

Determining the Combination of High-Speed Trains with Different Running Distance

Hong Zhang, Huiling Fu*, Madison Merrill Billy Berto

School of Traffic and Transportation
Beijing Jiaotong University
Beijing, China
hlfu@bjtu.edu.cn

Abstract—China High-speed Railway currently operates a number of long-distance trains. However, it has a negative influence on dispatching and operation efficiency of EMU (Electric Multiple Unit) and crew. This paper aims at determining the combination of trains with various running distance. By designing crucial indicators that refer to the matching relation between trains and passengers, we formulate a closed-loop optimization procedure. It includes calculating the matching indicators, selecting to-be-adjusted trains, adjusting train distance, distributing passenger flow on trains, calculating the matching indicators again and giving feedback on adjusting train distance iteratively, until the efficiency of train operation plan is improved as expected. We apply the proposed approach taking Beijing-Guangzhou high-speed railway as an example. Compared with the original train operation scheme, the result shows that matching indicators for some trains are improved.

Keywords—*High-speed railway; Combination of train distance; Passenger flow distribution*

I. INTRODUCTION

Long routes, numerous train stations, and complex passenger demand are the features of China high-speed railway network. To satisfy various passenger travel demands, China high-speed railway operates a number of long-distance trains to reduce passenger transfer time and to increase train loading factor. However, excessive long-distance trains have a negative influence on dispatching and the efficiency of EMU and crew operation. Dispatchers will have difficulty in adjusting operation in the case long-distance trains are behind schedule. There is a reasonable combination problem existing in the proportion of trains with different transport distances. This paper is about determining the reasonable distance structure of high-speed trains under the condition of minimizing the impact on nonstop passenger flows. Organization work for transfer at stations, train dispatching and operation efficiency of EMU and crew will also be considered.

Over the years, numerous studies have been devoted to optimizing train operation plan. For line frequency optimization, reference [1] presented a linear multi-objective programming suitable for the short HSR line without branches in Taiwan. Reference [2] formulated problems of integrated train stop services and associated frequency optimization as a nonlinear mixed-integer programming model to reduce total costs. For determining train lines, reference [3] and Reference [4] both proposed cut-and-branch approaches and reported

their computations on real world railway data. Reference [5] used a branch-and-cut approach within reasonable time. Reference [6] presented a Dantzig-Wolfe decomposition approach to route travelers through an expanded network called “change and go” network to minimize the number of transfers or transfer time. Reference [7] proposed a multi-commodity model setting in the city public transport network. Reference [8] proposed a heuristic cost approach. The approach ignores the upper bounds of the problem and lines are added successively on the track with required minimum frequencies. However, previous studies mostly paid attention on optimizing train routes, train stopping patterns, train frequencies, etc. Few studies have considered the combination of trains with different distances. This is the main concern of this study.

II. THE “TRAIN-PASSENGER FLOW MATCHING” INDICATORS

There are many aspects to be considered when analyzing whether the combination of transport distance of multiple trains is reasonable. For example, the number of EMU can be reduced when trains with various distances have a reasonable combination (the same effect applies to the crew circulation). By reducing the proportion of long-distance trains, the train dispatching work complexity will be lower in the case of the trains deviating from the original schedule. On the other hand, the number of transfer passengers may be increased.

The factors above, however, are difficult to measure quantitatively. Therefore, this paper designs three quantitative indicators for the problem of combining transport distances of multiple trains. The Train-Passenger Flow Matching Indicators are selected including three specific indicators which are average passenger load rate, average traveling time lost to passengers riding on trains, and the transport distance matching level of the flow between trains and passengers. The formulate for indicators are stated as follows.

A. Average Passenger Load Factor

This is the turnover of passenger traffic divided by the turnover of passenger load, as shown in (1):

$$\lambda_{pl} = M_p / M_{pl} \times 100\% \quad (1)$$

Where, M_p is the turnover of passenger traffic. M_{pl} is the turnover of passenger load.

B. Average Travel Time Lost to Passengers

It refers to passengers time lost riding if a train makes intermediate stops. The calculation is as in equation below:

$$T_i = \sum_{r \leq i < s, r, s, v, e \in V} q_i \tag{2}$$

Where, q_i is the dwelling time of a train at a given station i . i is the station between passengers original station and destination station.

C. Distance Matching Level

This is the matching level between passenger flows at different travel distances and trains at different distance:

$$\gamma = \bar{L}_p / L_t \tag{3}$$

Where, \bar{L}_p is average passenger travel distance. L_t is the distance covered by the train.

III. COMBINATION OF HIGH SPEED TRAINS WITH DIFFERENT RAILWAY DISTANCES

A. Calculation Procedure

The determination of high-speed trains' reasonable running distance structure should start from analyzing the existing ticket data. At first, we calculate the 'train-passenger flow matching' indicators for every train according to (1), (2) and (3). Then we select different types of trains and adjusted according to the calculated results. After that, we provide some possible combination schemes to adjust train running distance. To justify the distance structure of adjusted trains, the passenger flow is distributed in a new scheme, and matching indicators are calculated once more according to the results of passenger flow distribution on trains. The final distance structure of trains will be decided when satisfactory indicators are obtained after many times in the loop iteration. The specific calculation procedure is as in Fig. 1.

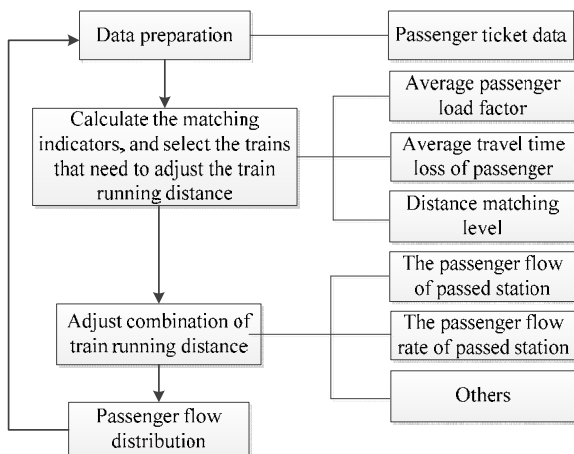


Fig. 1. Calculation procedure.

B. Selecting Trains Whose Distances Are to Be Adjusted

According to related previous analysis and the available statistical data, the distance of trains with more balanced passenger flow and more reasonable structure is two or three times of the average passenger travel distance. Trains with distance matching level of below 0.5 should be included in trains to be adjusted based on this conclusion.

For average passenger load rate and average passenger traveling time lost, we can draw a scatter chart of distance-train load scatter plots and distance-average time loss. We set 95 percent confidence interval for selection as in the formula below:

$$N_\lambda / N_z = 95\% \tag{4}$$

where, N_λ is the sum of trains with passenger load rate not lower than λ .

N_z is the sum of trains in the study rail network.

Similarly, we have the following formula:

$$N_T / N_z = 95\% \tag{5}$$

Where, N_T is the sum of trains with the average passenger loss time not higher than T .

Those trains with above two indicators according to (4) and (5) lower than 95 percent are considered as to be adjusted trains.

C. Adjusting Train Distance

The adjustment principles are:

- Reduce the impact on nonstop passenger trains as much as possible.
- The newly selected origin stations should meet the technical conditions of starting or terminating trains.

The passenger flow of passed stations refers to the passengers who traverse a station but do not board nor alight at the station. The passenger flow rate of passed stations refers to the result of the passenger flow of passed stations divided by the total amount of passengers arriving the destination station. The formula is shown in (6):

$$a = q_p / q \tag{6}$$

Where, q_p is the passenger flow of passed stations.

q is the total number of passengers arriving the destination station.

We will select a breakpoint station for a to be adjusted train if the value of this indicator according to (6) is low.

D. Distributing Passenger Flow on Trains

After we adjust the distance covered by the train, we then assign passenger flow on trains to verify the improvement level of the matching indicators.

A service network is at first constructed according to the updated train operation plan. How to construct this type of service network is described in detail in reference [6]. After that, we decide the impedance of passengers' travel itinerary and calculate its generalized cost, such that the probability of passengers' choice of one train can be calculated using the Logit model, which is as follows:

$$p_{n,k}(r,s) = \exp[-C_{n,k}(r,s)] / \sum_{p \in K_n(r,s)} \exp[-C_{n,p}(r,s)] \quad (7)$$

where, $C_{n,k}(r,s)$ is the decision items of generalized cost of a travel itinerary.

Next, we distribute passenger flow on the train service network. The overall calculation procedure is briefly described below.

Step 1: Calculate the passenger flow of type n from station r to s , denoted as $q_n(r,s)$, according to the passenger ticket data.

Step 2: Decide the travel itinerary collection $K_n(r,s)$ for every OD pair.

Step 3: Calculate the generalized cost $C_{n,k}(r,s)$ of every optional travel itinerary.

Step 4: Calculate the probability of optional travel itinerary for every OD pairs according to (7), which can also be called distribution proportion $p_{n,k}(r,s)$.

Step 5: Assign passenger flow on every optional travel itinerary according to $p_{n,k}(r,s)$.

Step 6: Examine whether the passenger flow of every OD pair is distributed, if yes, turn to step 7, otherwise turn to step 2.

Step 7: Calculate the passenger flow on each train.

When the above process is finished, we then calculate the average passenger load rate, average passenger traveling time lost, and distance matching level. If the values of those indicators are satisfactory as expected, then end the algorithm or else see the second step in the procedure in Fig. 1.

IV. CASE ANALYSIS

This section takes the Beijing-Guangzhou high-speed railway in China as an example to verify this approach. This railway line is one of the four vertical and four horizontal high-speed rail network in China. It connects the cities of Beijing and Guangzhou. After preprocessing the real world passenger flow data, our case study includes 28 stations and 100 trains.

The average passenger load factor of every train, the average traveling time lost by passengers, and the distance

matching level are first calculated. According to the results, 73 trains need to be involved in the procedure of passenger flow distribution. The trains require distance adjustment and related indicators are shown in Table I below.

The distances covered by long distance trains are shortened by altering train destination stations. Accordingly, we add trains on the shortened train running sections. The train distance adjustment scheme obtained in the first iteration is shown in Table II where new added trains are denoted in italic (note that abbreviation of station names is applied).

We carry out passenger flow distribution based on above train operation plan using the Matlab software. According to the obtained results, the matching indicators are calculated. The same procedure is repeated 36 times until we get a combination with better indicator values as expected. The part of final result is shown in Table III. It should be noted that, when adjusting train running distance, the original passenger demand which is not predicted is still used as the distributed passenger flow. Therefore, the OD pairs and the amount of passenger flow are not changed, and the average passenger load rate is the same as the overall level of the original scheme, that is the reason why this indicator is not included in Table III.

TABLE I. TO BE ADJUSTED TRAINS AND RELATED INDICATORS

Train number	Distance matching level	Average passenger load factor	Average travel time lost to passengers(min)
G6731	0.43	0.14	1.28
G8909	0.31	0.14	0.63
G6121	0.43	0.12	2.78
G1019	0.49	0.71	10.05
G1021	0.47	0.73	11.12
G551	0.33	0.52	10.03
G505	0.38	0.63	11.87
G65	0.30	0.78	10.54

TABLE II. THE TRAIN DISTANCE ADJUSTMENT SCHEME

Train Number	Train stopping scheme
G6731	BJ W-BD E-SJZ-HD E-ZZ E
G8909	CS S-HY E-CZ W-GZ S-SZN
G6121	CS S-ZZ W-CZ W-SG-GZ S-SZ N
G1019	CS H-SW-LY W-SG-GZ S-HM-SZ N
G1019N	WH-XN N-YY E-CS S
G1021	CS S-ZZ W-LY W-SG-QY-GZ S-HM-SZ N
G1021N	WH-XN N-YY E-CS S
G551	WH-XN N-YY E-CS S-ZZ W-HY E-CZ W-SG-QY-GZ S-SZN
G551N	LH W-ZMD W-XY E-WH
G505	BJ W-BD E-SJZ-XT E-HD E-AY E-XX E-ZZ E-ZMD W-XY E-WH
G505N	WH-XN N-YY E-CS S
G65	BJ W-BD E-SJZ-XT E-AY E-ZZ E-XC E-XY E-WH-XN N-YY E-CS S
G65N	CS S-HS W-CZ W-QY-GZ S

TABLE III. CALCULATION OF MATCHING INDICATORS BASED ON THE RESULTS OF PASSENGER FLOW ASSIGNMENT

	G6751	G8827	G1161	G6731	G1571	G8905
γ	1	0.27	0.85	0.35	0.35	0.69
T_i (min)	0	0	0	1.86	2.1	6.61
	G6105	G6121	G6011	G6013	G6017	G6019
γ	0.25	0.75	0.28	0.11	0.21	0.2
T_i (min)	0	13.18	0.51	0.65	2.41	1.89
	G1103	G1105	G1109	G1117	G1123	G1125
γ	0.18	0.46	0.28	0.2	0.2	0.1
T_i (min)	1.21	9.57	0.9	2.16	2.16	0.17
	G1154	G1001	G1003	G1005	G1007	G1009
γ	0.28	0.45	0.53	0.14	0.07	0.18
T_i (min)	4.71	10.52	11.27	1.74	0.3	3.88
	G519	G523	G527	G557	G573	G507
γ	0.12	0.28	0.18	0.12	0.37	0.18
T_i (min)	1.07	8.58	2.49	1.16	10.34	2.49
	G505N	G541	G547	G549	G73	G75
γ	0.31	0.1	0.36	0.3	0.48	0.29
T_i (min)	0	0.17	10.09	6.8	22.4	12.07
	G563	G565	G6101	G6103	G6031	G6015
γ	0.2	0.26	0.24	0.21	0.21	0.13
T_i (min)	0.65	2.11	0	0.85	0	3.23
	G8909	G1563	G1565	G6021	G6027	G6029
γ	0.58	0.23	0.76	0.43	0.13	0.16
T_i (min)	5.89	1.64	13.88	5.17	0.1	1.99
	G1127	G1129	G1133	G1013	G1015	G1019
γ	0.3	0.18	0.17	0.13	0.09	0.14
T_i (min)	1.78	2.21	1.58	2.14	0.72	1.25

Through optimization, the train operation plan with the new combination of trains with different distances is obtained. Compared with the original scheme, the new plan contains 68 trains with the distance matching level lower than 0.5 which is reduced by three trains. The train seating capacity utilization rate is also improved. The traveling time lost by passengers is 4.69 min on the average, which is reduced by 0.53 min. Resulting from this, the passengers' traveling speed can be increased.

After the combination of trains with different running distances is determined, the efficiency of EMU operation is also increased. Reducing the number of long-distance trains appropriately can increase the flexibility of circulating EMU trains, because an EMU can continuously serve transport tasks for a variety of running distances. At the same time, the reduction of long-distance trains can reduce the complexity of train dispatching work. The train dispatching measure is more flexible to handle the disruption of trains to reduce further delay.

However, the obtained combination of trains with different distances also has disadvantages. Although the goal of minimizing the number of transfers is considered when selecting the combination, the number of transfers still increases by 7,597 after adjusting the train running distance. However, by estimating the remaining seat capacity of other trains, approximately 60.8 percent of passengers can still choose other trains to accomplish their trips without transfers.

V. CONCLUSION

This paper proposed an approach for determining the combination of high-speed trains with different running distances. Three matching indicators were designed, including the average passenger load rate, the average traveling time lost by passengers, and the running distance matching level. We formulated a closed-loop optimization procedure. Firstly, the matching indicators were calculated, and to be adjusted trains can be accordingly selected. After that, we adjusted running distances of some trains, and distributed passenger flow on trains. Afterwards, the matching indicators were calculated again. Until the indicator values were at satisfactory level as expected, we ended the calculation and obtained the final train operation plan.

Taking Beijing-Guangzhou high-speed railway in China as an example, we verified the effectiveness of our proposed approach. Compared with the original train operation scheme, the results showed that two important matching indicators for some of the trains were improved after optimizing the combination of their running distances.

In the future, we will further quantitatively involve the influencing factors of EMU and crew circulation, and the complexity of train dispatching work in our approach in order to analyze comprehensive the economic benefits.

ACKNOWLEDGMENT

This research was jointly supported by the National Natural Science Foundation of China (No. 61703030), Beijing Training program for the Excellent Talents (No. 2017000020124G003), and the Fundamental Research Funds for the Central Universities (No. 2015JBZ011). We thank them for their support.

REFERENCES

- [1] Y. Chang, C. Yeh, and C. Shen, "A multi-objective model for passenger train services planning: application to Taiwan's high-speed rail line," *Transportation Research Part B*, vol. 34, pp. 91-106, 2000.
- [2] Y.Y. Ulusoy, S. Chien, and C. Wei, "Optimal all-stop, short-turn, and express transit services under heterogeneous demand," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2197, pp. 8-18, 2010.
- [3] M. Claessens, N. van Dijk, and P. Zwaneveld, "Cost optimal allocation of passenger lines," *European Journal of Operational Research*, vol. 110, pp. 474-489, 1998.
- [4] M. Bussieck, P. Kreuzer, and U. Zimmermann, "Optimal lines for railway systems," *European Journal of Operational Research*, vol. 96, pp. 54-63, 1997.
- [5] J.W. Goossens and L. Kroon, "A branch-and-cut approach for solving railway line planning problems," *Transportation Science*, vol. 38, pp. 379-393, 2004.
- [6] S. Scholl, "Customer-oriented line planning," Ph.D dissertation, University of Kaiserslautern, Kaiserslautern, Germany, 2005.
- [7] R. Borndörfer, M. Grötschel, and M.E. Pfetsch, "A column-generation approach to line planning in public transport," *Transportation Science*, vol. 41, pp. 123-132, 2007.
- [8] M. Goerigk, M. Schachtebeck, and A. Schöbel, "Evaluating line concepts using travel times and robustness," *Public Transport*, vol. 5, pp. 267-284, 2013.