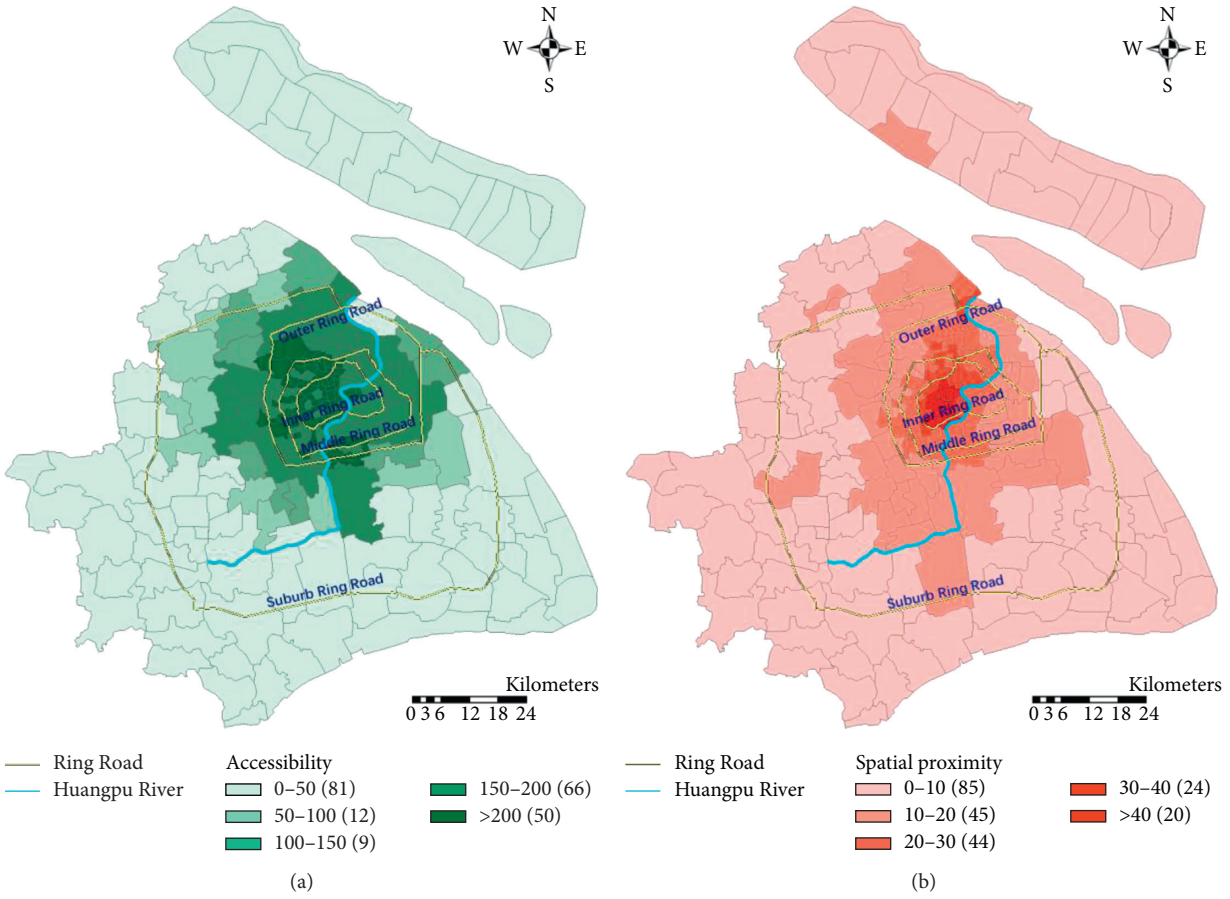


Figure 4: Spatial distributions of attributes for 218 census tracts. (a) Distribution of public transport-based accessibility. (b) Distribution of spatial proximity to health facilities.

accessibility measures are very low. Given Figure 4(a), we can measure and compare the weighted numbers of health facilities within 30 minutes by public transport in different tracts. However, do the tracts with large numbers of health facilities really have sufficient public transport-based accessibility to health facilities? Whether the short of health

facilities or the bad public transport services lead to the low accessibility in suburb areas? In fact, we cannot answer these questions by directly comparing the accessibility measures without considering the spatial heterogeneity.

To take the spatial heterogeneity of health facilities into account, we calculate the spatial proximity indexes of the 218 census tracts, respectively, as shown in Figure 4(b), which measures the spatial concentration degree of all available health facilities. The distribution trend of the spatial proximity index is very similar to that of accessibility. It reflects that most health facilities are mainly concentrated in the central city area, especially within the Inner Ring Road. In other words, the census tracts within the Inner Ring Road may have greater potential to access the health facilities if the public transport services are sufficient. Given Figure 4(b), we can only know which tracts are short of health facilities, but we still cannot decide whether the public transport services for health facilities in these tracts are sufficient or not.

To further evaluate the adequacy of public transport-based accessibility to health facilities in different census tracts considering the spatial heterogeneity, we use the accessibility measures and spatial proximity indexes to fit the benchmark curve based on equation (4). The calibration results are listed in Table 3, which indicates that all the estimated parameters are significant, and the function is effective. Thus, the benchmark accessibility can be calculated using this function.

The fitted curve and the actual data points (A_i, R_i) of the 218 census tracts are plotted in Figure 3, where the red dots represent the tracts whose accessibility is lower than the benchmark accessibility, whereas the blue ones represent the opposite. If A_i is lower than the benchmark accessibility at a fixed value of R_i, then the accessibility in the census tract (i) is worse than the average level among these tracts with equal spatial proximity. Therefore, we can conclude that the public transport service for health facility in the census tract(i) should be improved.

Furthermore, we present the relative evaluation index (E_i) of the 218 census tracts, respectively, on the map for comparison, as shown in Figure 5. The census tracts in cold colors (blue) may not have sufficient public transport-based accessibility to health facilities. It shows that above half of 218 census tracts cannot reach the benchmark accessibility. To better understand the results, the subway network which has been built in Shanghai and the bus stops are also plotted in Figure 5. We can find that the tracts not covered by dense subways or bus stations, especially the areas outside the Suburb Ring Road, usually have insufficient accessibility to the health facilities. Thus, it is urgent for these areas to improve their public transport services. However, even some census tracts within the Inner Ring Road may not have adequate public transport-based accessibility to health facilities. Although there are more facilities around these areas, the public transport services may not match with the high level of spatial proximity. Consequently, improvements to public transport services are also needed in some central city area. Although for some tracts around the Outer Ring Road there are limited health facilities nearby, their public transport services are good enough to connect these

Table 3: Calibration results of logistic function.

Parameters	Estimate	Std. error	t value	Pr(>|t|)
α	203.6747	3.7496	54.32	$<2e-16^{***}$
β	11.6744	0.3041	38.39	$<2e-16^{***}$
χ	2.5914	0.2470	10.49	$<2e-16^{***}$

Note: Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1.

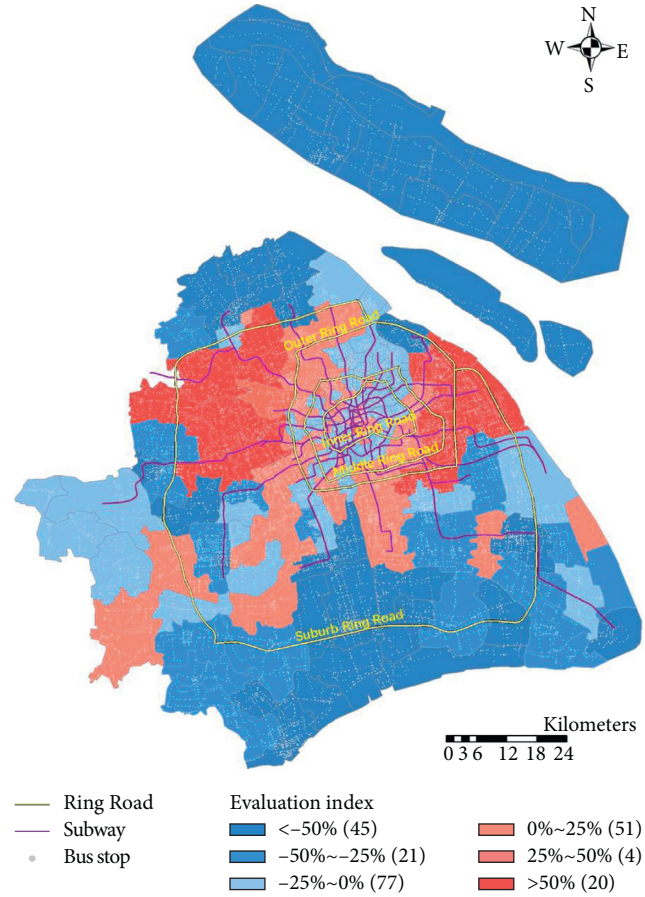


Figure 5: Distribution of evaluation indexes of public transport-based accessibility to health facilities.

available facilities. Therefore, it is not necessary to improve their public transport services. For this condition, only increasing the spatial proximity can further improve their public transport-based accessibility to health facilities.

6. Conclusion

This study proposed a novel method to evaluate the adequacy of public transport-based accessibility to health facilities by exploring the relationship between accessibility and spatial proximity of health facilities and establishing a benchmark curve for evaluation criterion. A case study of 218 census tracts in Shanghai was conducted to verify this evaluation method. With a relative evaluation index derived from the benchmark curve, we can compare and grade the adequacy of public transport-based accessibility of different census tracts with different spatial proximity. As a result, we

successfully identified the census tracts where the public transport-based accessibility to health facilities are insufficient. We find that even some census tracts within the central city areas are still short of public transport-based accessibility to health facilities, whereas some tracts in the urban periphery may have adequate public transport-based accessibility even though there are limited health facilities nearby. However, previous studies mostly failed to recognize these counterintuitive results mainly because of neglecting the spatial heterogeneity in different areas.

To sum up, the primary contribution of this research is to introduce a novel method to evaluate the adequacy of public transport-based accessibility to health facilities by exploring the relationship between accessibility and spatial proximity of health facilities. Specifically, we discover that the growth law of the accessibility vs. spatial proximity can be described by the logistic function, with which a benchmark curve can be established for the accessibility evaluation. In this way, we can efficiently recognize the geographical areas where the public transport-based accessibility to health facilities is insufficient. It is the first time that the effect of public transport on public health has been separated from the accessibility evaluation. Therefore, this research can not only help the government to better evaluate the public transport-based accessibility to health facilities but also provide a clear direction for decision-maker to implement the accessibility and healthcare improvement policies.

However, a few limitations still exist in this study, which also motivates a few future research directions. First, the measurement of accessibility only used a relatively simple and easy method (isochrones approach) because of the data availability. In future studies, a more comprehensive approach to measuring the public transport-based accessibility, which also incorporates the spatial heterogeneity of public facilities, would be proposed. Second, the logistic function is calibrated using only 218 census tracts' data. If the spatial scale of each zone can be made much smaller, the results may be even more accurate. What is more, other public facilities and additional cases should be applied to verify the effectiveness and applicability of this approach.

Data Availability

The mobile phone data used to support the findings of this study have not been made available because we do not have right to share this to the public. But we can provide part of the encrypted data. The POI 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 article.

Acknowledgments

This study was supported by the projects of the National Natural Science Foundation of China (grant numbers: 71774118 and 91546115) and the project of Science and Technology Commission of Shanghai Municipality (grant number: 16511105204).

References

[1] K. T. Geurs and B. van Wee, "Accessibility evaluation of land-use and transport strategies: review and research directions," *Journal of Transport Geography*, vol. 12, no. 2, pp. 127–140, 2004.

[2] M. Q. Dalvi and K. M. Martin, "The measurement of accessibility: some preliminary results," *Transportation*, vol. 5, no. 1, pp. 17–42, 1976.

[3] W. G. Hansen, "How accessibility shapes land use," *Journal of the American Institute of Planners*, vol. 25, no. 2, pp. 73–76, 1959.

[4] L. D. Burns, *Transportation, Temporal, and Spatial Components of Accessibility*, Lexington Books, Lexington, MA, USA, 1979.

[5] M. Ben-Akiva and S R. Lerman, "Disaggregate travel and mobility choice models and measures of accessibility," in *Proceedings of the 3rd International Conference on Behavioral Travel Modelling*, Tanunda, South Australia, 1979.

[6] S. Baradaran and F. Ramjerdi, "Performance of accessibility measures in europe," *Journal of Transportation and Statistics*, vol. 4, no. 3, 2001.

[7] S. L. Handy and D. A. Niemeier, "Measuring accessibility: an exploration of issues and alternatives," *Environment and Planning A: Economy and Space*, vol. 29, no. 7, pp. 1175–1194, 1997.

[8] X. Dong, M. E. Ben-Akiva, J. L. Bowman, and J. L. Walker, "Moving from trip-based to activity-based measures of accessibility," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 2, pp. 163–180, 2006.

[9] W. Li, Y. Li, X. Ban, H. Deng, H. Shu, and D. Xie, "Exploring the relationships between the non-work trip frequency and accessibility based on mobile phone data," in *Proceedings of the 97th Annual Meeting of Transportation Research Board*, Washington, DC USA, January 2018.

[10] W. Luo and F. Wang, "Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region," *Environment and Planning B: Planning and Design*, vol. 30, no. 6, pp. 865–884, 2003.

[11] W. Luo and Y. Qi, "An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians," *Health & Place*, vol. 15, no. 4, pp. 1100–1107, 2009.

[12] C. Davy, S. Harfield, A. McArthur, Z. Munn, and A. Brown, "Access to primary health care services for indigenous peoples: a framework synthesis," *International Journal for Equity in Health*, vol. 15, p. 163, 2016.

[13] F. Agbenyo, A. Marshall Nunbogu, and A. Dongzagla, "Accessibility mapping of health facilities in rural Ghana," *Journal of Transport & Health*, vol. 6, pp. 73–83, 2017.

[14] G. Higgs, R. Zahnow, J. Corcoran, M. Langford, and R. Fry, "Modelling spatial access to general practitioner surgeries: does public transport availability matter?" *Journal of Transport & Health*, vol. 6, pp. 143–154, 2017.

[15] S. Hou and H. Jiang, "An analysis on accessibility of hospitals in Changchun based on urban public transportation, an analysis on accessibility of hospitals in Changchun based on urban public transportation," *Geographical Research*, vol. 33, no. 5, pp. 915–925, 2014.

[16] J. J. LaMondia, C. E. Blackmar, and C. R. Bhat, "Comparing transit accessibility measures: a case study of access to healthcare facilities," in *Proceedings of the 90th Annual Meeting of Transport Research Board*, Washington, DC, USA, 2011.

[17] T. A. Litman, *Evaluating Accessibility for Transport Planning: Measuring People's Ability to Reach Desired Goods and Activities*, Victoria Transport Policy Institute, Victoria, Canada, 2017.

[18] T. Litman, "Measuring transportation: traffic, mobility and accessibility," *Institute of Transportation Engineers Journal*, vol. 73, no. 10, pp. 28–32, 2003.

[19] M. A. Joyce, "Proposed methodology for measuring public transport accessibility to employment sites in the auckland CBD," in *Proceedings of the 32nd Australasian Transport Research Forum*, Auckland, New Zealand, 2009.

[20] S. Mavoa, K. Witten, T. McCreanor, and D. O'Sullivan, "GIS based destination accessibility via public transit and walking in Auckland, New Zealand," *Journal of Transport Geography*, vol. 20, no. 1, pp. 15–22, 2012.

[21] Transport for London, *Measuring Public Transport Accessibility Levels*, Transport for London, London, UK, 2010.

[22] I. Benenson, K. Martens, Y. Rofé, and A. Kwartler, "Public transport versus private car GIS-based estimation of accessibility applied to the tel aviv metropolitan area," *The Annals of Regional Science*, vol. 47, no. 3, pp. 499–515, 2011.

[23] E. Blumenberg and P. Ong, "Cars, buses, and jobs: welfare participants and employment access in los angeles," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1756, no. 1, pp. 22–31, 2001.

[24] D. B. Hess, "Access to employment for adults in poverty in the buffalo-niagara region," *Urban Studies*, vol. 42, no. 7, pp. 1177–1200, 2005.

[25] M. Kawabata and Q. Shen, "Commuting inequality between cars and public transit: the case of the san francisco bay area, 1990–2000," *Urban Studies*, vol. 44, no. 9, pp. 1759–1780, 2007.

[26] M. Kawabata and Q. Shen, "Job accessibility as an indicator of auto-oriented urban structure: a comparison of boston and los angeles with tokyo," *Environment and Planning B: Planning and Design*, vol. 33, no. 1, pp. 115–130, 2006.

[27] C. Yan-yan, W. Pan-yi, L. Jian-hui, F. Guo-chen, L. Xin, and G. Yi, "An evaluating method of public transit accessibility for urban areas based on GIS," *Procedia Engineering*, vol. 137, pp. 132–140, 2016.

[28] I. Minocha, P. Sriraj, P. Metaxatos, and P. Thakuriah, "Analysis of transit quality of service and employment accessibility for the greater Chicago, Illinois, Region," *Transportation research record: Journal of the Transportation Research Board*, vol. 2042, pp. 20–29, 2008.

[29] W. Li, Y. Li, J. Fan, and H. Deng, "Siting of carsharing stations based on spatial multi-criteria evaluation: a case study of Shanghai EVCARD," *Sustainability*, vol. 9, no. 1, p. 152, 2017.

[30] C. Chen, L. Bian, and J. Ma, "From traces to trajectories: how well can we guess activity locations from mobile phone traces?" *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 326–337, 2014.

[31] Z. Yao, T. Xu, Y. Chen, Y. Jiang, and B. Ran, "Dynamic platoon dispersion model based on real-time link travel time," *IET Intelligent Transport Systems*, vol. 13, no. 11, pp. 1694–1700, 2019.

[32] W. Li, Z. Pu, Y. Li, and X. Ban, "Characterization of ridesplitting based on observed data: a case study of Chengdu, China," *Transportation Research Part C: Emerging Technologies*, vol. 100, pp. 330–353, 2019.

[33] P. J. Taylor and S. Openshaw, "Distance decay in spatial interactions," in *Proceedings of the Concepts and Techniques in Modern Geography*, Norwich, UK, 1975.

[34] A. S. A. Fotheringham, "New set of spatial-interaction models: the theory of competing destinations," *Environment and Planning A: Economy and Space*, vol. 15, no. 1, pp. 15–36, 1983.

[35] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding intra-urban trip patterns from taxi trajectory data," *Journal of Geographical Systems*, vol. 14, no. 4, pp. 463–483, 2012.

[36] A. Tsoularis and J. Wallace, "Analysis of logistic growth models," *Mathematical Biosciences*, vol. 179, no. 1, pp. 21–55, 2002.

[37] W. Li, L. Bao, L. Wang, Y. Li, and X. Mai, "Comparative evaluation of global low-carbon urban transport," *Technological Forecasting and Social Change*, vol. 143, pp. 14–26, 2019.

[38] Z. Yao, L. Shen, R. Liu, Y. Jiang, and X. Yang, "A dynamic predictive traffic signal control framework in a cross-sectional vehicle infrastructure integration environment," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2019.

[39] Z. Yao, Y. Jiang, B. Zhao, X. Luo, and B. Peng, "A dynamic optimization method for adaptive signal control in a connected vehicle environment," *Journal of Intelligent Transportation Systems*, pp. 1–17, 2019.