Journal of Advanced Transportation

Simulation and Optimization for Railway Operations Management

Lead Guest Editor: Andrea D'Ariano
Guest Editors: Francesco Corman, Taku Fujiyama, Lingyun Meng, and Paola Pellegrini

Simulation and Optimization for Railway Operations Management

Journal of Advanced Transportation

Simulation and Optimization for Railway Operations Management

Lead Guest Editor: Andrea D'Ariano
Guest Editors: Francesco Corman, Taku Fujiyama, Lingyun Meng, and Paola Pellegrini

Editorial Board

Contents

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 4896748, 3 pages
https://doi.org/10.1155/2018/4896748

Editorial

Simulation and Optimization for Railway Operations Management

Andrea D'Ariano,[1] **Francesco Corman**,[2] **Taku Fujiyama**,[3] **Lingyun Meng**,[4] **and Paola Pellegrini**[5]

[1]*Roma Tre University, Rome, Italy*
[2]*ETH Zurich, Zurich, Switzerland*
[3]*University College London, London, UK*
[4]*Beijing Jiaotong University, Beijing, China*
[5]*Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux, Villeneuve d'Ascq Cedex, France*

Correspondence should be addressed to Andrea D'Ariano; dariano@ing.uniroma3.it

Received 26 June 2018; Accepted 26 June 2018; Published 2 August 2018

In the forthcoming decades, railway transport is expected to face a significant growth of traffic flows which will mostly have to be accommodated over the existing infrastructures. An increase of capacity utilization is needed to avoid reduction of reliability and punctuality of transport services. Furthermore, energy-efficient operations and multimodality opportunities are also topics of growing interest.

Intelligent railway operations management requires accurate modelling and simulation of train and pedestrian traffic flows and optimal management of key decisions at strategic, tactical, and operational levels. This special issue focuses on advanced mathematical modelling and optimal control in the railway domain and related multimodal transport networks. We aim to identify new methods for improving the effectiveness and efficiency of railway operations, including the development of advanced algorithmic techniques for timetabling, capacity management, infrastructure management, and traffic and passengers flow management.

This special issue has selected a compendium of research papers addressing recent theoretical and practical advances on railway operations management. 32 papers were submitted to this special issue, 14 of which were accepted for publication. As the guest editors of this special issue, we next summarize the 14 accepted papers.

In the paper titled "A Short Turning Strategy for Train Scheduling Optimization in an Urban Rail Transit Line: The Case of Beijing Subway Line 4" by M. Zhang et al., the authors propose a mixed-integer nonlinear programming model for the train scheduling with consideration of short turning strategy and train circulation plan. Mixed-integer linear programming approach is used to solve this optimization problem. Two case studies are carried out based on the data of Beijing subway line 4. The simulation results show that operation pattern with short turning train services can acquire a better train schedule to meet passenger demand and a better train circulation plan.

In the paper titled "Integrated Optimization on Train Control and Timetable to Minimize Net Energy Consumption of Metro Lines," Y. Zhou et al. present an integrated optimization model on train control and timetable to minimize the net energy consumption with consideration of utilizing regenerative energy. An improved model and algorithms on train control are proposed to attain energy-efficient speed profiles. Case studies on Beijing Metro Line 5 illustrate that the improved train control approach can save traction energy consumption by 20% in comparison with the commonly adopted train control sequence in timetable optimization.

In the paper titled "Application of Data Clustering to Railway Delay Pattern Recognition" by F. Cerreto et al., the authors employ K-means clustering to identify recurrent delay patterns on a high traffic railway line north of Copenhagen, Denmark. The clusters identify behavioral patterns

in the very large ("big data") data sets generated automatically and continuously by the railway signal system. The results reveal where corrective actions are necessary, showing where recurrent delay patterns take place. The demonstrated methodology is scalable and can be potentially transferred to any system of transport.

In the paper titled "Defining Reserve Times for Metro Systems: An Analytical Approach" by L. D'Acierno et al., the authors provide an analytical approach for determining operational parameters for metro systems so as to support the planning and implementation of energy-saving strategies. They develop a suitable methodology for estimating reserve times that represent the main rate of extra time needed to put eco-driving strategies in place. The approach proposed by the authors is applied in Line 1 of the Naples metro system, whose service frequency was duly taken into account, to analyze operation configurations and to quantify the amount of saved energy.

In the paper titled "Using Smart Card Data Trimmed by Train Schedule to Analyze Metro Passenger Route Choice with Synchronous Clustering," W. Li et al. analyze smart card data in Shanghai metro systems to understand mobility patterns. To do that, they cluster travelers and smart card data transactions by looking at the pure travel time. This is then appropriately converted to a model of route choice throughout the network. They found out that those steps can improve substantially the amount of insight into the travelers' behavior.

In "A Simulation Platform for Combined Rail/Road Transport in Multiyards Intermodal Terminals," X. Chen et al. tackle multimodal freight terminals where the combination between road and rail transport is investigated by means of simulation in terminals featuring multiple rail yards. The key features of the proposed approach based on Time Petri Nets are an increased realism of train operations and container movements in Qianchang railway terminal, including train routing dispatching rules. The validation on historical data allows evaluating influence of design parameters on container operations.

In the paper titled "On Individual Repositioning Distance along Platform during Train Waiting," F. Leurent and X. Xie propose a stochastic model in which user's journey is decomposed into phases of, successively, walking in the access station, platform positioning, waiting for boarding, train riding, and walking in the egress station. Egress times and exit instants are random variables that are characterized by distribution and mass probability functions under closed form for both single and distributed walking speed. Maximum likelihood estimation is proposed and applied to a case study of commuter rail line in Paris, France, with satisfactory numerical results.

In the paper titled "Efficiency Assessment of Transit-Oriented Development by Data Envelopment Analysis: Case Study on the Den-en Toshi Line in Japan," J. Guo et al. assess the efficiency of Transit-Oriented Development (TOD), that is, an urban planning approach that encourages a modal shift from private to public transportation, by applying the Data Envelopment Analysis (DEA) method. The ridership of public transportation is considered as the direct output characteristic of TOD efficiency, and nine indicators of ridership are selected as inputs of TOD. The Tokyo Den-en Toshi Line in Japan is investigated as a typical case of TOD.

In the paper titled "Stop Plan of Express and Local Train for Regional Rail Transit Line" by Q. Luo et al., the Logit model is used to analyze the behavior of passengers choosing trains by considering the sensitivity of travel time and travel distance. Based on the composition of passenger travel time, an integer programming optimization model for train stop scheme is proposed, which aims at minimizing the total passenger travel time for a certain regional rail line in Shenzhen. Genetic Algorithms are used to solve the problem. The simulation result shows the feasibility of the proposed model and the efficiency of the proposed algorithm.

In the paper titled "Fuzzy Approach in Rail Track Degradation Prediction" by M. Karimpour et al., an adaptive network-based fuzzy inference system (ANFIS) model is proposed to estimate rail track degradation for the curves and straight sections of Melbourne tram track system. A fuzzy approach is proposed due to the nonlinear and noisy nature of the data according to the data that were available on the Melbourne tram network. Experimental results demonstrate that the developed model is capable of estimating the long-term behavior of rail tracks and predicting the gauge values with an R2 of 0.6 and 0.78 for curves and straights, respectively.

In the paper titled "PULSim: User-Based Adaptable Simulation Tool for Railway Planning and Operations," Y. Cui et al. introduce a user-based, customizable platform to provide the ability of defining sophisticated workflows for users. As the preconditions of the platform, the design aspects for modelling the components of a German railway system and building the workflow of railway simulation are elaborated. Based on the model and the workflow, an integrated simulation platform with open interfaces is developed. Users and researchers gain the ability to rapidly develop their own algorithms, supported by the tailored simulation process in a flexible manner.

In the paper titled "The Planners' Perspective on Train Timetable Errors in Sweden" by C.-W. Palmqvist et al., typical errors in train timetables of railways, relevant reasons, and potential benefits of new tools and processes are investigated and reported through reviewing the state of practice and the state of the art in timetable planning, studying the research literature and railway management documents from several European countries, and conducting interviews with timetable planners in Southern Sweden.

In the paper titled "Calculation Method for Load Capacity of Urban Rail Transit Station Considering Cascading Failure," J. Huang et al. quantify the capacity of urban rail transit stations in terms of number of passengers. Their paper relies on an association network of facilities, set up based on the analysis of passenger service chain in station. The analysis of a case study of Lujiabang Road Station in Shanghai shows that this algorithm can search for capacity bottlenecks and help trace the load variation of facilities in different scenarios, providing support for passenger flow organization.

In the paper titled "Optimal Operation of High-Speed Trains Using Hybrid Model Predictive Control" by Y. Yang et al., a control framework is proposed for the controller

design of an automatic train operation (ATO) system. Firstly, a piecewise linear system is constructed through piecewise linearization of the Chinese high-speed train's nonlinear dynamics. Secondly, the piecewise linear system is transformed into a mixed logical dynamical system. For the latter system, a hybrid model predictive controller is designed to realize the precise control. Simulation results validate the effectiveness of the framework.

Taken together, these 14 papers point to interesting research directions in mathematical modeling, simulation, and particularly optimization to tackle practically oriented railway operations management challenges in several countries worldwide, including China, Denmark, Italy, France, Japan, Australia, Germany, and Sweden.

Andrea D'Ariano
Francesco Corman
Taku Fujiyama
Lingyun Meng
Paola Pellegrini

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 5367295, 19 pages
https://doi.org/10.1155/2018/5367295

Research Article

A Short Turning Strategy for Train Scheduling Optimization in an Urban Rail Transit Line: The Case of Beijing Subway Line 4

Miao Zhang, Yihui Wang, Shuai Su, Tao Tang, and Bin Ning

State Key Laboratory of Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Yihui Wang; yihui.wang@bjtu.edu.cn

Received 24 November 2017; Revised 28 February 2018; Accepted 10 May 2018; Published 2 August 2018

Academic Editor: Jose E. Naranjo

In urban rail transit systems, train scheduling plays an important role in improving the transport capacity to alleviate the urban traffic pressure of huge passenger demand and reducing the operation costs for operators. This paper considers the train scheduling with short turning strategy for an urban rail transit line with multiple depots. In addition, the utilization of trains is also taken into consideration. First, we develop a mixed integer nonlinear programming (MINLP) model for the train scheduling, where short turning train services and full-length train services are optimized based on the predefined headway obtained by the passenger demand analysis. The MINLP model is then transformed into a mixed integer linear programming (MILP) model according to several transformation properties. The resulting MILP problem can be solved efficiently by existing solvers, e.g., CPLEX. Two case studies with different scales are constructed to assess the performance of train schedules with the short turning strategy based on the data of Beijing Subway line 4. The simulation results show that the reduction of the utilization of trains is about 20.69%.

1. Introduction

With the rapid development of the urban rail transit systems, more and more researches pay attention to train scheduling. Researchers have devoted themselves to optimize the train schedule to assist help dispatchers make better decisions meeting passengers and operators expectations. The objective functions are most concentrated on enhancing passengers satisfaction in terms of degree of crowdedness [1–3], passengers waiting time [2, 4, 5] and passengers travel time [6, 7] or reducing operation costs in terms of energy consumption [6, 8–10], trains [11, 12], train travel time [13], etc. Not only does the urban rail transit systems expand fast, but also the passenger demand rises significantly. In the 15 urban rail lines operated by Beijing Mass Transit Railway Operation Corporation Limited, the numbers of sections (i.e., the trip between two consecutive stations) where load factors are over 100% and 120% are up to 38 and 9, respectively. Note that the load factor is a parameter to assess the performance of a transport system. It is defined as the dimensionless ratio of passenger-kilometers traveled to seat-kilometers available. In 2016, China's volume of passenger traffic reaches 16.09 billion per year except for five regional express rail lines and eight modern trams, which shows an increase by 16.6% compared to last year. Figure 1 shows a comparison between the passenger flow of a quantity of Chinese cities in 2015, 2016, and 2017 [14]. Under such huge pressure of large passenger demand, several strategies are adopted for operators to tackle it. For example, reducing the headway of departure times to increase the number of train services, which is limited by the capacity of urban rail transit line and minimum turnaround time required by trains, adopting the capacity restriction method like closing several stations that have large passenger flow, offering plenty of discounts during low peak period, and so forth. However, drawbacks remain in these strategies such as the shortage of trains and the dissatisfaction of passenger for inconvenience caused by using other means of transportation. In particular, the passenger demand of an urban rail transit line is not evenly distributed at all stations but shows a tendency such as a convex curve that increases gradually and then drops from the maximal volume point to the end of the line or a curve like a ladder [15]. Due to the unbalanced circumstance, train operation pattern with short turning strategy is better to satisfy the unevenly distributed passenger demand [16].

Table 1: Summary of the relevant studies with short turning train scheduling.

Publication	Objective	Decision variables	Solution approach
Wang [26]	minimize the number of trains	short turning service headway and cycles	fuzzy optimization
Wang and Ni [23]	minimize the passenger travel cost and enterprise operation cost	short turning offsets, turn back station	ideal point method
Canca et al. [21]	reduce the passenger waiting time	turn back points location, departure and arrival times, short turning offsets	MILP
Bai et al. [24]	minimize the number of trains	turn back point location, headway, turning times	enumeration method
Ghaemi et al. [22]	minimize the delays and canceled services after blockage	short turning stations routes and platform tracks	MILP

In past decades, plenty of models and algorithms have been proposed to solve the train scheduling problem with short turning strategy (Table 1). Furth [17] proposed schedule coordination mode to denote the ratio between the number of short turning train services and full-length train services. They illustrated that the operation pattern with 1:1 schedule coordination mode can provide a better performance than train schedules with full-length train services for an OD matrix obtained from a survey, which means that a short turning service appears between two full-length services can meet passenger demand better. Tirachini et al. [18] developed a short turning model to optimize the frequencies, vehicle sizes, and turnaround stations for the short turning and full-length train services on a bus corridor. Only a single operation period is taken into consideration in this work, while Site and Filippi [19] focused on service patterns over different operation periods for a bus corridor to minimize the cost for users and operators by taking short turning strategy and variable vehicle sizes into consideration. Leffler et al. [20] dealt with the problem to determine where the turnaround station for a short turning service is and when a bus is a short turning service on a single bidirectional bus line in real-time. Canca et al. [21] considered inserting short turning services to deal with a disruption situation such as infrastructure incidences for rapid transit systems with the objective of diminishing the passenger waiting time. Ghaemi et al. [22] also considered a situation with a complete blockage and presented a mixed integer linear programming model to compute the train schedule by adopting short turning strategy. As for the optimization problem of train scheduling with short turning strategy in urban rail transit systems, Wang and Ni [23] set a nonlinear mixed integer programming model with the objective of minimizing the passengers' travel costs, and the operating costs. Bai et al. [24] established a model aiming at minimizing the trains according to the spatial distribution of passengers to optimize and balance train operation scheme. According to the prediction of passenger flow and the amount of trains, Li et al. [25] discussed the different operation patterns with different schedule coordination mode and the choice of the turnaround station based on the Line 1 in Shanghai, which is the first urban rail transit line with short turning strategy. Wang [26] took the passenger flow and trains into consideration with short turning strategy based on a single line in urban rail transit network and establish a multiobjective model to optimize the train schedule by minimizing the number of trains and the variance of the actual passenger capacity and the expected passenger capacity. Hu [27] analyzed the cause of disequilibrium passenger flow and the feasibility of operation pattern with short turning strategy and evaluates the train schedule by trains and the factor of loading.

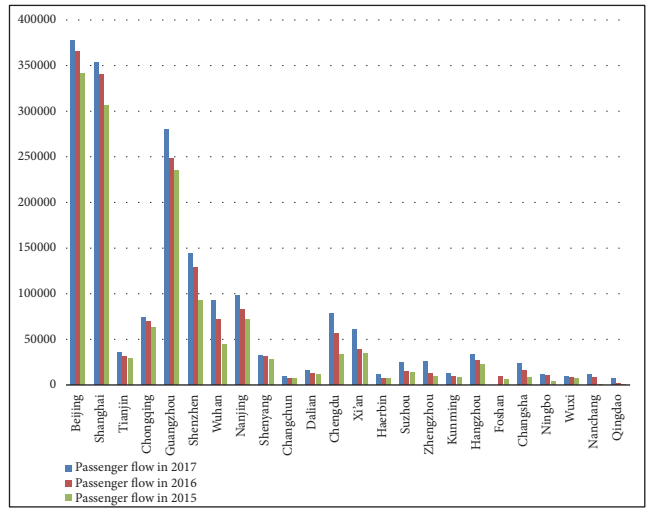


Figure 1: Comparison of passenger flows in 2015, 2016, and 2017 for the metro systems in the cities of China.

Train circulation plan is also important for the train scheduling because the number of the trains is limited. Chang et al. [28] proposed an integrated optimization model for train scheduling and utilization planning problems. Peeters and Kroon [29] developed an integer programming model to find an efficient railway train circulation on a set of interacting train lines by a branch-and-price algorithm. Abbink et al. [30] proposed a model to obtain an optimal allocation of railway train with the objective to minimize the shortages of capacity during the rush hours, where the optimal solution is more effective than the manually planned one. Lai et al. [31] developed a model which covered a lot of characteristics, such as train, inspection, utilization path, and operation. Corts et

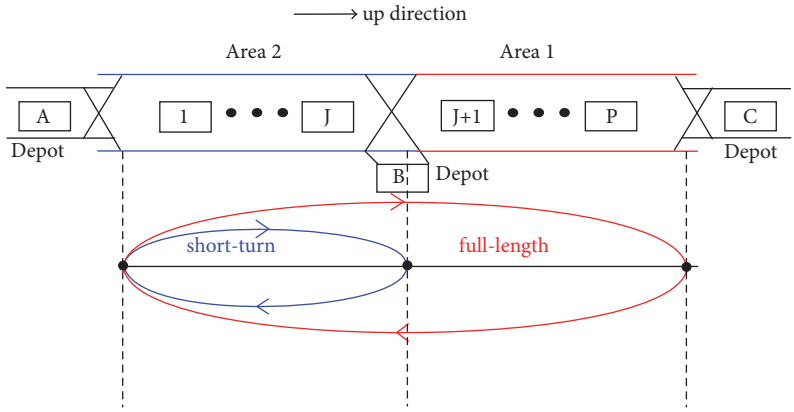


Figure 2: The layout of an urban rail transit line.

al. [32] took deadheading into consideration and proposed a model to set the optimal value of frequencies and capacity of vehicles, combining deadheading and short turning into an integrated fleet management strategy.

The contribution of this paper is a train scheduling model adopting short turning strategy to satisfy the passenger demand for an urban rail transit line with multiple depots. We integrate headway deviation and train circulation plan for all train services to establish a mixed integer linear programming model. We also take departure times, dwell times, and train orders into consideration.

The remainder of the paper is structured as follows. Section 2 introduces train operation and presents the assumptions and notations for the train scheduling model. Section 3 formulates the optimal model with decision variables, objective function, and constraints for the optimization problem. In Section 4, the optimization problem is transformed into an MILP problem through three transformation properties. In Section 5, a case study is implemented based on the data of Beijing Subway line 4. Finally, the conclusion and future work are presented in Section 6.

2. Problem Description

This section illustrates the concepts that are relevant to train operation and the train scheduling with a short turning strategy in an urban rail transit line. The assumptions and notations used for the train scheduling model with short turning strategy are also described.

2.1. Train Operation. The urban rail transit line in this paper is defined as a double-track rail line as shown in Figure 2 and the operation of trains in one direction is not influenced by the other. Train services in different directions are indexed differently. Three depots A, B, and C are connected with station 1, station J, and station P, respectively. The direction from station 1 to station P is referred to as the up direction, and the direction from station P to station 1 is referred to as the down direction. Block sections are divided into two areas, where area 1 includes the block section from station J+1 to station P and area 2 denotes the block section from station 1 to station J. Both short turning train services and full-length train services can run through area 2, but for area 1, only can full-length train services operate in it. In particular, the short turning train services are defined as train services only run in area 2. We divide the operation time of the urban rail transit line into several time intervals according to the passenger traffic flow, which is indexed by k. We introduce i and l to index a train service in the up and down direction with $i \in S^{\text{up}}_{\text{service}} = \{1, 2, \ldots, I^{\text{up}}_{\text{total}}\}$ and $l \in S^{\text{dn}}_{\text{service}} = \{1, 2, \ldots, I^{\text{dn}}_{\text{total}}\}$, where $I^{\text{up}}_{\text{total}}$ and $I^{\text{dn}}_{\text{total}}$ denote the total number of train services for both directions.

$$I^{\text{up}}_{\text{total}} = \sum_{k=1}^{K} I_{\text{up},k}, \tag{1}$$

$$I^{\text{dn}}_{\text{total}} = \sum_{k=1}^{K} I_{\text{dn},k}, \tag{2}$$

where K is the number of time intervals and $I_{\text{up},k}$ and $I_{\text{dn},k}$ denote the number of train services during the k^{th} time interval in the up and down direction, respectively.

There are several situations for train operations on the urban rail transit line. Figure 3(a) is to demonstrate the possible operations for a short turning train service in the up direction that comes out from Depot A. Situation 1 depicts a short turning train service that goes back to Depot B, and in Situation 2, a short turning train service turnaround at the destination station and then the connection train service goes back to Depot A. As for Situation 3, not only does the short turning train service turnaround at the station J, but also the connection train service in the down direction turnaround at its destination station, that is station 1. For full-length train services operating in the urban rail transit line, Figure 3(b) also shows several situations that are similar to Figure 3(a). Accordingly, there are also 6 situations for operation of train services in the down direction.

2.2. Optimal Train Scheduling Problem. As stated by Meng and Zhou [33], to improve the competitive advantages of rail operators, providing punctual and reliable train services is a fundamental goal. A train schedule typically contains

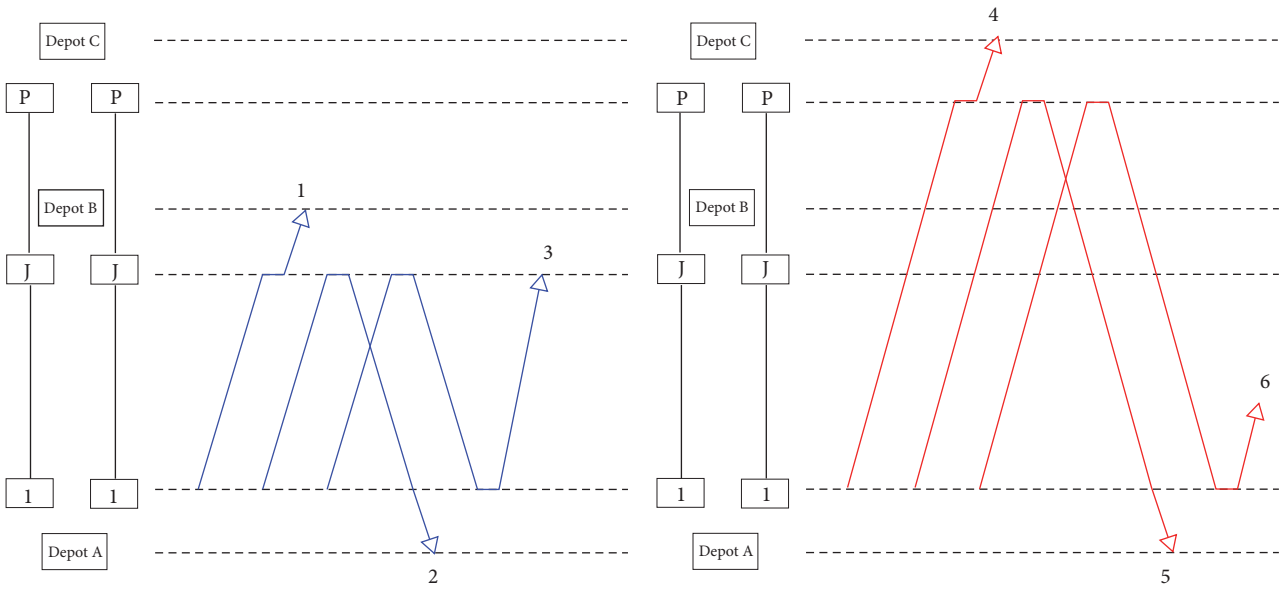


Figure 3: Possible situations for operations of train services in the up direction.

detailed departure times, arrival times, train orders, and train circulation plan for each train service in the urban rail transit line. We aim to minimize the total headway deviation between the actual headway and the predefined headway of departure and arrival times. Meanwhile, the trains utilization can also be optimized to reduce the size of trains and provide more train services.

The main decision variables of the train scheduling optimization problem are

(1) departure times and arrival times of train services at the origin and destination station for both directions;

(2) types of train services, which means that a train service is a short turning one or a full-length one;

(3) connection relationship between train services in the up direction and down direction.

As for constraints, departure and arrival times of each train services, train orders of train services, headways, including headways for train services in area 1 and area 2, and train circulation plan should be included.

Note that we only consider the headways at the origin station and the destination station, because the running times and dwell times are constants that the headway at the interstation should be equal to the headway at the origin station.

It is assumed that 5 train services should be served and there are different passenger demands during area 1 and area 2, where the values are 360s and 180s, respectively. Passengers in area 2 hope to get on train with the headway equaling to 180s, while passengers in area 1 only need a headway about 360s. For train operation pattern without short turning strategy, two patterns can be constructed with a single frequency. Pattern 1 is to satisfy the passenger demand

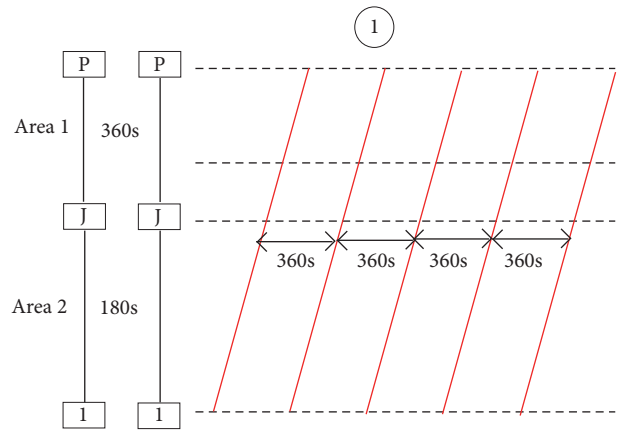


Figure 4: Only passenger demand in area 1 can be satisfied.

in area 1, that the headway of departure time is 360s as shown in Figure 4. It can cause 180s×5 increasing in headway, which is obtained by the sum of headway difference in area 2. Pattern 2 focuses on the passenger demand in area 2 and the headway of departure time is 180s as shown in Figure 5. It also has the same influence with Pattern 1. However, when the short turning train services depart between two consecutive full-length train services like Figure 6, the passenger demand for both areas can be satisfied well.

2.3. Assumptions about the Train Operation. The train scheduling model with short turning strategy mainly consider the departure times of train services for both directions, headways, turnaround operations and train circulation plans. In this paper, we make the following assumptions to simplify the optimal problem.

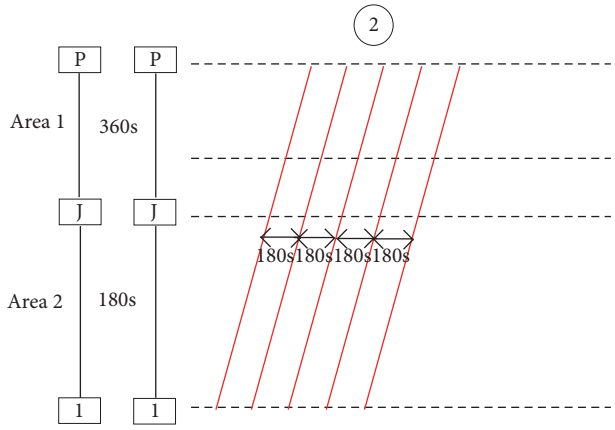


Figure 5: Only passenger demands in area 2 can be satisfied.

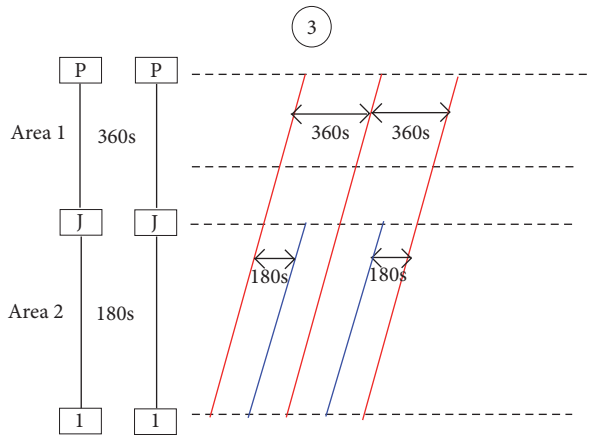


Figure 6: Both passenger demands in areas 1 and 2 can be satisfied.

(i) A1: the station capacity per direction is equal to 1, which means that there is no overtaking at any point in the urban rail transit line.

(ii) A2: there are two short turning train services at most between two full-length train services.

(iii) A3: the running times and dwell times between stations are constants.

(iv) A4: there are enough trains to serve the train services for three depots.

Assumption A1 guarantees the first-in first-out order for train services, which is valid for most urban rail transit lines. Assumption A2 is introduced to reduce the waiting time of passengers whose destinations are not in the short turning area. The fixed running times and dwell times in assumption A3 ensure that departure times of train services can obtain when the headway and the departure time of the first train service are predefined. Assumption A4 means that the depot capacity is not included in this paper.

2.4. Notations. As mentioned in assumption A3, the following parameters should be given: the running times and dwell times for short turning train services and full-length train services in both directions. To describe the detailed modeling methods, subscripts and parameters are shown in Table 2 and decision variables are shown in Table 3.

3. Modeling Formulation

To describe the detailed mathematical formulation, the decision variables are introduced as given in Table 2.

The variables $d_{o,i}^{\mathrm{up}}$ and $d_{o,l}^{\mathrm{dn}}$ are used to directly describe the time-stamps when train service i and l departs from the origin station. $d_{d,i}^{\mathrm{up}}$ and $d_{d,l}^{\mathrm{dn}}$ denote the time-stamps when train service i and l departs from the destination station, respectively. Actually, in this paper, arrival times are calculated by departure times plus the dwell times instead of introducing other variables to denote the arrival times. The departure time at the destination station equals the departure time at the origin station plus the running time and the dwell time at the intermediate stations.

The objective is to minimize the headway deviation and the number of trains as mentioned in Section 2.2. The headway deviation can be divided into two parts; one is the headway deviation for train services in area 1, which is represented by $h_{i,1}^{up}$ and $h_{l,1}^{dn}$. The other is the headway deviation for train services in area 2 and it is denoted by $h_{i,2}^{up}$ and $h_{l,2}^{dn}$. The key problem is to define whether a train service is a short turning one or a full-length one for both directions, which is solved by introducing these two binary variables δ_i and γ_l as shown in the following expressions:

$$\delta_i = \begin{cases} 1, & \text{train service } i \text{ is a short turning train service,} \\ 0, & \text{train service } i \text{ is a full-length train service,} \end{cases} \quad (3)$$

and

$$\gamma_l = \begin{cases} 1, & \text{train service } l \text{ is a short turning train service,} \\ 0, & \text{train service } l \text{ is a full-length train service.} \end{cases} \quad (4)$$

Actually, the key point to calculate $h_{i,1}^{up}$ and $h_{l,1}^{dn}$ is to find the departure times of two consecutive full-length train services. Based on Assumption A2, there are three circumstances C1, C2, and C3 as shown in Figure 7. Assume that a train service i is a full-length train service, which means that $\delta_i = 0$. When there is no short turning train service between two full-length train services as C1, then train service $i+1$ must be a full-length one with $\delta_{i+1} = 0$ and $h_{i,1}^{up}$ must be equal to $(d_{o,i+1}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}})$. If there is only one short turning train service between two full-length train service, which means $\delta_{i+1} = 1$, $\delta_{i+2} = 0$, then $h_{i,1}^{up}$ should be calculated by $(d_{o,i+2}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}})$. As for C3, it is obvious that $h_{i,1}^{up}$ should be calculated by $(d_{o,i+3}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}})$ with $\delta_{i+1} = 1, \delta_{i+2} = 1$, and $\delta_{i+3} = 0$. The circumstances are the same for train services in the down direction.

Table 2: Subscripts and parameters.

symbol	description
i, i'	index for train services in the up direction
l, l'	index for train services in the down direction
o	origin
d	destination
A,B,C	depots
J	total number of stations for short-turning
P	total number of stations for whole line
K	total number of time intervals
up	up operation direction
dn	down operation direction
$I_{\text{total}}^{\text{up}}$	total number of train services in the up direction
$I_{\text{total}}^{\text{dn}}$	total number of train services in the down direction
$S_{\text{service}}^{\text{up}}$	set of train services in the up direction $\{1, 2, ..., I_{\text{total}}^{\text{up}}\}$
$S_{\text{service}}^{\text{dn}}$	set of train services in the down direction $\{1, 2, ..., I_{\text{total}}^{\text{dn}}\}$
r_s^{up}	running time for short turning trains in the up direction
r_f^{up}	running time for full-length trains in the up direction
r_s^{dn}	running time for short turning trains in the down direction
r_f^{dn}	running time for full-length trains in the down direction
τ_s^{up}	dwell time for short turning trains in the up direction
τ_f^{up}	dwell time for full-length trains in the up direction
τ_s^{dn}	dwell time for short turning trains in the down direction
τ_f^{dn}	dwell time for full-length trains in the down direction
$I_{\text{up},k}$	the number of train services in the k^{th} time interval for up direction
$I_{\text{dn},k}$	the number of train services in the k^{th} time interval for down direction
$d_{o,1}^{\text{pre}}$	predetermined departure time for the first train service at the origin station
$d_{o,\text{total}}^{\text{pre}}$	predetermined departure time for the last train service at the origin station
$h_{\min}$	minimal headway
$h_{\max}$	maximal headway
$H_{2,k}^{\text{up}}$	expected headway of train services in area 2 during the k^{th} time interval in the up direction
$H_{1,k}^{\text{up}}$	expected headway of train services in area 1 during the k^{th} time interval in the up direction
$H_{2,k}^{\text{dn}}$	expected headway of train services in area 2 during the k^{th} time interval in the down direction
$H_{1,k}^{\text{dn}}$	expected headway of train services in area 1 during the k^{th} time interval in the down direction
$r_{\min}^{\text{turn}}$	minimal turnaround time
$r_{\max}^{\text{turn}}$	maximal turnaround time
M	a sufficiently large positive number

Table 3: Decision variables.

symbol	description
$d_{o,i}^{\text{up}}$	departure time of train service i in the up direction at the origin
$d_{d,i}^{\text{up}}$	departure time of train service i in the up direction at the destination
$d_{o,l}^{\text{dn}}$	departure time of train service i in the down direction at the origin
$d_{d,l}^{\text{dn}}$	departure time of train service i in the down direction at the destination
δ_i	0-1 binary variables, if δ_i=1, then train service i is a short turning train service
γ_l	0-1 binary variables, if γ_l=1, then train service l is a short turning train service
$\varepsilon(i, l)$	0-1 binary variables, if $\varepsilon(i, l)$=1, then train service l is the connection train of train service i at the destination station
$\eta(l, i)$	0-1 binary variables, if $\eta(l, i)$=1, then train service i is the connection train of train service l at the destination station
α_i^{up}	0-1 binary variables, if train i in the up direction comes out from the depot, α_i^{up}=0, otherwise, α_i^{up}=1
α_l^{dn}	0-1 binary variables, if train l in the down direction comes out from the depot, α_l^{dn}=0, otherwise, α_l^{dn}=1
β_i^{up}	0-1 binary variables, if train i in the up direction goes back to the depot, β_i^{up}=0, otherwise, β_i^{up}=1
β_l^{dn}	0-1 binary variables, if train l in the down direction goes back to the depot, β_l^{dn}=0, otherwise, β_l^{dn}=1

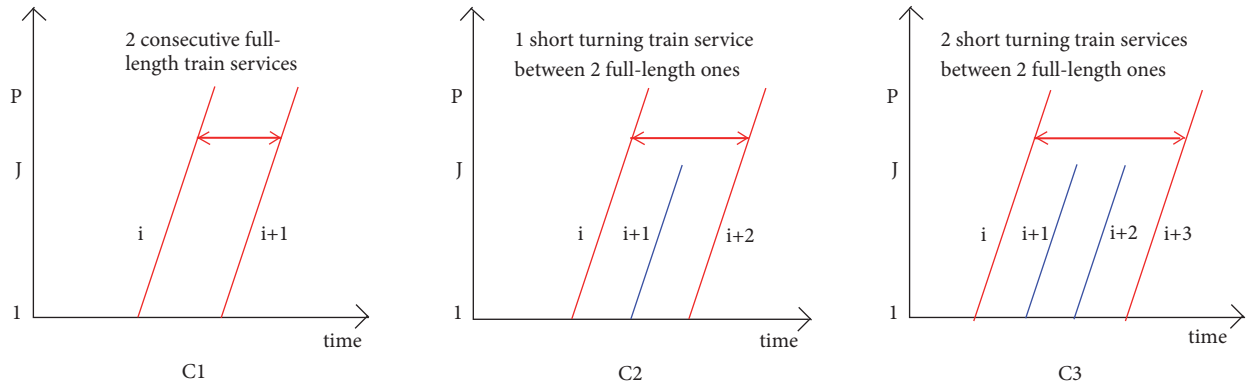


Figure 7: Possible combinations of short turning and full-length train services in the up direction.

Based on the above circumstances, $h_{i,1}^{up}$ and $h_{l,1}^{dn}$ can be calculated as follows:

$$\begin{aligned}
h_{i,1}^{\mathrm{up}} &= (1-\delta_i)(1-\delta_{i+1})\left(d_{o,i+1}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}}\right) \\
&\quad + (1-\delta_i)\delta_{i+1}(1-\delta_{i+2})\left(d_{o,i+2}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}}\right) \\
&\quad + (1-\delta_i)\delta_{i+1}\delta_{i+2}(1-\delta_{i+3})\left(d_{o,i+3}^{\mathrm{up}} - d_{o,i}^{\mathrm{up}}\right) \\
&= \left(d_{o,i+1}^{up} - d_{o,i}^{up}\right) - \delta_i\left(d_{o,i+1}^{up} - d_{o,i}^{up}\right) \\
&\quad - \delta_{i+1}\left(d_{o,i+1}^{up} - d_{o,i}^{up}\right) + \delta_i\delta_{i+1}\left(d_{o,i+1}^{up} - d_{o,i}^{up}\right) \\
&\quad + \delta_{i+1}\left(d_{o,i+2}^{up} - d_{o,i}^{up}\right) - \delta_i\delta_{i+1}\left(d_{o,i+2}^{up} - d_{o,i}^{up}\right) \\
&\quad - \delta_{i+1}\delta_{i+2}\left(d_{o,i+2}^{up} - d_{o,i}^{up}\right) \\
&\quad + \delta_i\delta_{i+1}\delta_{i+2}\left(d_{o,i+2}^{up} - d_{o,i}^{up}\right) \\
&\quad + \delta_{i+1}\delta_{i+2}\left(d_{o,i+3}^{up} - d_{o,i}^{up}\right) \\
&\quad - \delta_i\delta_{i+1}\delta_{i+2}\left(d_{o,i+3}^{up} - d_{o,i}^{up}\right) \\
&\quad - \delta_{i+1}\delta_{i+2}\delta_{i+3}\left(d_{o,i+3}^{up} - d_{o,i}^{up}\right) \\
&\quad + \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}\left(d_{o,i+3}^{up} - d_{o,i}^{up}\right),
\end{aligned} \tag{5}$$

and

$$\begin{aligned}
h_{l,1}^{\mathrm{dn}} &= (1-\gamma_l)(1-\gamma_{l+1})\left(d_{o,l+1}^{\mathrm{dn}} - d_{o,l}^{\mathrm{dn}}\right) \\
&\quad + (1-\gamma_l)\gamma_{l+1}(1-\gamma_{l+2})\left(d_{o,l+2}^{\mathrm{dn}} - d_{o,l}^{\mathrm{dn}}\right) \\
&\quad + (1-\gamma_l)\gamma_{l+1}\gamma_{l+2}(1-\gamma_{l+3})\left(d_{o,l+3}^{\mathrm{dn}} - d_{o,l}^{\mathrm{dn}}\right) \\
&= \left(d_{o,l+1}^{dn} - d_{o,l}^{dn}\right) - \gamma_l\left(d_{o,l+1}^{dn} - d_{o,l}^{dn}\right) \\
&\quad - \gamma_{l+1}\left(d_{o,l+1}^{dn} - d_{o,l}^{dn}\right) + \gamma_l\gamma_{l+1}\left(d_{o,l+1}^{dn} - d_{o,l}^{dn}\right) \\
&\quad + \gamma_{l+1}\left(d_{o,l+2}^{dn} - d_{o,l}^{dn}\right) - \gamma_l\gamma_{l+1}\left(d_{o,l+2}^{dn} - d_{o,l}^{dn}\right) \\
&\quad - \gamma_{l+1}\gamma_{l+2}\left(d_{o,l+2}^{dn} - d_{o,l}^{dn}\right) \\
&\quad + \gamma_l\gamma_{l+1}\gamma_{l+2}\left(d_{o,l+2}^{dn} - d_{o,l}^{dn}\right) \\
&\quad + \gamma_{l+1}\gamma_{l+2}\left(d_{o,l+3}^{dn} - d_{o,l}^{dn}\right) \\
&\quad - \gamma_l\gamma_{l+1}\gamma_{l+2}\left(d_{o,l+3}^{dn} - d_{o,l}^{dn}\right) \\
&\quad - \gamma_{l+1}\gamma_{l+2}\gamma_{l+3}\left(d_{o,l+3}^{dn} - d_{o,l}^{dn}\right) \\
&\quad + \gamma_l\gamma_{l+1}\gamma_{l+2}\gamma_{l+3}\left(d_{o,l+3}^{dn} - d_{o,l}^{dn}\right).
\end{aligned} \tag{6}$$

Accordingly, $h_{i,2}^{up}$ is the headway between departure times of all consecutive train services, which can be calculated as follows:

$$h_{i,2}^{up} = d_{o,i+1}^{up} - d_{o,i}^{up}. \tag{7}$$

As short turning train services and full-length train services in the down direction do not have fixed origins, $h_{l,2}^{dn}$ should be calculated by the following equation:

$$h_{l,2}^{dn} = d_{d,l+1}^{dn} - d_{d,l}^{dn}. \tag{8}$$

As for minimizing the number of trains, it is equivalent to maximize the number of train services that are served by the same train. We introduce two binary variables $\varepsilon(i,l)$ and $\eta(l,i)$ to express whether a train service i and a train service l are served by the same train or not for both directions.

$$\varepsilon(i,l) = \begin{cases} 1, & \text{train service } l \text{ is the connection train service of train service } i, \\ 0, & \text{train service } l \text{ is not the connection train service of train service } i. \end{cases} \tag{9}$$

and

$$\eta(l,i) = \begin{cases} 1, & \text{train service } i \text{ is the connection train service of train service } l, \\ 0, & \text{train service } i \text{ is not the connection train service of train service } l. \end{cases} \tag{10}$$

Based on above descriptions, the optimal problem is now formulated as the following problem:

$$Z = \min w_1 \cdot \frac{f_1}{F_1} + w_2 \cdot \frac{f_2}{F_2} \tag{11}$$

where

$$\begin{aligned} F_1 &= \min f_1 \\ &= \min \sum_{k=1}^{K} \sum_{i=(\sum_{m=0}^{k-1} I_{\mathrm{up},m})+1}^{\sum_{m=1}^{k} I_{\mathrm{up},m}} \left[\left| h_{i,2}^{\mathrm{up}} - H_{1,k}^{\mathrm{up}} \right| + \left| h_{i,1}^{\mathrm{up}} - (1-\delta_i) H_{2,k}^{\mathrm{up}} \right| \right] \\ &\quad + \sum_{k=1}^{K} \sum_{l=(\sum_{n=0}^{k-1} I_{\mathrm{dn},n})+1}^{\sum_{n=1}^{k} I_{\mathrm{dn},n}} \left[\left| h_{l,2}^{\mathrm{dn}} - H_{1,k}^{\mathrm{dn}} \right| + \left| h_{l,1}^{\mathrm{dn}} - (1-\gamma_l) H_{2,k}^{\mathrm{dn}} \right| \right], \end{aligned} \tag{12}$$

$$F_2 = \min f_2 = -\min \sum_{i=1}^{I_{\mathrm{total}}^{\mathrm{up}}} \sum_{l=1}^{I_{\mathrm{total}}^{\mathrm{dn}}} \left[\varepsilon(i,l) + \eta(l,i) \right], \tag{13}$$

and

$$w_1 + w_2 = 1. \tag{14}$$

The formula above should be subject to the following groups constraints:

Group 1 (departure time constraints). To describe the relationship between the departure times of train services at the origin and the destination in different directions, the following two constraints should be satisfied:

$$d_{d,i}^{\mathrm{up}} = d_{o,i}^{\mathrm{up}} + \delta_i \times \left(r_s^{\mathrm{up}} + \tau_s^{\mathrm{up}} \right) + \left(1 - \delta_i \right) \times \left(r_f^{\mathrm{up}} + \tau_f^{\mathrm{up}} \right), \tag{15}$$

$$d_{d,l}^{\mathrm{dn}} = d_{o,l}^{\mathrm{dn}} + \gamma_l \times \left(r_s^{\mathrm{dn}} + \tau_s^{\mathrm{dn}} \right) + \left(1 - \gamma_l \right) \times \left(r_f^{\mathrm{dn}} + \tau_f^{\mathrm{dn}} \right). \tag{16}$$

Group 2 (train order constraints). As shown in Figure 8, the train order of train services in the same direction is fixed in this paper, which means that train service i must depart before train service $i+1$ and train service l must arrive before train service $l+1$.

Group 3 (headway constraints). The headway for departure times of two consecutive train services must be larger than the minimal headway and smaller than the maximal headway.

Headway of train services in area 1 is

$$h_{\min} - M \times \left(1 - \delta_i \right) \le h_{i,1}^{\mathrm{up}} \le h_{\max} + M \times \left(1 - \delta_i \right), \tag{17}$$

and

$$h_{\min} - M \times \left(1 - \gamma_l \right) \le h_{l,1}^{\mathrm{dn}} \le h_{\max} - M \times \left(1 - \gamma_l \right), \tag{18}$$

where M is a sufficiently large positive number. If train service i or l is a full-length one, that is $\delta_i = 0$ or $\gamma_l = 0$, then the headway of train services in area 1 is larger than the minimum headway and smaller than the maximum headway. If i or l is a short turning train service, then (17) and (18) can be satisfied automatically.

Headway of train services in area 2 is

$$h_{\min} \le h_{i,2}^{\mathrm{up}} \le h_{\max}, \tag{19}$$

and

$$h_{\min} \le h_{l,2}^{dn} \le h_{\max}. \tag{20}$$

Group 4 (train circulation plan constraints). According to the layout of the urban rail transit line, a train service may come out from the depot or be served by a train that just finishes a previous train service. After arriving at the destination station, a train service could go back to the depot or just turnaround and the same train serves another train service. α_i^{up}, α_l^{dn}, β_i^{up}, and β_l^{dn} are introduced to describe the above operation of train services for both directions.

$$\alpha_i^{\mathrm{up}} = \begin{cases} 1, & \text{train service } i \text{ does not come out from the depot,} \\ 0, & \text{train service } i \text{ comes out from the depot.} \end{cases} \tag{21}$$

$$\alpha_l^{\mathrm{dn}} = \begin{cases} 1, & \text{train service } l \text{ does not come out from the depot,} \\ 0, & \text{train service } l \text{ comes out from the depot.} \end{cases} \tag{22}$$

$$\beta_i^{\mathrm{up}} = \begin{cases} 1, & \text{train service } i \text{ does not go back to the depot,} \\ 0, & \text{train service } i \text{ goes back to the depot.} \end{cases} \tag{23}$$

$$\beta_l^{\mathrm{dn}} = \begin{cases} 1, & \text{train service } l \text{ does not go back to the depot,} \\ 0, & \text{train service } l \text{ goes back to the depot.} \end{cases} \tag{24}$$

Based on the above definition of operation for train services, the following constraints should be satisfied:

$$\alpha_i^{\mathrm{up}} = \sum_{l \in S_{\mathrm{service}}^{\mathrm{dn}}} \eta(l,i), \tag{25}$$

$$\alpha_l^{\mathrm{dn}} = \sum_{i \in S_{\mathrm{service}}^{\mathrm{up}}} \varepsilon(i,l), \tag{26}$$

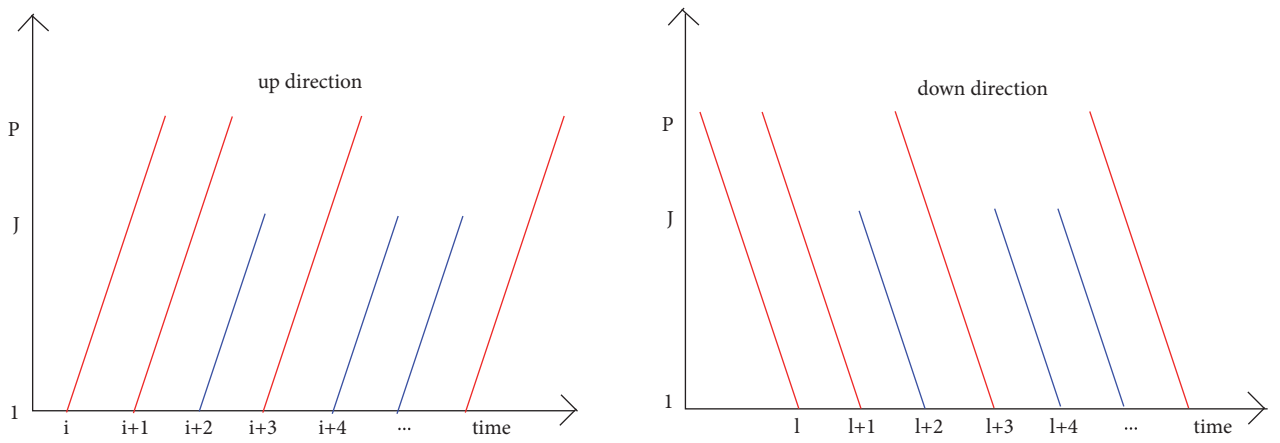


Figure 8: The order of train services in the up and down directions.

$$\beta_i^{\mathrm{up}} = \sum_{l \in S_{\mathrm{service}}^{\mathrm{dn}}} \varepsilon(i,l), \tag{27}$$

$$\beta_l^{\mathrm{dn}} = \sum_{i \in S_{\mathrm{service}}^{\mathrm{up}}} \eta(l,i), \tag{28}$$

where (25) means that if a train service i in the up direction does not come out from the depot, that is $\alpha_i^{\mathrm{up}} = 1$, then there must be a train service l in the down direction that turnarounds at its destination station and train service i is served by the same train, which means that $\eta(l,i) = 1$ and $\beta_l^{\mathrm{dn}} = 1$. Accordingly, (26) denotes that if a train service l does not come out from the depot with $\alpha_l^{\mathrm{dn}} = 1$, then there must be train service i turnarounds and the train which serves the train service i will serve the train service l; that is, $\varepsilon(i,l) = 1$ and $\beta_i^{\mathrm{up}} = 1$. As for (27), when a train service i does not go back to the depot, which means that $\beta_i^{\mathrm{up}} = 1$, then the train must turnaround and serve another train service l with $\varepsilon(i,l) = 1$ and $\alpha_l^{\mathrm{dn}} = 1$. Similarly, if a train service l does not go back to the depot with $\beta_l^{\mathrm{dn}} = 1$, then a train service i must be served by the same train with train service l; that is, $\eta(l,i) = 1$ and $\alpha_i^{\mathrm{up}} = 1$.

In addition, the following constraint must be involved to ensure that a short turning train service in the up direction will not connect with a full-length train service in the down direction or a full-length train service in the up direction will not connect with a short turning train service in the down direction:

$$\varepsilon(i,l) \leq 1 + 2\delta_i\gamma_l - \delta_i - \gamma_l \tag{29}$$

Furthermore, if a train service turnarounds at the destination station, then the actual turnaround time must be larger than the minimal turnaround time and smaller than the maximal turnaround time, which can be formulated as follows:

$$\begin{aligned} r_{\min}^{\mathrm{turn}} - M \times (1 - \varepsilon(i,l)) &\leqslant d_{o,l}^{\mathrm{dn}} - d_{d,i}^{\mathrm{up}} \\ &\leqslant r_{\max}^{\mathrm{turn}} + M \times (1 - \varepsilon(i,l)), \end{aligned} \tag{30}$$

$$\begin{aligned} r_{\min}^{\mathrm{turn}} - M \times (1 - \eta(l,i)) &\leqslant d_{o,i}^{\mathrm{up}} - d_{d,l}^{\mathrm{dn}} \\ &\leqslant r_{\max}^{\mathrm{turn}} + M \times (1 - \eta(l,i)). \end{aligned} \tag{31}$$

Note that M is a big value in this paper.

4. Solution Approach: MILP Approach

The mixed integer nonlinear model can be transformed into a mixed integer linear programming MILP problem based on several transformation properties. We introduce three properties to linearize the nonlinear section in the model proposed above according to [34].

(i) Property I is described as follows: a new auxiliary real-valued variable z is introduced to replace $\delta \times \tilde{f}$, where δ is a logical variable and $\tilde{f}$ is a real-value variable. When $\delta = 0 \Longrightarrow z = 0$, and $\delta = 1 \Longrightarrow z = \tilde{f}$, then $z = \delta \times \tilde{f}$ is equivalent to

$$\begin{aligned} z &\leq \tilde{f}_{max} \times \delta, \\ z &\geq \tilde{f}_{min} \times \delta, \\ z &\leq \tilde{f} - \tilde{f}_{min} \times (1 - \delta), \\ z &\geq \tilde{f} - \tilde{f}_{max} \times (1 - \delta), \end{aligned} \tag{32}$$

where $\tilde{f}_{max}$ is the maximal value of $\tilde{f}$ and $\tilde{f}_{min}$ is the minimal value of $\tilde{f}$.

(ii) Property *II*: there are two logical variables multiplied, i.e., $\delta_1 \times \delta_2$. It can be replaced by an auxiliary logical variable δ_3, which is equivalent to

$$\begin{aligned} -\delta_1 + \delta_3 &\le 0, \\ -\delta_2 + \delta_3 &\le 0, \\ \delta_1 + \delta_2 - \delta_3 &\le 1, \end{aligned} \tag{33}$$

where $\delta_1 = 0, \delta_2 = 0 \Longrightarrow \delta_3 = 0$; $\delta_1 = 0, \delta_2 = 1$ or $\delta_1 = 1, \delta_2 = 0 \Longrightarrow \delta_3 = 0$; and $\delta_1 = 1, \delta_2 = 1 \Longrightarrow \delta_3 = 1$.

(iii) Property *III*: for an optimal problem with the objective to minimize f_{obj}, where $f_{obj} = |\widetilde{f}|$, we introduce a new auxiliary real-valued variable $\widetilde{w}$ to replace f_{obj} and transform $|\widetilde{f}|$ as follows:

$$\begin{aligned} \widetilde{w} &\ge \widetilde{f} \\ \widetilde{w} &\ge -\widetilde{f}. \end{aligned} \tag{34}$$

Taking nonlinear terms in constraints that related to train services in the up direction for example, the following terms in (5) should be transformed according to property *I*:

We introduce $z_{1,i} = \delta_i(d^{up}_{o,i+1} - d^{up}_{o,i})$, $z_{2,i} = \delta_{i+1}(d^{up}_{o,i+1} - d^{up}_{o,i})$, $z_{3,i} = \delta_i\delta_{i+1}(d^{up}_{o,i+1} - d^{up}_{o,i})$, $z_{4,i} = \delta_{i+1}(d^{up}_{o,i+2} - d^{up}_{o,i})$, $z_{5,i} = \delta_i\delta_{i+1}(d^{up}_{o,i+2} - d^{up}_{o,i})$, $z_{6,i} = \delta_{i+1}\delta_{i+2}(d^{up}_{o,i+2} - d^{up}_{o,i})$, $z_{7,i} = \delta_i\delta_{i+1}\delta_{i+2}(d^{up}_{o,i+2} - d^{up}_{o,i})$, $z_{8,i} = \delta_{i+1}\delta_{i+2}(d^{up}_{o,i+3} - d^{up}_{o,i})$, $z_{9,i} = \delta_i\delta_{i+1}\delta_{i+2}(d^{up}_{o,i+3} - d^{up}_{o,i})$, $z_{10,i} = \delta_{i+1}\delta_{i+2}\delta_{i+3}(d^{up}_{o,i+3} - d^{up}_{o,i})$ and $z_{11,i} = \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}(d^{up}_{o,i+3} - d^{up}_{o,i})$

$$\begin{aligned} z_{1,i} &\le h_{\max} \cdot \delta_i, \\ z_{1,i} &\ge h_{\min} \cdot \delta_i, \\ z_{1,i} &\le \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i), \\ z_{1,i} &\ge \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i). \end{aligned} \tag{35}$$

$$\begin{aligned} z_{2,i} &\le h_{\max} \cdot \delta_{i+1}, \\ z_{2,i} &\ge h_{\min} \cdot \delta_{i+1}, \\ z_{2,i} &\le \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_{i+1}), \\ z_{2,i} &\ge \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_{i+1}). \end{aligned} \tag{36}$$

$$\begin{aligned} z_{3,i} &\le h_{\max} \cdot \delta_i\delta_{i+1}, \\ z_{3,i} &\ge h_{\min} \cdot \delta_i\delta_{i+1}, \\ z_{3,i} &\le \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i\delta_{i+1}), \\ z_{3,i} &\ge \left(d^{\text{up}}_{o,i+1} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i\delta_{i+1}). \end{aligned} \tag{37}$$

$$\begin{aligned} z_{4,i} &\le h_{\max} \cdot \delta_{i+1}, \\ z_{4,i} &\ge h_{\min} \cdot \delta_{i+1}, \\ z_{4,i} &\le \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_{i+1}), \\ z_{4,i} &\ge \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_{i+1}). \end{aligned} \tag{38}$$

$$\begin{aligned} z_{5,i} &\le h_{\max} \cdot \delta_i\delta_{i+1}, \\ z_{5,i} &\ge h_{\min} \cdot \delta_i\delta_{i+1}, \\ z_{5,i} &\le \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i\delta_{i+1}), \\ z_{5,i} &\ge \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i\delta_{i+1}). \end{aligned} \tag{39}$$

$$\begin{aligned} z_{6,i} &\le h_{\max} \cdot \delta_{i+1}\delta_{i+2}, \\ z_{6,i} &\ge h_{\min} \cdot \delta_{i+1}\delta_{i+2}, \\ z_{6,i} &\le \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_{i+1}\delta_{i+2}), \\ z_{6,i} &\ge \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_{i+1}\delta_{i+2}). \end{aligned} \tag{40}$$

$$\begin{aligned} z_{7,i} &\le h_{\max} \cdot \delta_i\delta_{i+1}\delta_{i+2}, \\ z_{7,i} &\ge h_{\min} \cdot \delta_i\delta_{i+1}\delta_{i+2}, \\ z_{7,i} &\le \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}), \\ z_{7,i} &\ge \left(d^{\text{up}}_{o,i+2} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}). \end{aligned} \tag{41}$$

$$\begin{aligned} z_{8,i} &\le h_{\max} \cdot \delta_{i+1}\delta_{i+2}, \\ z_{8,i} &\ge h_{\min} \cdot \delta_{i+1}\delta_{i+2}, \\ z_{8,i} &\le \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_{i+1}\delta_{i+2}), \\ z_{8,i} &\ge \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_{i+1}\delta_{i+2}). \end{aligned} \tag{42}$$

$$\begin{aligned} z_{9,i} &\le h_{\max} \cdot \delta_i\delta_{i+1}\delta_{i+2}, \\ z_{9,i} &\ge h_{\min} \cdot \delta_i\delta_{i+1}\delta_{i+2}, \\ z_{9,i} &\le \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}), \\ z_{9,i} &\ge \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}). \end{aligned} \tag{43}$$

$$\begin{aligned} z_{10,i} &\le h_{\max} \cdot \delta_{i+1}\delta_{i+2}\delta_{i+3}, \\ z_{10,i} &\ge h_{\min} \cdot \delta_{i+1}\delta_{i+2}\delta_{i+3}, \\ z_{10,i} &\le \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_{i+1}\delta_{i+2}\delta_{i+3}), \\ z_{10,i} &\ge \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_{i+1}\delta_{i+2}\delta_{i+3}). \end{aligned} \tag{44}$$

$$\begin{aligned} z_{11,i} &\le h_{\max} \cdot \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}, \\ z_{11,i} &\ge h_{\min} \cdot \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}, \\ z_{11,i} &\le \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\min} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}), \\ z_{11,i} &\ge \left(d^{\text{up}}_{o,i+3} - d^{\text{up}}_{o,i}\right) - h_{\max} \cdot (1 - \delta_i\delta_{i+1}\delta_{i+2}\delta_{i+3}). \end{aligned} \tag{45}$$

Table 4: Operation information of line 4 in Beijing.

up direction			down direction		
Station	Running time[s]	Dwell time[s]	Station	Running time[s]	Dwell time[s]
AnheqiaoNorth	3750	30	Tiangongyuan	2405	35
Xingong	2430	30	Xingong	3762	35
Tiangongyuan	-	30	AnheqiaoNorth	-	30

Table 5: Input parameters for case study-A.

Time interval	I_{up}	I_{dn}	$H^{\text{up}}_{2,k}[s]$	$H^{\text{up}}_{1,k}[s]$	$H^{\text{dn}}_{2,k}[s]$	$H^{\text{dn}}_{1,k}[s]$
5:00-7:00	13	13	600	600	600	600
7:00-9:00	24	24	600	300	600	300

Note that some terms in (35)-(45) also should be linearized according to property *II*. $\delta_i \cdot \delta_{i+1}, \delta_i \cdot \delta_{i+1} \cdot \delta_{i+2}, \delta_i \cdot \delta_{i+1} \cdot \delta_{i+2} \cdot \delta_{i+3}$ can be translated into the following constraints:

$$\begin{aligned} -\delta_i + W_{1,i} &\le 0, \\ -\delta_{i+1} + W_{1,i} &\le 0, \\ \delta_i + \delta_{i+1} - W_{1,i} &\le 1. \end{aligned} \tag{46}$$

$$\begin{aligned} -W_{1,i} + W_{2,i} &\le 0, \\ -\delta_{i+2} + W_{2,i} &\le 0, \\ W_{1,i} + \delta_{i+2} - W_{2,i} &\le 1. \end{aligned} \tag{47}$$

$$\begin{aligned} -W_{1,i} + W_{3,i} &\le 0, \\ -W_{1,i+1} + W_{3,i} &\le 0, \\ W_{1,i} + W_{1,i+1} - W_{3,i} &\le 1, \end{aligned} \tag{48}$$

where $W_{1,i} = \delta_i \cdot \delta_{i+1}$, $W_{2,i} = W_{1,i} \cdot \delta_{i+2}$, and $W_{3,i} = F_{1,i} \cdot W_{1,i+1}$. All these terms should be replaced according to the previous two properties. As for the absolute value terms in the objective function, they are equivalent to the following constraints:

$$\begin{aligned} x_{i,k} &\ge h^{\text{up}}_{2,i} - H^{\text{up}}_{2,k}, \\ x_{i,k} &\ge -\left[h^{\text{up}}_{2,i} - H^{\text{up}}_{2,k}\right]. \end{aligned} \tag{49}$$

$$\begin{aligned} y_{l,k} &\ge h^{\text{dn}}_{2,l} - H^{\text{dn}}_{2,k}, \\ y_{l,k} &\ge -\left[h^{\text{dn}}_{2,l} - H^{\text{dn}}_{2,k}\right]. \end{aligned} \tag{50}$$

$$\begin{aligned} u_{i,k} &\ge h^{\text{up}}_{i,1} - (1-\delta_i) H^{\text{up}}_{1,k}, \\ u_{i,k} &\ge -\left[h^{\text{up}}_{i,1} - (1-\delta_i) H^{\text{up}}_{1,k}\right]. \end{aligned} \tag{51}$$

$$\begin{aligned} v_{l,k} &\ge h^{\text{dn}}_{l,1} - (1-\gamma_l) H^{\text{dn}}_{1,k}, \\ v_{l,k} &\ge -\left[h^{\text{dn}}_{l,1} - (1-\gamma_l) H^{\text{dn}}_{1,k}\right]. \end{aligned} \tag{52}$$

where $x_{i,k} = |h^{\text{up}}_{2,i} - H^{\text{up}}_{2,k}|$, $y_{l,k} = |h^{\text{dn}}_{2,l} - H^{\text{dn}}_{2,k}|$, $u_{i,k} = |h^{\text{up}}_{i,1} - (1-\delta_i)H^{\text{up}}_{1,k}|$, and $v_{l,k} = |h^{\text{dn}}_{l,1} - (1-\gamma_l)H^{\text{dn}}_{1,k}|$. The corresponding term f_1 in the objective function will be reformulated as the following equation:

$$\begin{aligned} f_1 = &\sum_{k=1}^{K} \sum_{i=(\sum_{m=0}^{k-1} I_{\text{up},m})+1}^{\sum_{m=1}^{k} I_{\text{up},m}} \left(x_{i,k} + u_{i,k}\right) \\ &+ \sum_{k=1}^{K} \sum_{l=(\sum_{n=0}^{k-1} I_{\text{dn},n})+1}^{\sum_{n=1}^{k} I_{\text{dn},n}} \left(y_{l,k} + v_{l,k}\right). \end{aligned} \tag{53}$$

The MILP problem can be solved by several existing commercial and free solvers, such as CPLEX, Xpress-MP, and GLPK.

5. Case Study

In order to assess the performance of the proposed MILP approach, we use two case studies to compare two operation patterns with or without short turning strategy based on the data of Beijing Subway line 4, where the layout is shown in Figure 9. Three depots connect with Anheqiao North Station, Xingong Station, and Tiantongyuan Station, respectively. Area 1 includes the section between Xingong Station and Tiantongyuan Station. Section from Xingong Station to Anheqiao North Station is depicted as area 2. In addition, the parameters of train operation in line 4 also are given in Table 4. The MILP problem is solved by the commercial solver CPLEX with version 12.6 on a 64-bit windows 10 platform.

Note that the value of dwell times at stations inside short turning circle in the up direction is 875s and outside short turning circle in the up direction is 1175s. For train services in the down direction, the dwell time at stations in area 1 and area 2 are 1190s and 900s, respectively.

5.1. Case Study-A. In case study-A, the input parameters are given in Table 5. The total number of train services for both directions is 37, respectively. In particular, the maximal and minimal turnaround times are 720s and 120s. As for the headway of departure times must larger than 120s and smaller than 660s. Note that the departure time of the first train service in both directions is 5:00, and the departure time of the last train service for both directions is 9:00.

Figure 9: The layout of the Beijing Subway line 4.

In this case study, a small scale of simulation is conducted. We assume the operation time of the urban rail transit line is 5:00 to 9:00 and the time period is divided into 2 time intervals as shown in Table 5. The urban rail transit line serves different operation pattern during two time intervals. During 5:00 to 7:00, the ratio of headway in area 2 and headway in area 1 is 1, which means that there would be only full-length train services operated. From 7:00 to 9:00, headway in area 2 is twice headway in area 1; then there would be a short turning train service served between two full-length train services.

Firstly, we calculate the value of F_1 and F_2 by (12) and (13), which also should meet all constraints. We obtain the values of F_1 and F_2 which are 894 and -46, respectively. There are 15377 constrains and 4757 variables. Then the comparison of optimization results is shown in Table 6, where we set several different values for weight coefficients w_1 and w_2. Computing times, objective function values, and gap are considered in this result. The result of computing time shows a decreasing tendency where the value of w_1 increases and w_2 decreases at the same time. Also the objective function value is on the rise during this process. The gap can reach the maximum value among these results when the weight of headway deviation equals the weight of train circulation plan. All the gaps are not too large because there only are 74 train services in total that should be served. Furthermore, the trains used in both directions are 16 and 14, which is smaller than trains used

Table 6: Comparison with different value of weight coefficients for case study-A.

w_1	w_2	Computing time[s]	Objective function value	Gap
0.1	0.9	1.76	-0.657	0
0.3	0.7	1.55	-0.059	0.58%
0.5	0.5	1.31	0.536	2.03%
0.7	0.3	1.29	1.124	0.41%
0.9	0.1	1.25	1.713	0

Table 7: Input parameters for case study-B.

Time interval	I_{up}	I_{dn}	$H_{2,k}^{\text{up}}[s]$	$H_{1,k}^{\text{up}}[s]$	$H_{2,k}^{\text{dn}}[s]$	$H_{1,k}^{\text{dn}}[s]$
5:00-7:00	13	13	600	600	600	600
7:00-9:00	60	60	360	120	360	120
9:00-13:00	60	60	480	240	480	240
13:00-15:00	12	12	600	600	600	600

without short turning strategy. The optimal train schedule obtained when $w_1 = 0.1$ and $w_2 = 0.9$ is shown as Figure 10, where the lines with same color mean that the train services are served by the same train.

5.2. Case Study- B. In case study-B, the input parameters are given in Table 7. The total number of train services for both directions is 145, respectively. The optimal train schedule with short turning strategy obtained by CPLEX solver is demonstrated in Figure 11; the line with same color means that these train services are served by the same train. The schedule coordination mode during each time interval is 1:0, 1:2, 1:1, and 1:0, where the former value is frequency for full-length train services and the latter one is frequency for short turning train services. The calculation results of values of F_1 and F_2 are 1680 and -223, respectively. In this case study, we set the value of w_1 and w_2 equal to 0.9 and 0.1 to use less computing time. The numbers of constraints and variables are 171269 and 50117. During the whole operation period, 43 trains come out from Depot A to serve the train services for the service pattern with short turning strategy, and 26 trains come out from Depot B or Depot C to serve the train service in the down direction.

To make a comparison, optimal train schedule without short turning strategy also is presented as shown in Figure 12. Under this situation, the same passenger demand should be satisfied, which means that values of the predefined headway are identical to the corresponding values for operation pattern with short turning strategy. To serve 145 train services in the up direction, 44 trains come out from Depot A. Meanwhile, 43 trains come out from Depot B or Depot C to serve the 145 train services in the down direction. According to the results, optimal train schedule with short turning strategy could provide a better train circulation plan with 20.69% reduction of trains in use. In particular, the number of trains in theory can be calculated by the following equation:

$$N_{train} = \frac{t_{cycle}}{h} \tag{54}$$

where N_{train} is the number of trains that can be computed theoretically, t_{cycle} is the cycling time of the train service, and h is the headway [15]. Based on (54), the number of trains to satisfy the passenger demand is 122.7 in single direction, which means the optimal train schedule can reduce about $[(122.7 - 43 - 26) + 122.7 - 44 - 43)]/122.7 \times 2 = 36.43\%$ of trains.

Furthermore, the objective value obtained by optimal train schedule with short turning strategy is $(2.634 - 0.801)/2.634 * 100\% = 69.59\%$ better than the train schedule without short turning strategy. Because the number of train services that should be served is calculated by dividing the length of the time interval by predefined headway in area 2, which means that satisfying the passenger demand inside short turning circle is the basis goal. To finish the whole operation period, train schedule without short turning train services can only meet the passenger demand in area 2 well, while load factor in area 1 is insufficient. For instance, during the second time interval, the predefined headways in area 1 and area 2 are 360s and 120s, respectively. The actual headway of two consecutive train services is close to 120s, which cannot make full use of the trains as passenger demand in area 1 does not match the transport capacity. As a result, there is a huge increase in headway deviation.

Figures 13 and 14 demonstrate the load factor of a train service at two certain stations in the up direction. Note that Qingyuan Road Station is a station in area 1, and East gate of Peking University Station is a station in area 2, which means that Qingyuan Road Station is not included in the short turning circle, while East gate of Peking University Station is. As shown in Figure 13, the values for train operation pattern with short turning strategy are larger than the ones without short turning strategy at Qingyuan Road Station. That is caused by the different passenger demand in area 1 and area 2, which is related to the predefined headway. Actually, the number of train services that should be served is calculated by the following equation [35]:

$$H_k = \frac{t_{k+1} - t_k}{I_k} \tag{55}$$

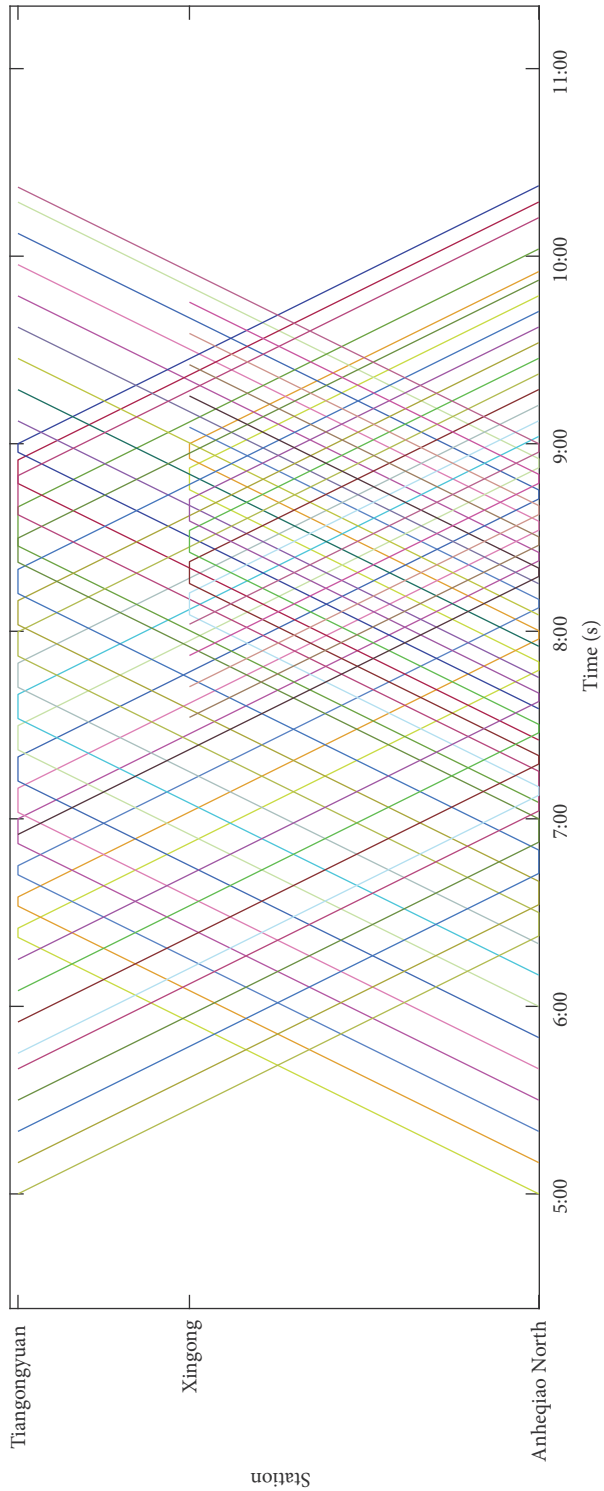


Figure 10: Optimal train schedule with short turning strategy for case study-A.

where H_k is the predefined headway for the kth time interval, t_k and t_k are the starting time and ending time of the kth time interval, and I_k is the number of train services that should be operated during the kth time interval. When trains are running without short turning strategy and the same number of train services that should be operated, the predefined headway is the same as the predefined headway inside short turning circle. That means the actual headway between two consecutive train services is closed to the headway in area 2. This leads to an excess of train services running in area 1 and making the load factor of stations in area 1 small. In addition, the value of the load factor is less than 0.3, which resulted

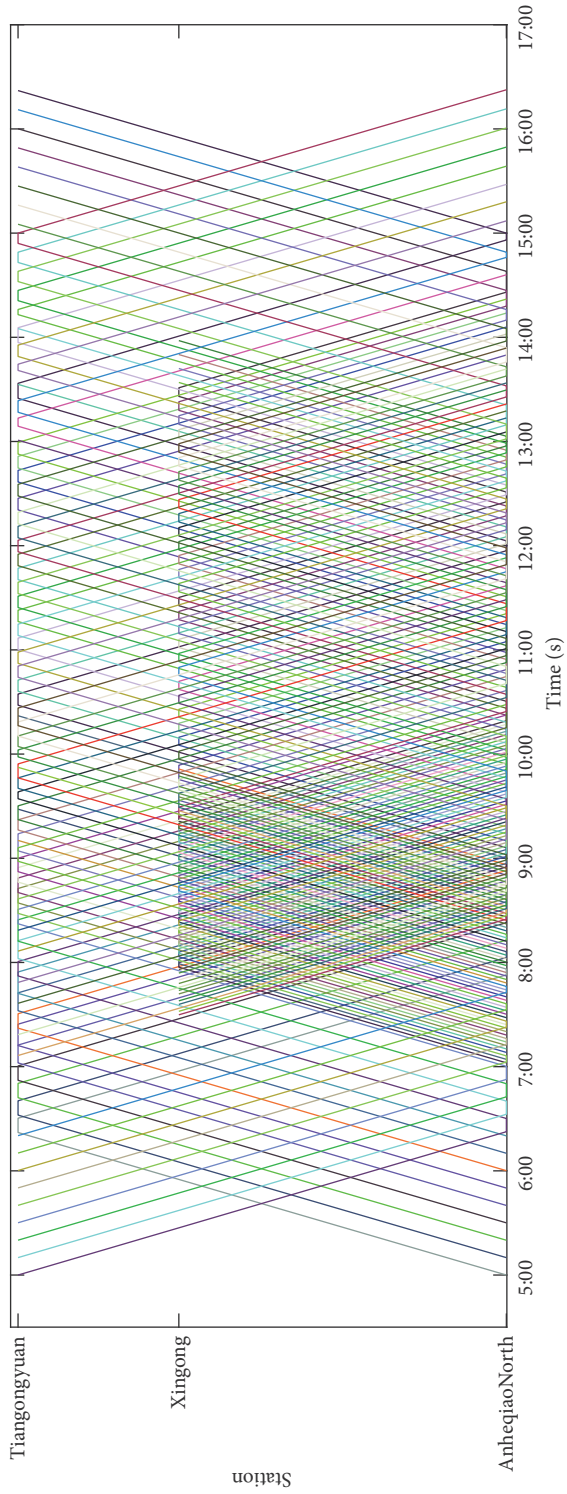


Figure 11: Optimal train schedule with short turning strategy for case study-B.

from the small quantity of passenger flow at this station. The passenger demand is so small that there is no need to establish the station. As for load factor at East gate of Peking University Station in Figure 14, if service pattern is operated with short turning strategy, the load factor is smaller but closer to 1 compared with the load factor of a train service without short turning strategy at East gate of Peking University Station during the early peak hour period. It indicates that the short turning strategy can reduce the crowdedness at stations in area 2. The load factor between 8:00 and 9:00 is near to 1.4, an

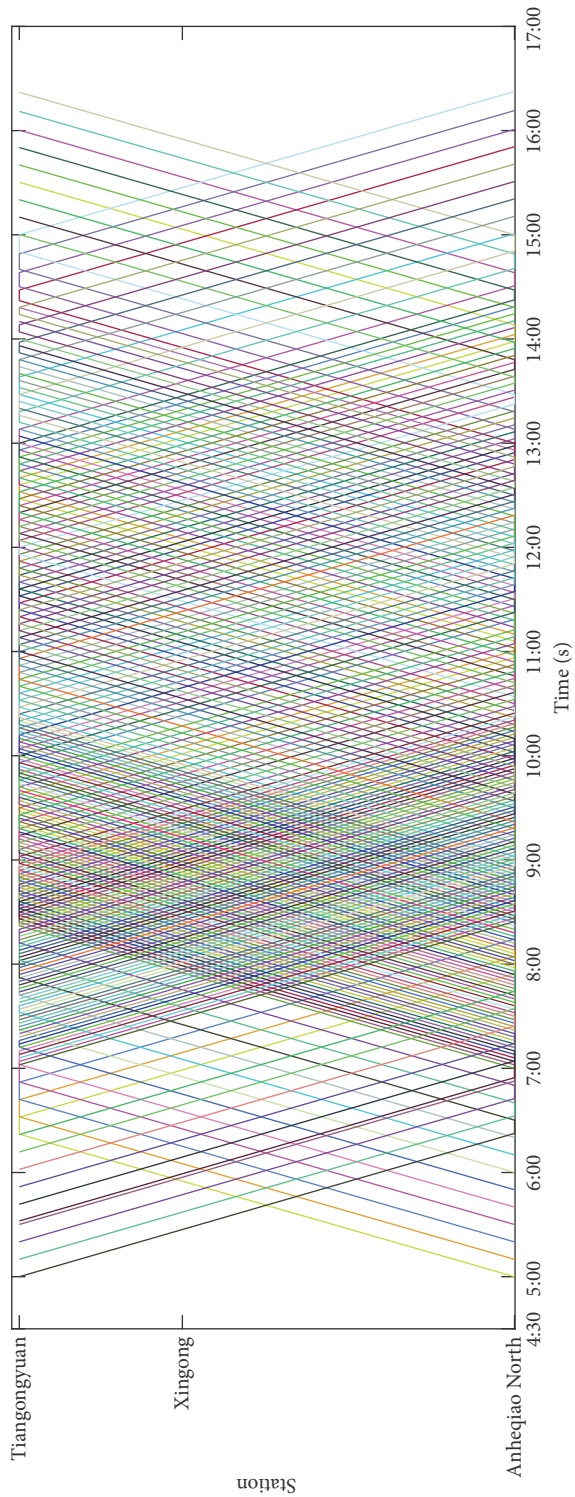


Figure 12: Optimal train schedule without short turning strategy for case study-B.

overcrowd circumstance that is abnormal. The reason for the phenomenon is that the passenger flow for line 4 is so large that only adopting the short turning strategy cannot satisfy the huge passenger demand. Other measures can be taken into consideration in future.

6. Conclusions and Future Work

In this paper, we have developed a new model for train scheduling optimization problem for an urban rail transit line with multiple depots. In this model, short turning strategy

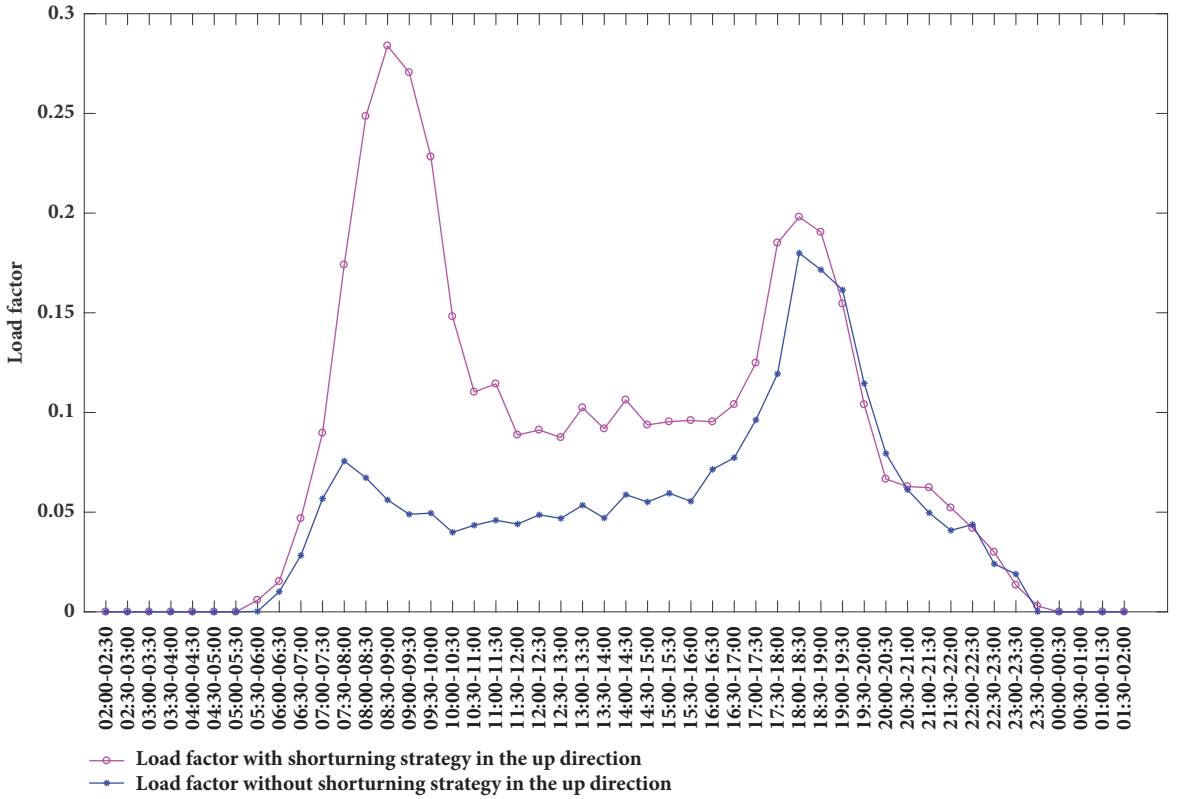


Figure 13: Load factor of a train service at Qingyuan Road Station.

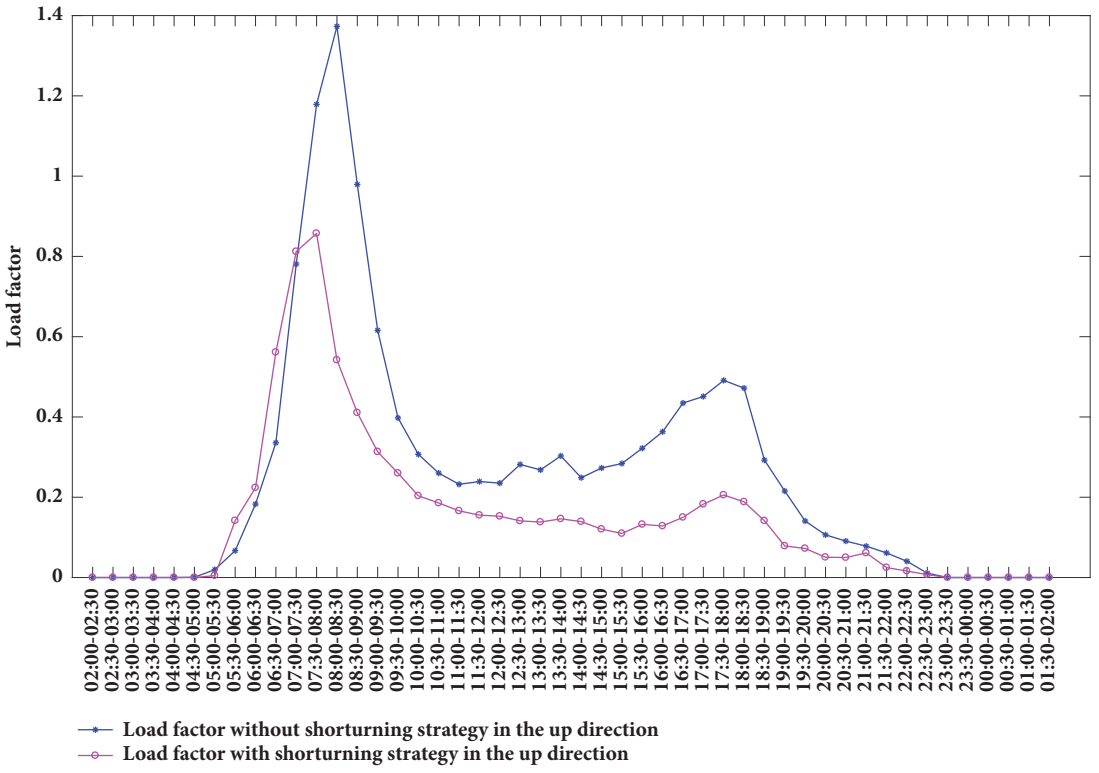


Figure 14: Load factor of a train service at east gate of Peking University Station.

is introduced to satisfy the unbalanced passenger demand. Groups of practical and typical constraints such as departure times, dwell times, headways, train orders, and train circulation plan are taken into consideration. Several transformation properties are involved to transform the model into a mixed integer linear programming one, which can be solved by existing solvers.

Simulation experiment is carried out based on the data of Beijing Subway line 4 to evaluate the performance of the proposed model and approach. Our results show that operation pattern with short turning strategy can acquire a better train schedule and train circulation plan.

For the future research steps, we will extend it by taking more realistic constraints into consideration, such as depot equilibrium and stochastic passenger flow. In particular, the large scale of train services such as train schedule in one day could be implemented for extended problem. Furthermore, other types layout of urban rail transit lines could be considered.

Conflicts of Interest

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

Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 61503020), the Beijing Natural Science Foundation (no. L171008), the Fundamental Research Funds for the Central Universities (no. 2017JBM076), and the Fundamental Research Funds for the Central Universities (no. 2017YJS004). The authors acknowledge the help of Yimeng Luo and Xingdong Zhao.

References

[1] X. Tian, *An optimization of train scheduling for urban rail transits under time-dependent conditions [Ph.D. thesis]*, Lanzhou Jiaotong University, 2013.

[2] L. Wang, *Research of urban rail transit scheduling optimization based on the demanding of passenger flow [Ph.D. thesis]*, Southeast University, 2015.

[3] X. Li and X. Zhang, "Optimizing rail program under the maximize satisfaction," *Technology and Economy in Areas of Communications*, pp. 69–72, 2015.

[4] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.

[5] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.

[6] Y. Wang, T. Tang, B. Ning, T. J. J. van den Boom, and B. De Schutter, "Passenger-demands-oriented train scheduling for an urban rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 1–23, 2015.

[7] K. Wu and Z. Wu, "A study of the model and algorithm of real-time train scheduling for urban rail transit . Technology and Economy in," *Areas of Communications*, vol. 19, pp. 30–35, 2017.

[8] S. Su, X. Li, T. Tang, and Z. Gao, "A subway train timetable optimization approach based on energy-efficient operation strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 883–893, 2013.

[9] X. Li and H. K. Lo, "Energy minimization in dynamic train scheduling and control for metro rail operations," *Transportation Research Part B: Methodological*, vol. 70, pp. 269–284, 2014.

[10] X. Yang, A. Chen, X. Li, B. Ning, and T. Tang, "An energy-efficient scheduling approach to improve the utilization of regenerative energy for metro systems," *Transportation Research Part C: Emerging Technologies*, vol. 57, pp. 13–29, 2015.

[11] S. Zhang, *Train scheduling and delay analysis of urban rail transit based on dwell time model [Ph.D. thesis]*, Beijing Jiaotong University, 2014.

[12] Y. Yue, J. Han, S. Wang, and X. Liu, "Integrated Train Timetabling and Rolling Stock Scheduling Model Based on Time-Dependent Demand for Urban Rail Transit," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp. 856–873, 2017.

[13] X. Zhou and M. Zhong, "Single-track train timetabling with guaranteed optimality: branch-and-bound algorithms with enhanced lower bounds," *Transportation Research Part B: Methodological*, vol. 41, no. 3, pp. 320–341, 2007.

[14] "Statistical analysis and reporting in 2016 for urban rail transit," http://mp.weixin.qq.com/s/fyEsfhcJxP6kdttTNVXcPg.

[15] G. Zhang, *Operation Organization of Urban Rail Transit*, Shanghai publisher of Science and Technology, Shanghai, China, 2012.

[16] A. Ceder, *Designing Transit Short-Turn Trips with the Elimination of Imbalanced Loads*, Springer Berlin Heidelberg, Heidelberg, Germany, 1988.

[17] P. G. Furth, *Short Turning on Transit Routes*, Eurailpress, Hamburg, Germany, 1987.

[18] 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.

[19] 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.

[20] D. Leffler, O. Cats, E. Jenelius, and W. Burghout, "Real-time short-turning in high frequency bus services based on passenger cost," in *Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017*, pp. 861–866, Italy, June 2017.

[21] 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.

[22] N. Ghaemi, O. Cats, and R. M. Goverde, "A microscopic model for optimal train short-turnings during complete blockages," *Transportation Research Part B: Methodological*, vol. 105, pp. 423–437, 2017.

[23] Y.-Y. Wang and S.-Q. Ni, "Optimization of train schedules of full-length & short-turn operation modes in urban rail transit," *Tiedao Xuebao/Journal of the China Railway Society*, vol. 35, no. 7, pp. 1–8, 2013.

[24] G. Bai, J. Guo, Y. Yang, and H. Shi, "Optimal train schedule for urban rail transit trains with short-turn strategy," *Urban Mass Transit*, vol. 18, pp. 45–50, 2015.

[25] S. Li, C. Guanghua, and C. Yongbing, "Practical analysis of long short routing operation mode for shanghai metro line 1," *Urban Mass Transit*, vol. 10, pp. 50–53, 2007.

[26] Y. Wang, "Study on optimization model of urban transit network organization and coordination with long and short routing," *Logistics Technology*, vol. 30, pp. 71–74, 2011.

[27] Q. Hu, "Study on mixed-routing in railway transit for disequilibrium passenger flow," *Value Engineering*, vol. 31, pp. 81–83, 2012.

[28] S. Chang, J. C. Jong, and Y. C. Lai, "Integrated optimization model for the train scheduling and ultilization planning problems in mass rapid transit systems," in *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis*, pp. 1–16, Tokyo, Japan, 2015.

[29] M. Peeters and L. Kroon, "Circulation of railway rolling stock: a branch-and-price approach," *Computers & Operations Research*, vol. 35, no. 2, pp. 538–556, 2008.

[30] E. Abbink, B. van den Berg, L. Kroon, and M. Salomon, "Allocation of railway rolling stock for passenger trains," *Transportation Science*, vol. 38, no. 1, pp. 33–41, 2004.

[31] Y.-C. Lai, D.-C. Fan, and K.-L. Huang, "Optimizing rolling stock assignment and maintenance plan for passenger railway operations," *Computers & Industrial Engineering*, vol. 85, pp. 284–295, 2015.

[32] 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.

[33] 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.

[34] H. P. Williams, "Model building in mathematical programming," *Journal of the Operational Research Society*, vol. 51, 4th edition, 1999.

[35] Y. Wang, T. Tang, B. Ning, and L. Meng, "Integrated optimization of regular train schedule and train circulation plan for urban rail transit lines," *Transportation Research Part E: Logistics and Transportation Review*, vol. 105, pp. 83–104, 2017.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 2710608, 13 pages
https://doi.org/10.1155/2018/2710608

Research Article

Using Smart Card Data Trimmed by Train Schedule to Analyze Metro Passenger Route Choice with Synchronous Clustering

Wei Li,[1,2] **Qin Luo**,[2,3] **Qing Cai**,[4] **and Xiongfei Zhang**[3]

[1]*Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Shenzhen, China*
[2]*Shenzhen Key Laboratory of Urban Rail Transit, Shenzhen University, Nanshan Ave 3688, Shenzhen, China*
[3]*College of Urban Traffic and Logistics, Shenzhen Technology University, Lantian Road 3002, Shenzhen, China*
[4]*Department of Civil, Environment and Construction Engineering, University of Central Florida, Orlando, Florida 32816, USA*

Correspondence should be addressed to Qin Luo; luoqin82@126.com

Received 23 November 2017; Revised 13 March 2018; Accepted 29 April 2018; Published 24 July 2018

Academic Editor: Francesco Corman

The metro passenger route choice, influenced by both train schedule and time constraints, is important to metro operation and management. Smart card data (Automatic Fare Collection (AFC) data in metro system) including inbound and outbound swiping time are useful for analysis of the characteristics of passengers' route choices in metro while they could not reflect the property of train schedule directly. Train schedule is used in this paper to trim smart card data through removing inbound and outbound walking time to/from platforms and waiting time. Thus, passengers' pure travel time in accord with trains' arrival and departure can be obtained. Synchronous clustering (SynC) algorithm is then applied to analyze these processed data to calculate passenger route choice probability. Finally, a case study was conducted to illustrate the effectiveness of the proposed algorithm. Results showed the proposed algorithm works well to analyze metro passenger route choice. It was shown that passenger route choice during both peak period and flat period could be clustered automatically, and noise data are isolated. The probability of route choice calculated through SynC algorithm can be used to revise traditional model results.

1. Introduction

Metro passenger route choice is vitally important to metro operation and management, such as passenger flow distribution and metro tickets clearing. It can provide useful data to help enhance train schedules to make full use of the train capacity. However, the metro passenger behavior is totally different from the car user behavior. The former one is largely influenced by both metro network structure and train schedule while the latter one is mostly decided by users themselves. On one hand, different metro network structures will lead to different route choices. For example, passengers would like to select those routes with few transfers. On the other hand, the train schedule will also influence passenger behaviors. Coordinated transit line could reduce passengers' waiting time in transfer stations. The routes with coordinated transit line should be more attractive than those without coordinated transit line.

So far, many scholars have modeled, analyzed, and studied the problem of passenger route choice behavior within private transportation, such as Kato et al. [1]. Unlike private transportation, metro trains are operated according to the train schedule, leading metro passengers' traveling to be restricted to the schedule. Therefore, traditional methods used in private transportation are not applicable for analyzing metro passenger behavior. Hence, the researchers tried to adopt some technologies widely used in metro transportation into the metro passenger behavior analysis. Among them, AFC (Automatic Fare Collection) system can collect these smart card data about passenger swipe inbound and outbound time of stations, which is useful for analyzing passenger behavior. A lot of research has been done to analyze passenger route choice based on smart card data. However, passengers with different walking time and waiting time may select the same route as metro trains' arrival and departure are dispersed. Hence, passengers walking time to/from platforms

and waiting time on platforms which were included in the smart card data should be useless for the analysis of passenger route choices.

This paper aims to propose a new method to analyze metro passenger route choice over travel periods based on smart card data and train schedule. Firstly, smart card data are trimmed using train schedule to eliminate walking time to/from platforms and waiting time. Then synchronous clustering algorithm, a kind of cluster algorithm, is applied to analyze passenger route choice based on these preprocessed data. Finally, a case study is carried out on the Shanghai metro network to validate the proposed algorithm.

2. Literature Review

Traditional methods on passenger behavior can be classified by Wardrop Law (Liu et al. [2]) as nonequilibrium model and equilibrium model (Smith et al. [3]). They believed that passengers' trip preference depends on travel time perception while individuals' perceptions are different. Some scholars put forward the stochastic user equilibrium model (stochastic user equilibrium (SUE)) to describe the problem. A simulation method was used to realize random users equilibrium model, and experiments were carried out in a large scale urban rail transit network (Kato et al. [1]). With the continuous expansion of parameter types and network sizes, SUE model has been becoming more and more complex for the reality (Thomas [3], Cascetta [4]). However, some scholars found that the traditional models may have some defects when they are applied in metro transportation. The main reason is that passengers' travel routes are affected by metro train schedule; that is to say, metro passengers' arrival and departure are limited to trains' arrival and departure. Thus the applicability of these traditional models is questioned.

The AFC system has been put into application in many metro systems worldwide. AFC system can record these data including passenger inbound swiping time, outbound swiping time, and some other related information. These data are useful in analyzing the passengers' route behaviors in metro. Pelletier [6] divided the usage of smart card data into three categories, long-term planning service, short-term planning service, and operation planning service. For example, swipe card data can be used to forecast the passenger flow OD matrix (Munizaga and Palma [7, 8]), to deal with demand analysis (Morency et al. [9]), to carry on operation and management of rail transit planning (Utsunomiya et al. [10]), etc.

Specifically, smart card data are getting more attention and more research has been made recently. Chan [11] put forward two research ideas based on London metro transit Oyster card data: one was to estimate the OD traffic matrix and the other was to build the metro transit service reliability matrix. This is the first time to use historical card data to make metro transit service quality evaluation. The main application of smart card is to analyze passenger travel behavior. For example, Kusakabe et al. [12] proposed a method to predict the specific trains that passengers choose to ride by using a vast number of long-term history swipe data and parameters. Zhu et al. [13] proposed a method to calibrate the metro passenger behavior model using the AFC data with the genetic algorithm and parameter estimation combining technology. Zhu et al. [14] presented a methodology for assigning passengers to individual trains using both smart card data and AVL data from train tracking systems; it can estimate the probability of the passenger boarding each feasible train and the probability distribution of the number of trains a passenger is unable to board due to capacity constraints. Ma et al. [15] developed a data mining method to identify the spatiotemporal commuting patterns of Beijing public transit riders using transit smart card data. Hong et al. [16] proposed a methodology for assigning passenger flows on a metro network based on Automatic Fare Collection (AFC) data and realized timetable. Briand et al. [17] analyzed the behavioral habits of public transport passengers using a real dataset of smart card data covering a period of five years. Faroqi et al. [18] investigated the relationship between passengers' spatial and temporal characteristics with a novel passenger-based perspective using smart card data. It is implemented for four-day smart card data including 80,000 passengers in Brisbane, Australia. Similarly, Zhu et al. [19] presented an integrated framework for estimating individual passenger's train choices through a data-driven approach with real timetable and Automatic Fare Collection (AFC) data. Besides, smart card data can also be used for estimation or prediction. For example, Hörcher et al. [20] presented a comprehensive method to estimate the user cost of crowding in terms of the equivalent travel time loss with large scale smart card, in a revealed preference route choice framework. Zhao et al. [21] developed a methodology for predicting daily individual trip making and trip attributes using transit smart card data, and the methods are tested using transit smart card data of 10,000 users in London. Also, smart card data are used to make metro train schedule. Zhang et al. [22] proposed a novel method to optimize the skip-stop scheme for bidirectional metro lines using the time-dependent passenger demand extracted from smart card data, so that the average passenger travel time can be minimized.

Some recent studies have made some progress on analyzing passenger behavior based on smart card data, part of which are useful for realistic size networks. The specific focus of this paper is to propose a method specifically aimed at using a small number of parameters, so that it can be easily used for large scale networks. Hence, this paper uses data analysis methods, i.e., ***cluster algorithm***, to analyze the passenger route choice behaviors on metro networks. The cluster algorithm is a method of multivariate statistical analysis. Data are classified according to individual characteristics so that the data in the same category have the highest homogeneity. On the other hand different categories should have relatively higher heterogeneity. The cluster algorithm aims to analyze and mine the intrinsic structure and rules of given data [23, 24]. In the process of data clustering, the clustering algorithm can automatically divide data points into different sets according to the attributes. These data with similar attributes are divided into the same set, while these data points with different attributes are divided into different sets [25]. Clustering algorithms can be divided into several types: clustering algorithms based on division (i.e., K-means),

clustering algorithms based on density (i.e., DBSCAN and OPTICS), affinity propagation clustering algorithm (affinity propagation (AP) algorithm), synchronous clustering algorithm (SynC algorithm), etc.

K-means algorithm is the most widely used clustering algorithm based on division. It has been nearly 60 years since it was proposed [26]. However, the biggest shortcoming of the K-means algorithm is to select the initial K value and the value of the selected K data points since the initial value may lead the convergence of the K-means algorithm to different results. Hence, many scholars proposed other new clustering algorithms, among which AP algorithm is one kind of typical clustering algorithms [27]. AP clustering algorithm does not need to specify the number of clusters in advance. ***Synchronous clustering algorithm*** (SynC algorithm) [28, 29] is another kind of clustering algorithm of which initial values are not sensitive. The main idea of synchronous clustering is that each data point is regarded as an independent individual, and similar individuals automatically get together to form clustering collections. Due to the characteristics of synchronous clustering algorithm, this algorithm has many advantages; for example, (1) the algorithm does not require given cluster centers in advance, (2) the algorithm is not sensitive to the initial value, and (3) the algorithm can well avoid noise interference data.

However, to our best knowledge, no studies adopted the SynC algorithm to analyze metro passenger route choices with smart card data trimmed by train schedules. Hence, taking the advantages of the synchronous clustering algorithm (SynC) into consideration, this paper adopts the SynC algorithm to analyze metro passenger behavior.

3. Methodology

3.1. Basic Assumptions. Some necessary assumptions and elements are firstly described as follows:

(1) All passengers' behaviors are assumed to be reasonable, and passengers would not stay in stations for a long time. But there are always some unreasonable data which spend a very long time or an extremely short time during given OD pairs. This proposed algorithm will regard these data as noise data in the dataset.

(2) Train congestion is not considered in data preprocessing. It means passengers can ride the first arriving train after they reach platforms.

(3) All trains are operated according to the train schedule strictly.

3.2. Definition of Train Schedule and Smart Card Data

3.2.1. Train Schedule. The metro train schedule contains necessary information of all trains running on the network, like train codes, arrival and departure time of trains at each station, etc. Figure 1 shows an example of a train schedule used by a metro line in Shanghai. Each red line represents a planned operation train.

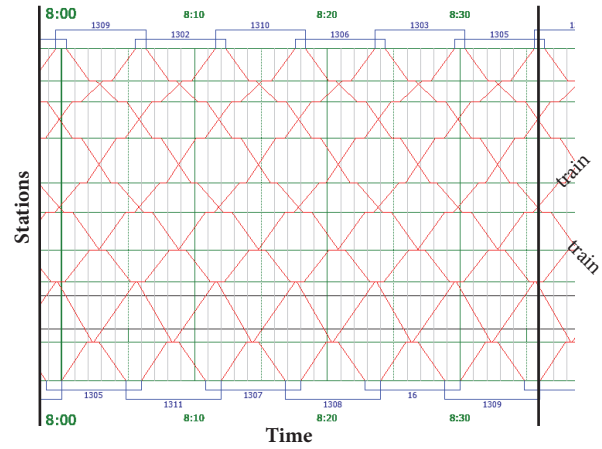


Figure 1: An example of the train schedule.

The definition of train schedule is described below: metro line is defined as $L = \{1, 2, \ldots, l, \ldots, N\}$, and the station collection on line l is $S_l = \{1, 2, \ldots, i, \ldots, M\}$. Then, station $S_{l,i}$ represents the station i in line l; $S_{l,i}(A^j_{l,i}, D^j_{l,i})$ defines the arrival time $A^j_{l,i}$ and departure time $D^j_{l,i}$ of train j at $S_{l,i}$. Thus, the trajectory of train j is described as $\{\forall i \in l \mid S_{l,i}(A^j_{l,i}, D^j_{l,i})\}$, and the network train schedule can be described as $T = \{\forall j, l, i \mid S_{l,i}(A^j_{l,i}, D^j_{l,i})\}$.

3.2.2. Smart Card Data. AFC system can record the original station (O is used in this paper), destination station (D is used in this paper), and their corresponding inbound and outbound time. These swiping data can be used to obtain the detailed passenger flow demand. Table 1 shows some examples of entry and exit swiping card data recorded by the AFC system, like card number, swiping date, inbound station code, inbound swiping time, outbound station code, outbound swiping time, etc.

Smart card data (AFC data) are defined as $OD(n, T^{(si)}, S^{(in)}, T^{(so)}, S^{(out)})$, in which n is the card ID, $T^{(ci)}$ is the inbound swiping time, $T^{(co)}$is the outbound swiping time, $S^{(in)}$ is the O station, and $S^{(out)}$ is the D station.

3.2.3. Passenger Travel Process on Metro. Figure 2 shows the metro passenger travel process. It displays typical metro passenger traveling, which mainly contains passengers' swiping card at entry gates, walking to platforms, waiting for coming trains, riding trains (transfer if it has), and finally walking out of station. As shown in the figure, symbol definition includes walking cost time (entry walking time, $T^{(in)}$), waiting cost time (waiting time on platforms, $T^{(w)}$), travel cost time (in-vehicle time, $T^{(v)}$), and walking out of station cost time (exit walking time, $T^{(out)}$). If a passenger makes a transfer, the additional transfer walking cost time (transfer walking time, $T^{(t)}$) and transfer waiting cost time (waiting time, $T^{(tw)}$) are required.

Table 1: Samples of smart card data.

Date	Card ID	O station	Inbound Time	D station	Outbound Time
2014-11-17	1416107917	0248	09:15:45	1056	09:36:00
2014-11-17	1282520204	0751	09:20:00	0727	09:36:17
2014-11-17	0934484109	1060	09:13:56	0248	09:36:22
2014-11-17	1069233288	0411	09:22:54	0750	09:36:41
.....					

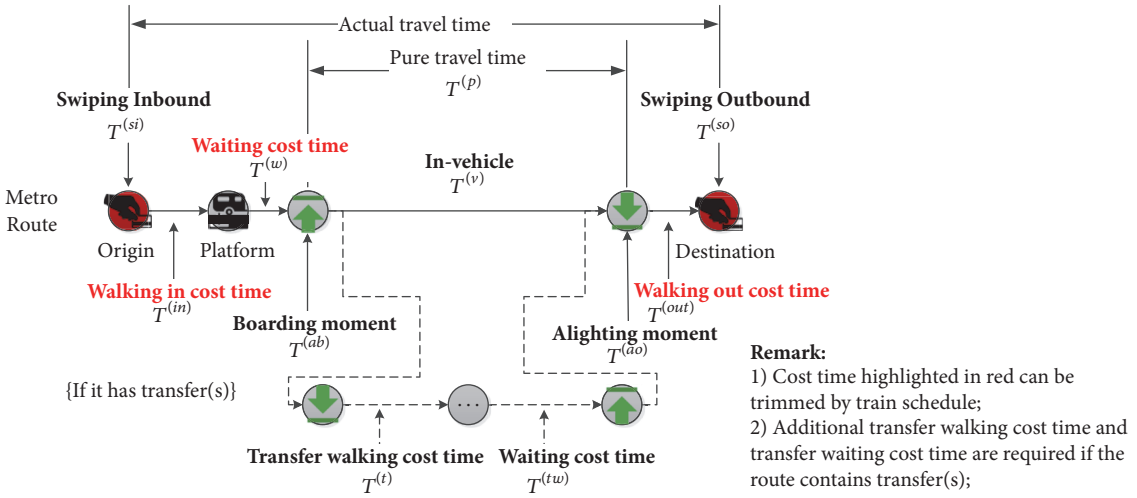


Figure 2: Passenger trip diagram by metro transit.

Here, $T^{(si)}$(inbound swiping time) is defined as the moment passengers swipe in stations. $T^{(so)}$(outbound swiping time) is defined as the moment passengers swipe out of stations. The difference between $T^{(si)}$ and $T^{(so)}$ is the passengers' actual travel time during metro. Besides, $T^{(ab)}$ (actual board time) is defined as the actual moment when passengers board trains, while $T^{(ao)}$ (actual alight time) refers to the actual moment when passengers alight trains. Then, the pure travel time (pure travel time, $T^{(p)}$) is the difference between $T^{(ab)}$ and $T^{(ao)}$. It is obvious that the values of $T^{(ab)}$ and $T^{(ao)}$ are limited to train arrival, which is related to the train schedule.

3.3. AFC Data Trimmed by Train Schedule. The passengers' travel time by metro (actual travel time is used in this paper) can be obtained from the difference between the inbound swiping time and the outbound swiping time from smart card data. Obviously, the actual travel time could be different in one OD pair if passengers select different route. When the difference of route travel time between OD pairs is large, passenger's selected route can be easily decided based on the travel time. However, smart card data contains inbound and outbound walking time and waiting time, which are useless information. Since trains' arrival at stations is dispersed, some passengers with different walking time may take the same trains. That is to say, some passengers may take the train just after they arrive at platforms, while some passengers may wait for a long interval for a train they just miss. Thus, the travel time without waiting time and walking time at O station and D station can present more useful information than the travel time with waiting and walking time.

We could use train schedule to trim smart card data by removing walking and waiting time at O stations and walking time at D stations. The trimmed result can be used in cluster algorithm, subsequently. Figure 3 shows some passenger travel time before and after using AFC data trimming algorithm. It can be seen that the original AFC data are out of order, while these data after trimming are orderly. The pure travel time could reflect some discrete characteristics of train arrival and departure.

The method to determine passengers' actual boarding and alighting time is shown in Figure 4. First, for each AFC data, its inbound station is set as $S^{(in)} = S_{l,i}$, and its inbound time is set as $T^{(si)}$. Find train j based on the following equation after searching all trains which run pass $S_{l,i}$ in order:

$$D_{l,i}^{j-1} \le T^{(si)} \le D_{l,i}^{j} \tag{1}$$

It means that passengers can ride train j to their destinations or transfer stations. Thus the possible actual board time $T^{(ab)}$ is

$$T^{(ab)} \longleftarrow D_{l,i}^{j} \tag{2}$$

Similarly, the actual alighting time can be obtained in the same way. Its outbound station is set as $S^{(out)} = S_{l,i'}$, while its outbound time is set as $T^{(so)}$. Find train j' with the following

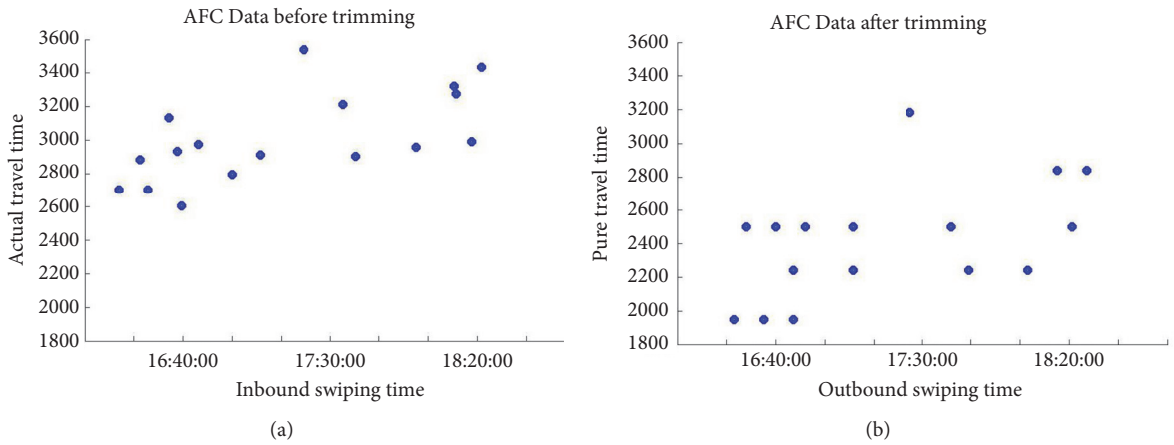


Figure 3: A sample of AFC data before and after trimming.

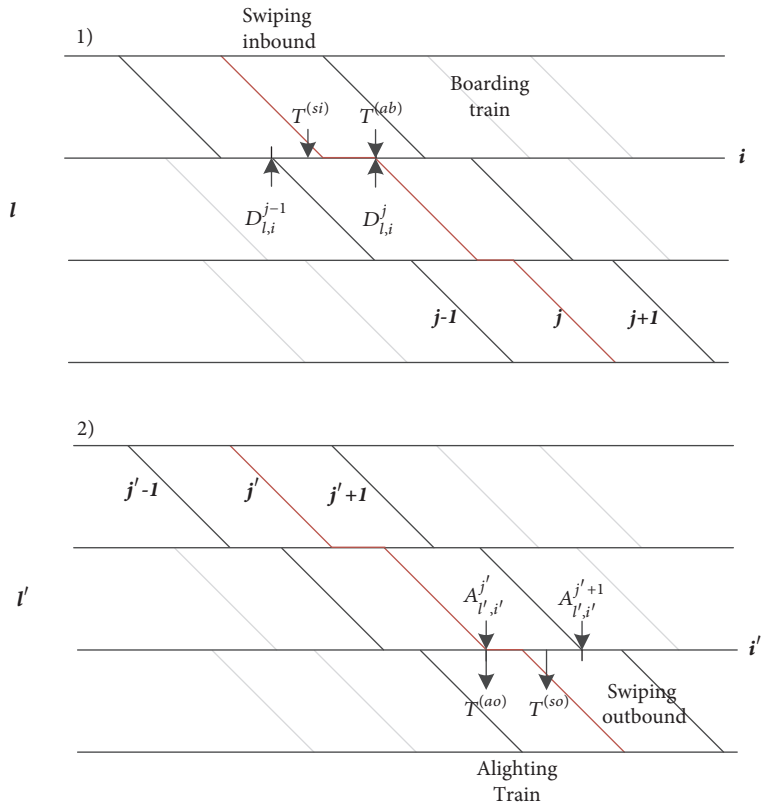


Figure 4: Determination of passengers' actual boarding (1) and alighting (2) time.

equation after searching all trains which run pass $S_{l',i'}$ in reverse order:

$$A^{j'}_{l',i'} \le T^{(so)} \le A^{j'+1}_{l',i'} \tag{3}$$

Thus the possible actual board time $T^{(ao)}$ is

$$T^{(ao)} \longleftarrow A^{j'}_{l',i'} \tag{4}$$

It should be noted that a least walking time is needed to enter into or exit from the platform by gates. The minimum time constraint ε is considered in $T^{(ab)}$ and $T^{(ao)}$ as follows:

$$D^{j-1}_{l,i} \le T^{(si)} + \varepsilon \le D^{j}_{l,i} \tag{5}$$

$$A^{j'}_{l',i'} \le T^{(so)} + \varepsilon \le A^{j'+1}_{l',i'} \tag{6}$$

Therefore, the pure travel time can be acquired by

$$T^{(p)} = T^{(ao)} - T^{(ab)} \tag{7}$$

3.4. SynC Algorithm. Based on the pure travel time, this paper applies SynC algorithm analysis to process these data. This part presents how to use the SynC algorithm to analyze metro passenger route choice.

3.4.1. Data Normalization. Before cluster, the data need to undergo normalization since data points may have different scales and dimensions which will affect the effectiveness of clustering algorithm. Data normalization is firstly adopted to make data fall into a certain range. This paper wants to make inbound swiping time and pure travel time into the same certain range to carry on the cluster.

Z-score normalization is used in this paper to carry on data normalization, which is based on the mean and standard deviations of attribute values. The advantage of Z-score normalization is that it does not need to compute the maximum and minimum values of the data set and has good effects on the normalization of outliers. Its formula is

$$\widetilde{v_i} = \frac{v_i - \overline{v}}{\sigma} \tag{8}$$

where $\overline{v}$ is the mean value of attribute value, and σ is the standard deviation of attribute values.

3.4.2. Synchronous Clustering Algorithm (SynC Algorithm). The main idea of SynC algorithm is to regard each data point as an individual, and the similar points would get clustered. The procedure of the algorithm is shown in Figure 5: firstly, data points are independent and move close to their similar data points, as shown in Figure 5(a)); secondly more and more data points will gather together to the one with same attribute, as shown in Figure 5(b)); finally, all similar data points are clustered together to form a cluster center, while some noise data are automatically isolated, as shown in Figure 5(c)).

Some equations should be given in SynC algorithm.

Definition 1 (domain distance ϵ). It means the maximum distance from the given point.

Definition 2 (Nb_ϵ (the ϵ collection of data point x)). Let x be a data point of data set D; Nb_ϵ means the data whose distance from x is smaller than ϵ:

$$Nb_\epsilon = \{y \in D \mid dist(y, x) \le \epsilon\} \tag{9}$$

where $dist(y, x)$ is the distance between data points x and y.

Definition 3 (Kuramoto Amplitude of data point x). Let x_i be the ith dimension of data point x. After it is influenced by other points in Nb_ϵ, the Kuramoto Amplitude of data point x_i can be described as

$$\frac{d_{x_i}}{dt} = \omega_i + \frac{S}{|Nb_\epsilon|} \sum_{y \in Nb_\epsilon(x)} \sin(y_i - x_i) \tag{10}$$

where ω can be ignored in this cluster algorithm, and S is a constant (equal to 1 in this part). Finally, the Kuramoto Amplitude can be rewritten as

$$x_i(t+1) = x_i(t) + \frac{1}{|Nb_\epsilon|} \sum_{y \in Nb_\epsilon(x)} \sin(y_i - x_i) \tag{11}$$

where t is the time step, and $t = 0$ represents the initial state.

Definition 4 (synchronous coordination parameter). It represents the degree of synchronous coordination of all data points in the data set at the current time step:

$$r_c = \frac{1}{N} \sum_{i=1}^{N} \sum_{y \in Nb_\epsilon(x)} e^{-(y_i - x_i)} \tag{12}$$

It can be seen that synchronous coordination parameter of the data set will increase gradually when more data points gather together. And after the parameter does not change for a long time, the data set achieves convergence within Nb_ϵ. It reaches a local synchronized status. Finally, when all data points gather together ($r_c \longrightarrow 1$), it reaches a global synchronized status.

Definition 5 (optimal domain distance ϵ). It means the cluster result is the best when ϵ is equal to a certain value. The optimal distance can be determined according to the SynC algorithm [28]:

$$M^j = \underset{j}{\operatorname{argmin}} L\left(D, M^j\right) \tag{13}$$

where M^j is the jth cluster center of the given data; argmin_j is the function that can calculate j which leads the value of $L(D, M^j)$ to be minimum.

$L(D, M^j)$ can be computed by following equations:

$$L\left(D, M^j\right) = L(M) + L(D \mid M) \tag{14}$$

$$L(M) = \sum_{i=1}^{K} \sum_{j=1}^{|C_i|} \log_2\left(\frac{N}{|C_i|}\right) + \sum_{i=1}^{K} \frac{p_i}{2} \log_2\left(\frac{N}{|C_i|}\right) \tag{15}$$

$$L(D \mid M) = -\sum_{i=1}^{K} \sum_{x \in C_i} \log_2\left(pdf(x)\right) \tag{16}$$

where K is the number of cluster centers; C_i is the ith cluster set; $|C_i|$ is the number of data points in C_i; p_i is the data dimension; $pdf(x)$ is the probability of data point x which belongs to C_i.

Therefore, the steps of synchronization clustering algorithm (SynC algorithm) are described as follows, while the flowchart of SynC algorithm is shown in Figure 6:

(1) Initial time step is set as $t = 0$, and all data points are regarded as independent cluster center.

(2) Set domain distance ϵ, and calculate Nb_ϵ of all data points.

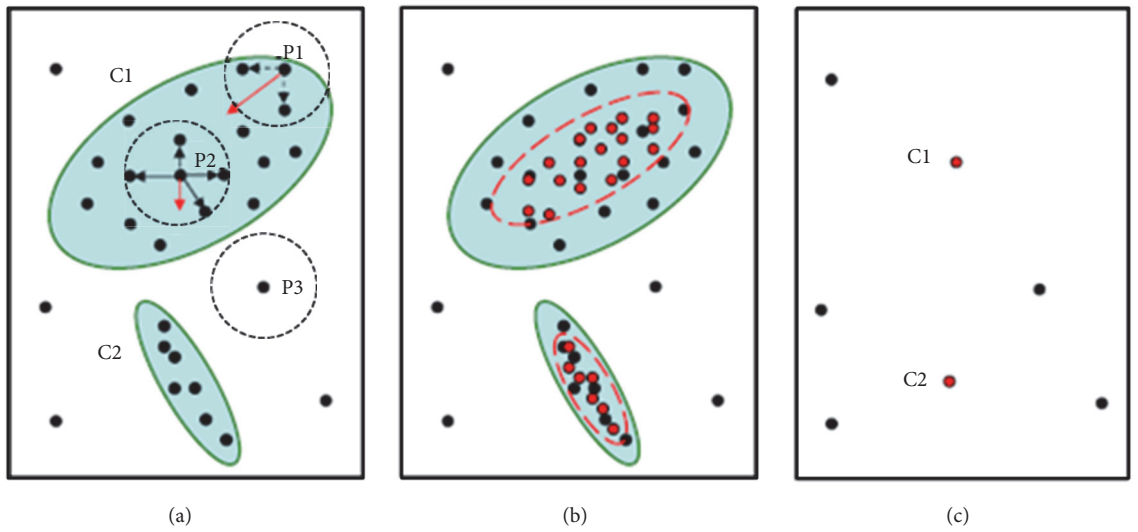


Figure 5: Sketch of synchronous clustering (SynC) algorithm process [28].

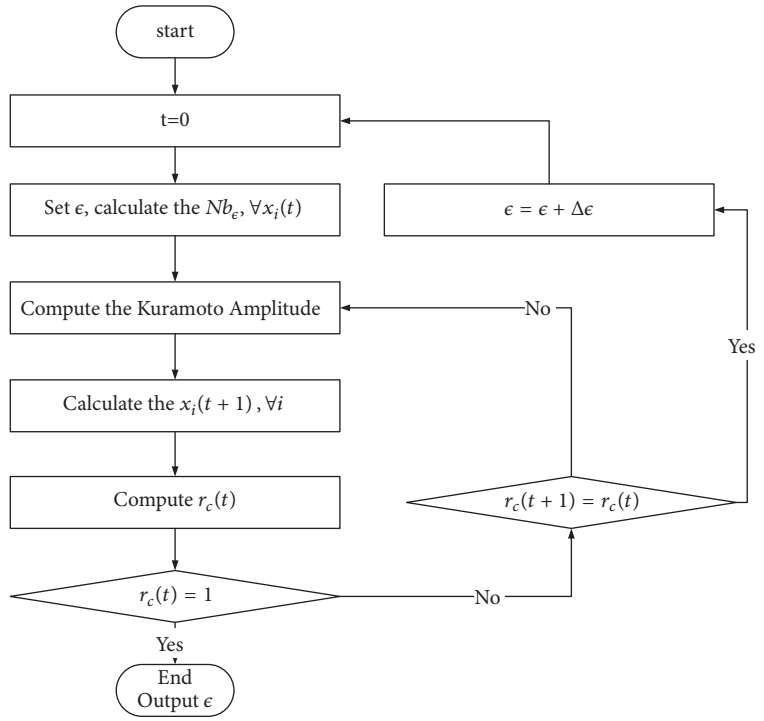


Figure 6: Flowchart of SynC algorithm.

(3) Compute the Kuramoto Amplitude of all data points using Nb_ϵ, and data points of $x_i(t+1)$ can be calculated when it moves to next time step ($t = t+1$).

(4) Compute the synchronous coordination parameter r_c of this data set at this time step.

(5) If $r_c = 1$, then it reaches a global synchronized status, the algorithm ends and the optimal domain distance ϵ can be computed. If this is not the case, the algorithm moves to step (6).

(6) If r_c remains the same ($r_c(t+1) = r_c(t)$), then it reaches a local synchronized status. Let $\epsilon = \epsilon + \Delta\epsilon$, $t = 0$, move to step (2), and start a new cluster. Otherwise, move to step (3) and continue this cluster.

4. Case Study

To evaluate the proposed algorithm of smart card data trimming and SynC, a real-life metro network (the Shanghai Metro system, shown in Figure 7) with a large number of

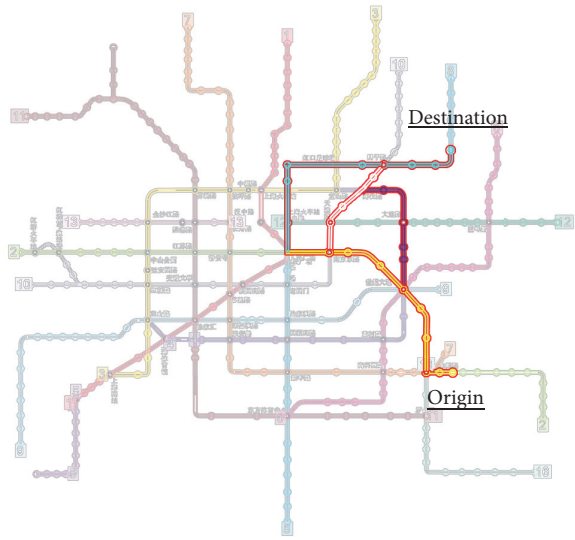


Figure 7: Shanghai metro network.

lines and stations is presented as a case study application. The network consists of 14 transit lines and each has an upstream direction and a downstream direction. There are totally 289 stations in the network, of which 42 stations are transfer stations. Jinke Road Station and Huang Xing Road Station are selected as O and D station in this case. Jinke Road Station is surrounded by working companies, while Huangxing road is located in the residential area. It leads to the fact that there are larger passenger flows in the OD pair during the evening peak.

4.1. Calculation Process

(1) OD Pair. Jinke Road Station (station code 0254) in line 2 is taken as O station and Huangxing Road Station (station code 0844) in line 8 is taken as D station. Whole week data from November 11, 2016, to November 15, 2016, are selected in this case, which had 199 data records in total. (The AFC data were obtained from the Shanghai Metro Company.)

(2) Train Schedules and AFC Data Trimming. To make the case study easy to program, the planned train schedule instead of actual schedule is used. And the planned train schedule using at weekday during November 2016 is applied in the case study, and all trains are assumed to operate according to the train schedule strictly. Train schedule is used to trim AFC data to obtain the pure travel time by removing entry/exit walking time and waiting time according to the proposed AFC data trimmed method. The process is shown in Table 2.

(3) Data Normalization. The inbound swiping time is selected as the X axis of cluster data set and the pure travel time is selected as the Y axis. However, due to the different dimensions of data points, data normalization is needed to get a better cluster result. The normalization example of the data points is shown in Table 3.

(4) Clustering Process. C#.Net programming language is applied to program coding to achieve the algorithm. Figure 8 shows the process of SynC algorithm. The X axis is inbound time after data normalization while the Y axis is pure travel time after data normalization. And two horizontal lines in each figure represent morning and evening peak period, respectively. Each part in Figure 8 represents a local synchronized status in SynC algorithm. At the first part, each data point is regarded as a cluster center/centroid. The data points automatically get together in local synchronized status, leading centroids to be merged slowly in the following parts. It can be seen that, with the clustering process, data points gradually merge to form cluster centers, and noisy data are isolated obviously at the same time, when reaching the optimal domain distance as (13)-(16). The final result is shown in Figure 9. Point color refers to the cluster they belong to. The more the data points of the same color, the higher the passenger flow this route has. Passenger route selection probability during both peak and flat period is easy to obtain with the result.

4.2. Algorithm Analysis.

The cluster algorithm applies the pure travel time which removes entry/exit walking time and waiting time using train schedule. Some comparative analyses are made in this part. Figure 10 shows the cluster results using both AFC data with trimming (Figure 10(a)) and AFC data without trimming (Figure 10(b)) by train schedule. It is indicated that the trimming results could present metro travel time characteristics clearly while the no-trimming results present passenger travel time disorderly. Thus, pure travel time trimmed by train schedules could represent some discrete characteristics of metro transportation since it could take train schedules into consideration.

4.3. Result Analysis.

Table 4 shows cluster results by the distinction of early peak, flat peak, and evening peak. This table shows passengers preference on route choice with different periods. And clusters with small passenger flow are regarded as noisy. Table 5 shows the route list of this OD pair in traditional model used in Shanghai Metro Company [5]. The candidate route sets are generated according to *the K-short algorithm* with route expected travel time, and the selection probability of each route is calculated by logistics model. It contains some possible routes that passengers may choose to follow and the corresponding selection probability of each route. This table is very important to metro operation since it is used to calculate the passenger flow distribution of the whole network. Also the allocation to each metro line is decided by the line passenger flow computed by the traditional model results.

The routes in Table 5 are used to link the cluster centers in Table 4 according to the comparison of travel time. Take Table 5 (route list) as a contrast; the following results can be summarized from Table 4 (cluster result):

(1) There are mainly two routes during morning peak period. About 60% of passengers choose the route with a long time but less transfer (Route No. 3 in Table 5), and 40% of passengers choose the route with

Table 2: Process of AFC data trimming.

Card ID	Inbound time	Outbound time		Card ID	Actual board time	Actual alighting time
2914517671	07:18:57	08:15:27		2914517671	07:24:53	08:11:50
1119250823	07:21:09	08:17:47	Trimming	1119250823	07:24:53	08:16:50
2663388071	07:22:15	08:20:28	⟶	2663388071	07:24:53	08:14:20
3525029280	10:09:43	11:04:26		3525029280	10:14:10	10:59:02
1454658848	16:27:52	17:15:59		1454658848	16:31:50	17:11:19
	.....				.....	

Table 3: Data normalization.

ID	Inbound time	Pure travel time		ID	Inbound time	Pure travel time
1	07:24:53	2817		1	-2.260	-0.382
2	07:24:53	3117	Normalization	2	-2.260	-0.268
3	07:24:53	2967	⟶	3	-2.260	-0.057
4	10:14:10	2692		4	-1.516	-0.652
5	16:31:50	2369		5	0.143	-1.352
	.....				.....	

Table 4: Result of synchronous clustering algorithm (Sync).

Clusters	Normalized X axis	Alighting time∗	Normalized Y axis	Pure travel time∗	Passenger flow (Data count)	Ratio	Remark
Morning Peak							
1	-2.165	7:46:28	0.225	3097.2	17	58.6%	
2	-2.055	8:11:22	-1.051	2508.0	12	41.4%	
Flat Period							
3	-1.301	11:03:05	-0.232	2886.0	9	36.0%	
4	-1.065	11:56:50	0.737	3333.7	3	12.0%	Noisy
5	-0.719	13:15:37	1.769	3810.3	2	8.0%	Noisy
6	-0.404	14:27:06	-0.328	2841.9	11	44.0%	
Evening Peak							
7	0.464	17:44:50	1.138	3519.0	19	13.1%	
8	0.474	17:46:59	3.923	4805.2	2	1.4%	Noisy
9	0.485	17:49:36	-0.033	2978.3	88	60.7%	
10	0.763	18:52:50	-0.947	2555.8	36	24.8%	

Table 5: Route list of the OD pair according to traditional model [5].

Route ID	Origin	Destination	Pass Line	Travel time	Selection probability
1	0254	0844	2-10-8	2509	50%
2	0254	0844	2-4-10-8	2621	33%
3	0254	0844	2-8	2981	17%

a short time (Route No. 1 in Table 5). This result is not similar to Table 5. It is a bit surprising that not all passengers selected the route with the shortest travel time. The possible reasons for selecting the route with a longer travel time but less transfer during peak period are that passengers may want to avoid station congestion. Station congestion may lead to the fat that they miss the first arrival train because of not enough space in vehicle and too many passengers on the platform. Thus, passengers may think the transfer could take them more time in their trips during morning peak.

(2) As shown in Table 4, the difference between cluster 3 and cluster 6 is small; thus these two clusters can actually be considered the same one. After linking clusters to routes, we could find that most passengers choose Route No. 1 or Route No. 2, while few choose Route 3 during flat period, which is in line with the result of Table 5. The results can reveal the fact that passengers do not expect shorter travel time but expect more comfortable service instead during flat period. Besides, many noisy points could be found during flat periods, which is relative long travel time with less passenger flow in metro system. It means

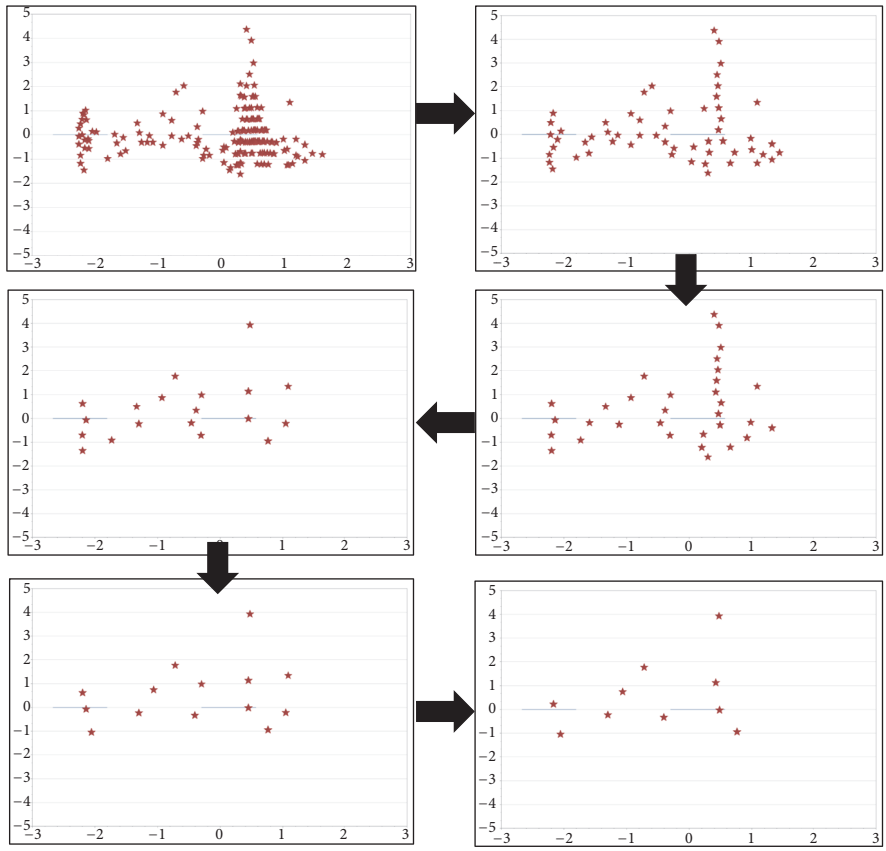


Figure 8: Process of synchronous clustering (SynC) algorithm.

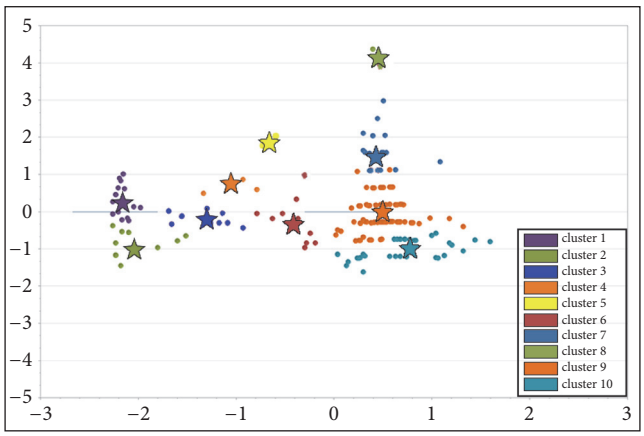


Figure 9: Result of synchronous clustering (Sync) algorithm.

passengers may not be in a hurry on their trips during flat periods.

(3) Passenger flow mainly occur during evening peak from Jinke road to Huangxing road, accounting for more than 70% of the total passenger flow on a whole day. The result shows that passengers are not sensitive to travel time during evening peak. A small number of passengers (about 25%) select Route No. 1 and Route No. 2, while the majority of the passengers (13%+61%) choose Route No. 3, which is different from the results in Table 5. It may be because that passenger prefer to travel with less transfers during evening peak.

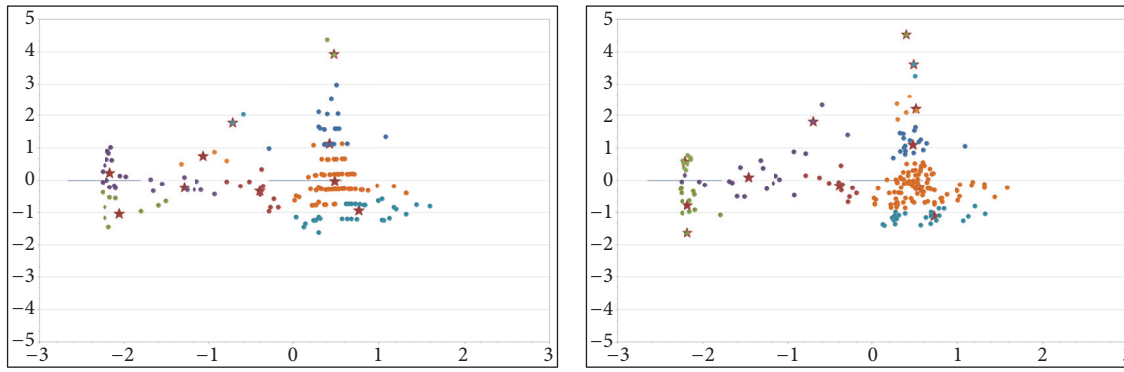
Metro passenger route choices could be various for their travel time and their inbound time, especially for peak and flat period. In this case study, passengers are more likely to choose Route No. 3 (with less transfer) during morning and evening peak while passengers are more likely to choose Route No. 1 or Route No. 2 (with less travel time) during flat period. Passengers' route choices may be influenced by both their travel moment and travel cost time.

It should be noted that smart card data with only a week range are used in the case study. The passenger route selection probability would be more reliable with more smart card data. Therefore, the result of the algorithm can be used to revise traditional model results like those of Table 5.

4.4. Algorithm Extension. The proposed algorithm can be applied to other OD pairs easily on metro network. But there are two limitations in the algorithm of AFC data trimmed by train schedule when it is used for other OD pairs. The one is that passengers can choose any transit line to finish their trips when origin station or destination station is a transfer station. The other one is that passengers can choose either upstream trains or downstream trains of the transit line to finish their trips when origin station or destination station

Table 6: Applicability of the proposed algorithm in Shanghai metro network.

Situation ID	Situation Name	OD pairs count	OD pairs percent	Passenger flow count	Passenger flow percent
		119918	100%	5313949	100%
1	OD pairs with unclear routes	23390	19.50%	606245	11.41%
2	OD pairs with similar travel time	7193	6.00%	98003	1.84%
3	OD pairs with small passenger flow	25737	21.46%	26810	0.50%
4	Noise data (Same in and out)	4112	3.43%	95712	1.80%
5	applicable OD pairs	53233	49.61%	4520205	84.44%



(a) (b)

Figure 10: Comparison of cluster results after and before data trimming.

is a normal station. These two kinds of OD pairs are called *unclear routes OD pairs*. The algorithm of AFC data trimmed by train schedule cannot be applied to these two kinds of OD pairs. The main reason is that passengers' walking in cost time or walking out cost time is not able to be removed from AFC data since it is not clear which transit line or which upstream/downstream trains of metro schedule should be chosen.

There are also two limitations in the SynC algorithm when it is used for other OD pairs. One is that it is useless to apply the algorithm into the OD pairs whose travel time of routes is similar. Routes cannot be clustered by these similar travel times. The other is that the SynC algorithm is useless when the passenger flow is very low between the OD pairs. The cluster algorithm cannot work with such less data.

Shanghai metro network is applied to discuss the applicability of the proposed algorithm, as shown in Table 6. The date of selected data is November 15, 2016. There are 119,918 OD pairs in this network and 5,313,949 passengers traveling on that day. Some interesting findings can be obtained as follows:

(1) There are about 20% OD pairs (23,390, 19.50%) in which passengers can ride more than one line to their destinations. These OD pairs contain 1) both upstream and downstream lines which are both feasible routes and 2) many routes which are feasible in original/destination stations when they are transfer ones; thus the proposed algorithm of AFC data trimming by train schedule cannot be used in this type of OD pairs, accounting for 11.41% passenger flow (606,245).

(2) There are about 6.00% OD pairs (7,193) having similar travel time. The SynC algorithm cannot use these data to cluster distinct points.

(3) However, more than 20% OD pairs (25,737) contain less than 5 passengers on a whole day. Such small passenger flow is useless for cluster. But there are only 26,810 passengers (accounting for 0.50%) traveling though these OD pairs.

(4) AFC data also contains some noisy data; for example, passengers may swipe in and swipe out from same stations. There are about 4,112 OD pairs (3.43%), accounting for passenger flow of 95,712 (1.80%). These data are useless for the analysis of passenger behavior among metro system.

(5) Therefore, besides those above data, the proposed algorithm can be used in about 50% OD pairs (53,233, 49.61%) to cluster passenger travel routes, while more than 85% AFC data (passenger flow, 4,520,205, 84.44%) can be used in clustering. That is to say, most

AFC data are useful for the analysis of passenger route behavior.

5. Conclusion

This paper studied metro passenger route choice with train schedule and cluster algorithm. On the basis of AFC data, the algorithm of AFC data trimmed by train schedule was proposed to obtain pure travel time. The results were then used in synchronous clustering algorithm to analyze the passenger route choice (selection probability) under time constraints. Then, a case study by using Shanghai metro data was conducted to validate the proposed algorithm. It was indicated that the probability of route choice can be calculated through SynC algorithm in different periods, and thus the algorithm can be used to revise traditional model results. The proposed algorithm can help to analyze passenger route preference with smart card data without traditional methods which contains a large number of parameters. And the passenger route preference would be relatively accurate with more smart card data.

However, there are some limitations in the proposed method which needs further research. (1) The journey time of different routes over different periods should be different. The travel time in Table 5 in the results of the paper has theoretical values, which are calculated by train section running time and passenger transfer cost time. The results do not consider congestion in trains and variable train operation headway in the calculation. Thus, further research should be made in the determination of dynamic travel time of passenger routes over periods. For example, congestion data in the train carriages and on the platforms of stations, which can be acquired by passenger flow detection devices based on image recognition, are useful for calculating dynamic journey time of passenger routes over periods. (2) Besides, how to link these clustering results to these travel routes automatically needs a further study, in order to make the data process more complete.

Conflicts of Interest

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

Acknowledgments

This paper is supported by the Research Projects of the Social Science and Humanity on Young Fund of the Ministry of Education under Grant no. 15YJCZH108 and the Research Projects of Natural Science Foundation of Guangdong Province under Grant no. 2015A030310341.

References

[1] 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.

[2] Y. Liu, J. Bunker, and L. Ferreira, "Transit userś route-choice modelling in transit assignment: a review," *Transport Reviews*, vol. 30, no. 6, pp. 753–769, 2010.

[3] R. Thomas, "Traffic assignment techniques," 1991.

[4] E. Cascetta, *Transportation Systems Analysis: Models and Applications*, Springer, 2009.

[5] R.-H. Xu, Q. Luo, and P. Gao, "Passenger flow distribution model and algorithm for urban rail transit network based on multi-route choice," *Journal of the China Railway Society*, vol. 31, no. 2, pp. 110–114, 2009.

[6] 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.

[7] M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport Origin-Destination matrix from passive smartcard data from Santiago, Chile," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 9–18, 2012.

[8] M. Munizaga, F. Devillaine, C. Navarrete, and D. Silva, "Validating travel behavior estimated from smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 44, pp. 70–79, 2014.

[9] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[10] M. Utsunomiya, J. Attanucci, and N. Wilson, "Potential uses of transit smart card registration and transaction data to improve transit planning," Transportation Research Record 1971, Transportation Research Board of the National Academies, Washington, DC, USA, 2006.

[11] J. Chan, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*, Massachusetts Institute of Technology, 2007.

[12] 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.

[13] W. Zhu, H. Hu, and Z. Huang, "Calibrating rail transit assignment models with genetic algorithm and automated fare collection data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 7, pp. 518–530, 2014.

[14] 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.

[15] X. Ma, C. Liu, H. Wen, Y. Wang, and Y.-J. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.

[16] 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, 2017.

[17] A.-S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 274–289, 2017.

[18] H. Faroqi, M. Mesbah, and J. Kim, "Spatial-temporal similarity correlation between public transit passengers using smart card data," *Journal of Advanced Transportation*, vol. 2017, 2017.

[19] W. Zhu, W. Wang, and Z. Huang, "Estimating train choices of rail transit passengers with real timetable and automatic fare

collection data," *Journal of Advanced Transportation*, vol. 2017, Article ID 5824051, 12 pages, 2017.

[20] D. Hörcher, D. J. Graham, and R. J. Anderson, "Crowding cost estimation with large scale smart card and vehicle location data," *Transportation Research Part B: Methodological*, vol. 95, pp. 105–125, 2017.

[21] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Individual mobility prediction using transit smart card data," in *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 19–34, 2018.

[22] P. Zhang, Z. Sun, and X. Liu, "Optimized Skip-Stop Metro Line Operation Using Smart Card Data," *Journal of Advanced Transportation*, 2017.

[23] K.-L. Du and M. N. S. Swamy, "Clustering i: Basic clustering models and algorithms," in *Neural Networks and Statistical Learning*, pp. 215–218, Springer, London, UK, 2014.

[24] K.-L. Du and M. N. S. Swamy, "Clustering II: Topics in clustering," in *Neural Networks and Statistical Learning*, pp. 259–297, Springer, London, UK, 2014.

[25] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, ASA-SIAM Series on Statistics and Applied Probability, SIAM American Statistical Association, Philadelphia, Pa, USA, 2007.

[26] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[27] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[28] C. Böhm, C. Plant, J. Shao, and Q. Yang, "Clustering by synchronization," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD-2010*, pp. 583–592, USA, July 2010.

[29] X. Chen, "A new clustering algorithm based on near neighbor influence," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7746–7758, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 3179321, 11 pages
https://doi.org/10.1155/2018/3179321

Research Article

Stop Plan of Express and Local Train for Regional Rail Transit Line

Qin Luo,[1,2] **Yufei Hou**,[2] **Wei Li**,[1,3] **and Xiongfei Zhang**[1]

[1]*College of Urban Traffic and Logistics, Shenzhen Technology University, Lantian Road 3002, Shenzhen, China*
[2]*Shenzhen Key Laboratory of Urban Rail Transit, Shenzhen University, Nanshan Ave 3688, Shenzhen 518000, China*
[3]*Key Laboratory of Optoelectronic Devices and Systems of Ministry of Education and Guangdong Province, College of Optoelectronic Engineering, Shenzhen University, Nanshan Ave 3688, Shenzhen, China*

Correspondence should be addressed to Wei Li; aliweib1@126.com

Received 20 November 2017; Revised 26 April 2018; Accepted 6 May 2018; Published 25 June 2018

Academic Editor: Taku Fujiyama

The urban rail transit line operating in the express and local train mode can solve the problem of disequilibrium passenger flow and space and meet the rapid arrival demand of long-distance passengers. In this paper, the Logit model is used to analyze the behavior of passengers choosing trains by considering the sensitivity of travel time and travel distance. Then, based on the composition of passenger travel time, an integer programming model for train stop scheme, aimed at minimizing the total passenger travel time, is proposed. Finally, combined with a certain regional rail line in Shenzhen, the plan is solved by genetic algorithm and evaluated through the time benefit, carrying capacity, and energy consumption efficiency. The simulation result shows that although the capacity is reduced by 6 trains, the optimized travel time per person is 10.34 min, and the energy consumption is saved by about 16%, which proves that the proposed model is efficient and feasible.

1. Introduction

With the continuous expansion of urban scale and the development of urban rail transit network, the round-trip passenger flow between city centers and suburbs grows gradually. As the central city is saturated basically, more and more companies and residential buildings tend to transfer to the suburbs, which further promotes the swelling of the traffic flow. Besides, due to the extension of urban rail transit line, it forms long-distance regional rail transit lines [1]. If the train stops frequently, it will inevitably cause the increase of travel time, thereby reducing the operational efficiency and being not conductive to the efficient development of rail transit. And long-distance commuters want to arrive their destinations as soon as possible and passengers at stations along the route wish to reach the destinations without transfer. Therefore, cities need to use express and local train operating mode to meet the demands for the commuters [2]. On the one hand, from the passenger's point of view, this model can shorten the travel time of commuting passenger flow and improve the passenger service level, but at the same time, it will increase the travel time of some local passengers to a certain extent and affect their travel experience. On the other hand, for operating enterprises, the mode will save the train energy consumption, thereby enhancing operational efficiency, but the system carrying capacity will be reduced. For this reason, how to ensure the stop scheme of express trains reasonably and save the passenger travel time to an extreme is a significant project at present.

In view of the demand for commuters, scholars mainly study the mode of the express and local train and the plan of the full-length and short-run route. Based on the express and local train operation mode, the literature mainly considers the travel demand of passengers. For instance, L. Wang [3] considered the shortest total travel time of the passengers to build up an optimization model. R. Tang [4] constructed a linear programming model with the goal of maximizing passenger satisfaction, which is reducing average passenger waiting time. Hassannayebi E et al. [5] studied the optimal time table of express/local train and built up

a model in order to reduce the passenger waiting time by using variance and penalty function. Chun J H et al. [6] come up with the express/local train mode, which can solve the problem of passenger crowding and train delay with building the model of minimizing the number of serviceable cars. Y Lai et al. [7] took transfer passengers into account and then established an innovative network structure to study the train stop plan. Y. Gao et al. [8] proposed a biobjective programming model to achieve the simultaneous minimization of the energy consumption and the travel time. However, the above literature sources do not consider the travel process comprehensively. In the combination mode, there is a situation that express trains cross through local trains. And this part of time should not be ignored. Therefore, Canca D et al. [9, 10] also mainly considered meeting the travel demand of the passenger flow from the passenger angle by adjusting the train schedule including the overtaking factor. However, they ignored the commuting passengers' sensitivity to travel time, whose travel purpose is to arrive the destination quickly. Therefore, D. Di et al. [11, 12] established a passenger path-choice model to determine the passenger principle in path choice by considering perception difference of time without considering the effect of travel distance.

In the condition of the full-length and short-run route, the recent studies mainly focused on the selection of short-run route, service level, traffic capacity, and rolling stock [13]. For example, J. Q. Liao [14], respectively, devised the operation scheme in condition of the single route and the full-length and short-run route aiming at the maximum travel saving time of all passengers. Q. Hu et al. [15] analyzed the section of short-run route from the use of the rolling stock and capacity. However, the operation scheme plan of mixed-routing cannot save the travel time for long-distance passengers and solve the problem of commuting effectively.

To sum up, based on the analysis of the composition of passenger travel time and passenger selection behavior, this paper makes a reasonable stop scheme of express trains aiming to reduce the total travel time to maximize the operational efficiency of express/local train for regional rail transit line.

The remainder of the paper is organized as follows. Section 2 analyzes the passenger flow characteristics and the effect of travel time and carrying capacity in the condition of express and local train operating mode. In Section 3, a stopping plan optimization model of trains based on minimizing the total passenger travel time is established, and then the model is solved by genetic algorithm. In Section 4, an example of a regional line in Shenzhen is provided to optimize the train stop plan by using the model and algorithm above and evaluate it by the travel time, the capacity, and the energy consumption compared with the general plan. The final section draws the conclusions and proposes some issues for future.

2. Analysis of Passenger Flow Characteristics and Effects

2.1. Passenger Flow Characteristics. The long-distance regional rail transit line is connected in series with urban areas, suburbs, and satellite towns, and the intercity passenger flow is not only large, but also imbalanced in time direction. During the morning rush, the traffic flow moves from satellite towns or suburbs to the downtown, and at the evening peak, the situation is opposite. Meanwhile, with different levels of development along the rail transit, the passenger volume of boarding, and alighting at stations are also inconsistent, and overall, they show the disproportion of the section flow. Consequently, two types of the fluctuating passenger flow have demands for regional express rails: commuters prefer to take the through train from downtown area to the suburban, while short-distance passengers want to reach any two stations directly.

2.2. Travel Time Cost. Although operating express/local train prolongs the total trip time of passengers who take local train because of overtaking, it saves the time of those passengers who take express train.

2.3. Carrying Capacity. Express and local train scheme will reduce the capacity of lines [16]. In the traditional pattern, the line capacity n_r is only related to the minimum headway I, and its calculating method is as formula (1). However, the carrying capacity of express/local train takes the deduction coefficient method for calculation and analysis. The deduction coefficient of the express train refers to the number of the local cars deducted from the parallel operation chart because of drawing an express train, and the formula is as (2) [17].

$$n_r = \frac{3600}{I} \tag{1}$$

$$\varepsilon_e = 1 + \left(\frac{t_d}{I}\right) - k \tag{2}$$

where ε_e is the deduction coefficient of the express train, t_d is the difference of travel time of express/local train, and k means the times which are crossed of slow trains.

2.4. Train Energy Consumption. In daily operation, in addition to the maintenance of infrastructure, the energy consumption during the train operation also accounts for a large proportion [18]. And compared to the normal train mode, where trains stop at all stations, the express train only stops at some stations, so the operating scheme of the express and local train reduces the energy consumption of traction and braking at start and stop, and it can achieve the energy saving and emission reduction. In terms of the operating enterprise, it has great benefit [19]. And the equations of energy consumption calculation and parameters (Table 1) are as follows:

$$E = E_t + E_b \tag{3}$$

$$E_t = \sum_{m=1}^{f_1+f_2} \sum_{i=1}^{n} \left(\int_0^{l_t} F_t dl \right) \tag{4}$$

$$E_b = \sum_{m=1}^{f_1+f_2} \sum_{i=1}^{n} \left(\int_0^{l_b} F_b dl - E_r \right) \tag{5}$$

$$E_r = \sum_{m=1}^{f_1+f_2} \sum_{i=1}^{n} \sum_{k=m+1}^{f_1+f_2} x_{m,k} \cdot \int_0^{l_t} F_b dl \cdot \xi \cdot \frac{t_c}{t_b} \tag{6}$$

where (3) means the total energy consumption, (4) is the train traction energy consumption, (5) is the energy consumption of train braking, and (6) shows the energy saved in regenerative braking.

3. Optimization Model and Algorithm

3.1. Problem Description. First, operating express/local train combination plan is mainly for commuting problem, so this paper chooses the morning peak as the optimization period to study. Second, two kinds of trains own the same technical parameters. Third, although the line is divided into the upstream and downstream direction, the model only discusses the one-way situation where the passenger flow is larger. And then, the regional rail transit line which this paper studies can meet the demand of passenger flow without the situation of passenger retention. The last assumption is that passengers' arriving can all be considered to follow the normal distribution.

3.2. Modeling. Passenger travel time t includes waiting time t_w and running time t_r like formula (7), in which the total passenger waiting time consists of two parts: the first part is the normal passenger waiting time when the foregoing local train is not overtook, and the second part is the increased passenger waiting time at stations caused by overtaking. And t_r includes the interval running time and station dwell time.

$$t = t_w + t_r. \tag{7}$$

3.2.1. Analysis of Passenger Selecting Behavior for Train. The research object is commuting passenger flow with early morning rush hour. Considering that passengers are generally unwilling to transfer, this paper assumes that passengers at most experience one transfer in the same line and same direction. Besides, when OD are both express train stations, passengers will not choose the transfer way from express to local train [20]. Therefore, according to the different nature of origin and destination, all kinds of the travel selection behavior are shown in Table 2.

The behavior of passengers selecting trains can be analyzed by Logit model like formula (8) [21]. And (9) shows the passenger flow who selects some kind of travel mode [22].

$$p_{ij}^{(k)} = \frac{\exp\left(-\theta \cdot c_{ij}^{(k)}\right)}{\sum_{k \in S} \exp\left(-\theta \cdot c_{ij}^{(k)}\right)} \tag{8}$$

where $p_{ij}^{(k)}$ is the passenger ratio choosing the mode k to travel from station i to j, θ is the passenger expectation variable based on travel impedance which is inversely proportional to the passenger expectation, $c_{ij}^{(k)}$ is the impedance choosing the mode k to travel from station i to j, and S is collection of travel modes.

$$q^{(k)} = q_{ij} \cdot P_{ij}^{(k)} \tag{9}$$

where $q^{(k)}$ is the passengers choosing mode k to travel and q_{ij} is the passengers flowing from i to j.

Passenger travel impedance is mainly affected by both the train running time t_y and travel distance l, and the combined impedance is determined by the weight of both.

(1) Train Running Time. When passengers travel from local to express station or just the reverse, they need to think about transferring. The principles are as follows.

(1) $t_{(trans)} > t_{(l)}$, where $t_{(trans)}$ is the total time of changing trains and $t_{(l)}$ only refers to the time of local train. In addition, the transfer time is as

$$t_{(trans)} = t_w^1 + t_r^1 + t_w^2 + t_r^2 \tag{10}$$

where t_w^1 and t_r^1, respectively, represent the waiting time and running time including dwell time of the first train for passenger; t_w^2 and t_r^2, respectively, represent the same meaning for the second train.

(2) When passengers consider taking local train to express, they not only need to take the total time into account, but also need to think of whether the middle has express station to supply to transfer.

(2) Travel Distance. The operation ratio of express and local train can be estimated by the number of passengers whose travel distance is greater than five sections and less than five sections, like the following formula [23]:

$$p = \frac{\sum_{|i-j|>5} q_{ij}}{\sum_{|i-j|\leq 5} q_{ij}}. \tag{11}$$

From the above two parts, the combined impedance is as (12) and the flow chart of passengers train-choice behavior is as shown in Figure 1.

$$c_{ij}^{(k)} = \mu_1 \cdot t_y + \mu_2 \cdot l \tag{12}$$

where μ_1 and μ_2 are the weight of train running time and travel distance, and they depend on the specific situation.

3.2.2. Objective Function. Total travel time t is calculated as formula (13) and the model parameters are listed in Table 3.

$$t = \left(\sum_{m=1}^{f_1} q_{ij-m}^{(l)} + x_q \cdot \sum_{m=1}^{f_2} q_{ij-m}^{(e)} \right) \cdot \left(\frac{\mathrm{T}}{2 \cdot (c_1 f_1 + c_2 f_2)} + \sum_{k=1}^{i-1} c_1 h_k x_k + t_{ij} + \sum_{w=i+1}^{j-1} \left(t_s + c_1 x_w h_w \right) \right). \tag{13}$$

In conclusion, when the number of stations is n, the paper establishes the following objective function to achieve the minimization of the total passenger travel time as follows:

$$\begin{aligned} \min Z &= \min \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} t \right) \\ s.t. \quad & q_{ij-m} \leq \eta_{\max} \cdot a \\ & f_d \geq 1 \quad (d = 1, 2, 3). \end{aligned} \tag{14}$$

Table 1: Energy consumption model parameters.

Parameters (unit)	Explanation
E (KJ)	The total energy consumption
E_t (KJ)	The traction energy consumption
E_b (KJ)	The braking energy consumption
E_r (KJ)	The regenerating energy
l_t (m)	The distance of traction
F_t (KN)	The net force of traction process
$x_{m,k}$	Boolean variable. When the m^{th} and k^{th} train are in the same power supply section, the value is 1; otherwise, the value is 0.
l_b (m)	The distance of braking
F_b (KN)	The net force of braking process
ξ	The energy transformation ratio
t_c (s)	The overlapping time of the acceleration time of the m^{th} and the braking time of the k^{th}.
t_b (s)	The braking time of the k^{th} train.

Table 2: Passenger selection behavior.

OD	Local train station	Express train station
Local train station	Local train	Local train
		Local train→Express train
Express train station	Local train	Local train
		Local train→Express train
	Express train→Local train	Express train

Table 3: Model parameters.

Parameters (unit)	Explanation
f_1 (time/h)	The operating frequency of the local train
f_2 (time/h)	The operating frequency of the express train
x_q	Boolean variable. When the i station and j station are both local stations, the value is zero.
$q_{ij-m}^{(l)}$ (person)	The number of passengers waiting for the m^{th} train which is the local train
$q_{ij-m}^{(e)}$ (person)	The number of passengers waiting to the m^{th} train which is the local train
T (min)	The length of the optimized period
c_1/c_2	Boolean variable. When passengers take the local train, the value of c_1 is one, c_2 is zero, and when passengers take the express train, the value of c_1 is zero, c_2 is one.
h_k (min)	The increased delay time because the slow train is crossed by the express train at preceding stations
x_k	Boolean variable. When the slow train is crossed by the express train at the k^{th} station, the value is one; otherwise, the value is zero.
t_{ij} (min)	The interval running time of trains
t_s (min)	The dwelling time at stations
x_w	Boolean variable. When the local train is crossed by the express train at the w^{th} station, the value is one; otherwise, the value is zero.
h_w (min)	The increased dwell time in the trains because the local train is crossed by the express train at this station
q_{ij-m} (min)	The number of passengers waiting for the m^{th} train
$\eta_{\max}$	Maximum section load factor
a	Passenger capacity of the train

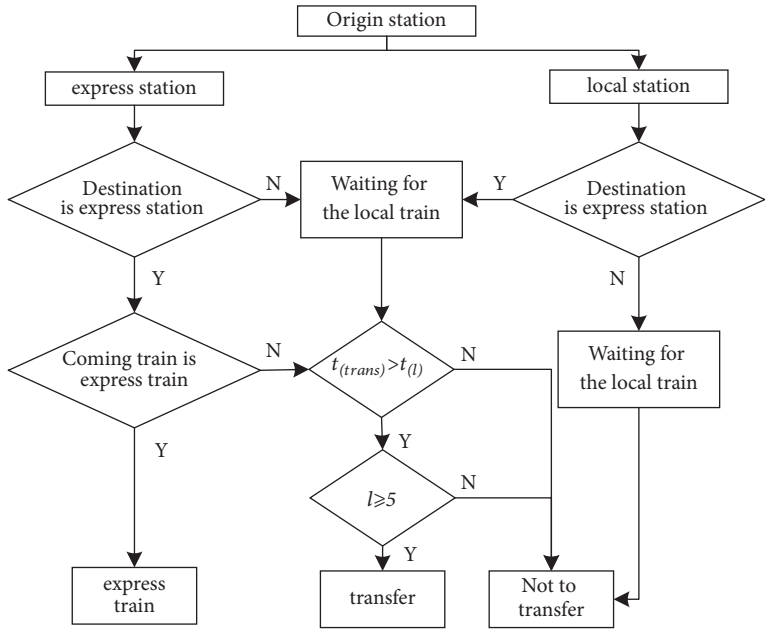


Figure 1: The flow chart of passenger train-choice behavior.

3.3. Solution Algorithm. The evolutionary algorithm, which is represented by genetic algorithm, uses crossover and mutation operator to realize the information exchange and local search between individuals in the group and provides each individual with the opportunity of optimization [24]. Through the competitive selection mechanism of survival of the fittest, the evolutionary algorithm guides the population to the better. Exactly, the unique binary coding rule of genetic algorithm is suitable for the value of 0-1 of the train stop strategy. Besides, instead of starting from a point and optimizing according to a fixed route, it searches the whole feasible solution space at the same time, so it can avoid falling into a local optimum and find the global optimal results. The steps are as follows.

(1) Population initialization. It generates N binary coded individuals randomly, and the length is the number of stations, where in code, 1 means that express train will stop at the corresponding station and 0 means no stopping. What is more, because the express train must stop at the original and terminal stations, the encoding is 1.

(2) Ensuring the fitness function. It constructs the fitness function as formula (15). The greater the fitness value, the better the individual.

$$u(x) = M - f(x) \tag{15}$$

where M means the total time of general mode and $f(x)$ is the time of combination mode with every plan.

(3) Taking the operations of selecting, crossover, and mutation. The selection operation adopts roulette method, the crossover operator is uniform crossover, and the mutation operator is a single point mutation.

(4) After an iteration, the worst individual is replaced by the best individual in the current generation, that is also called "survival of the fittest".

Figure 2: The regional rail transit line in Shenzhen.

(5) With the requirement of iterations, steps 4 and 5 are repeated.

4. Case Analysis

4.1. Line Profile. Taking a certain line in Shenzhen as an example (Figure 2), this paper analyzes and verifies the optimization of the express and local train stop problem. The data required in the example analysis is shown in Table 4, the OD passenger flow during the morning peak (8:00~9:00) is shown in Table 5, and the length of interval and pure runtime of trains are followed in Table 6. Besides, according to the OD data, the passenger flow of boarding and alighting and the distribution of section flow of up direction (GXB→HZN) and down direction (HZN→GXB) are, respectively, shown in Figures 3, 4, and 5.

Table 4: Values of case parameters.

Parameters (unit)	Value	Parameters (unit)	Value
Length (km)	53.07	Train composition	A8
The number of stations	18	Dwell time (s)	30
Minimum headway (s)	120	Maximum running time (km/h)	120
Passenger capacity (person)	2480	Maximum section load factor	100%
Passenger expectation variable θ	0.9	Weight μ_1	0.6
Weight μ_2	0.4		

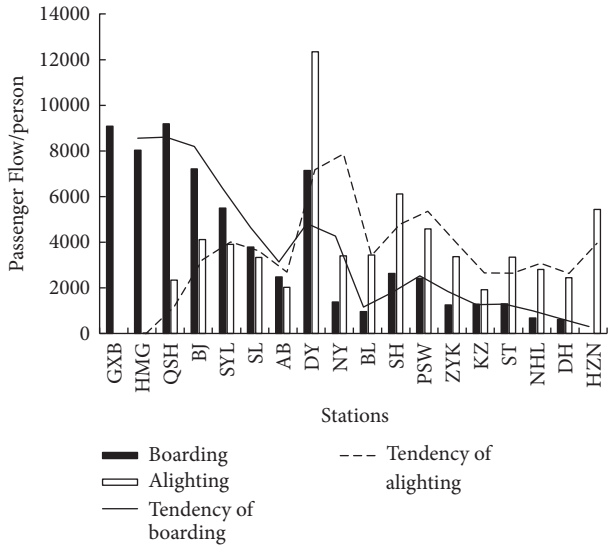


Figure 3: The passenger flow of boarding and alighting of up direction.

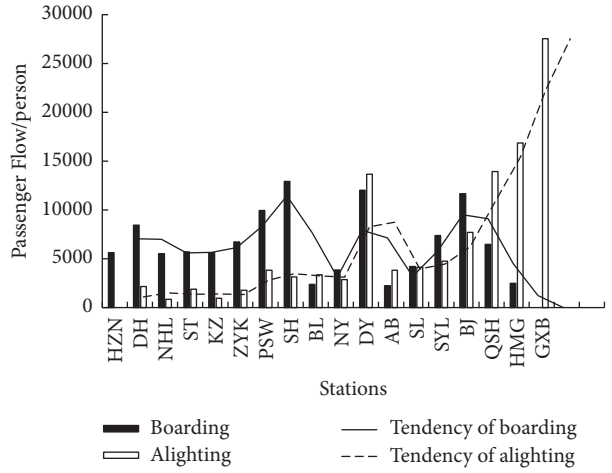


Figure 4: The passenger flow of boarding and alighting of down direction.

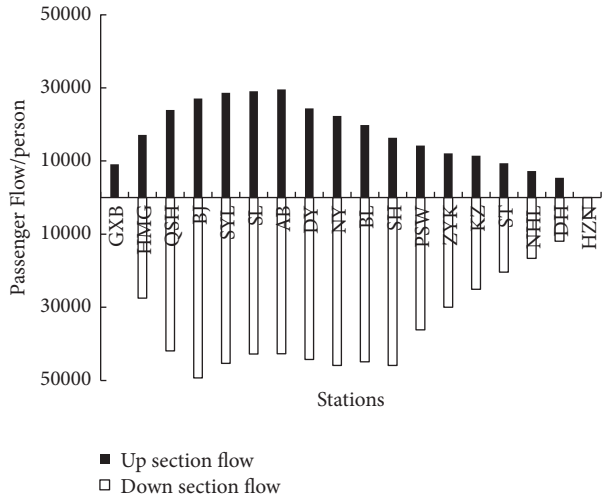


Figure 5: The distribution of section passenger flow.

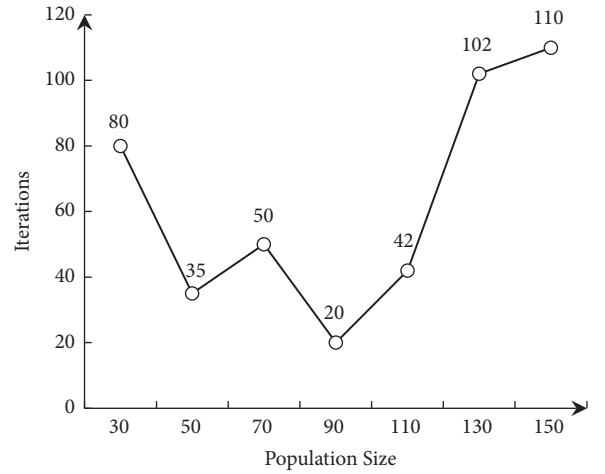


Figure 6: The relationship between the population size and number of iterations.

4.2. Solution. This paper uses the genetic algorithm to solve the model. Different parameters will affect the speed and results of the algorithm. The paper uses two of the population sizes, crossover probability and mutation probability, as the independent variables and the other two as the dependent variables to analyze the influence of the values on algorithm results. When the values of the population size were from 30 to 150, the results showed that the final fitness value did not change with the population size, while the rate of convergence will be affected as shown in Figure 6. And according to the current literature that the range of the crossover rate is 0.4 to 0.99 and 0.001 to 0.1, respectively [25], the paper supposes that the crossover rate steps and the mutation rate steps are 0.5 and 0.005. The results show that although the number of

Table 5: The value of OD passenger flow during the morning rush hour.

Station	GXB	HMG	QSH	BJ	SYL	SL	AB	DY	NY	BL	SH	PSW	ZYK	KZ	ST	NHL	DH	HZN
GXB	0	1133	1708	1464	597	420	179	2661	261	161	525	391	224	111	118	66	63	131
HMG	2480	0	624	1179	751	462	211	2509	282	185	580	419	238	120	122	74	147	132
QSH	5399	1085	0	1473	1941	543	325	1883	361	322	569	435	276	130	158	129	461	173
BJ	5914	3702	2044	0	621	961	514	1740	440	358	858	629	342	142	172	85	172	183
SYL	1763	1686	2959	945	0	954	488	1190	453	422	658	463	284	115	148	97	66	156
SL	935	807	777	843	829	0	307	1334	338	370	390	263	206	93	126	128	71	164
AB	444	391	395	417	337	246	0	1024	215	242	271	195	139	63	85	87	52	111
DY	3774	2956	1565	1030	528	939	1225	0	1054	1204	1717	918	739	204	456	230	146	474
NY	526	468	446	380	322	267	242	1193	0	180	318	284	162	68	98	93	57	126
BL	293	252	250	217	182	155	140	782	96	0	231	254	122	56	74	71	48	97
SH	1844	1725	1405	1229	766	561	524	3484	628	757	0	337	338	302	622	365	195	475
PSW	1116	1074	999	721	628	396	427	2680	669	853	391	0	302	317	489	433	247	643
ZYK	932	833	679	575	362	288	248	1549	316	389	270	280	0	206	353	240	129	324
KZ	428	375	419	259	186	284	247	868	279	349	568	913	430	0	324	317	175	471
ST	444	382	399	267	183	227	204	892	234	295	767	714	469	211	0	387	245	674
NHL	431	367	421	226	150	272	208	790	243	270	443	707	321	222	443	0	183	497
DH	562	576	971	426	196	352	259	980	307	328	511	852	385	333	834	572	0	610
HZN	226	186	192	156	83	117	90	443	108	116	188	354	167	174	603	274	2138	0

Table 6: The length of interval and pure runtime of trains.

Interval	Length (km)	Runtime (min)	Interval	Length (km)	Runtime (min)
GXB-HMG	2.905	2.563	BL-SH	4.213	2.922
HMG-QSH	4.279	3.363	SH-PSW	1.566	1.521
GSH-BJ	2.494	2.113	PSW-ZYK	2.007	1.714
BJ-SYL	3.535	2.84	ZYK-KZ	4.926	3.339
SYL-SL	6.364	4.173	KZ-ST	3.229	2.367
SL-AB	3.428	2.503	ST-NHL	1.623	1.634
AB-DY	2.413	1.915	NHL-DH	1.745	1.795
DY-NY	6.037	4.35	DH-HZN	1.238	1.241
NY-BL	2.072	1.756			

Table 7: Values of algorithm parameters

Parameters	Value	Parameters	Value
Population	50	Crossover probability	0.8
Iterations	150	Mutant probability	0.05

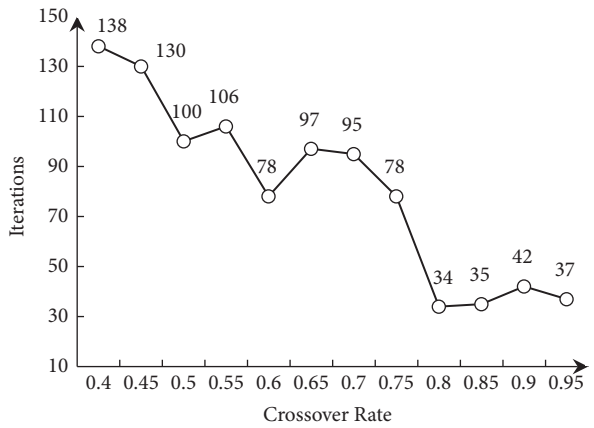


Figure 7: The relationship between the crossover rate and number of iterations.

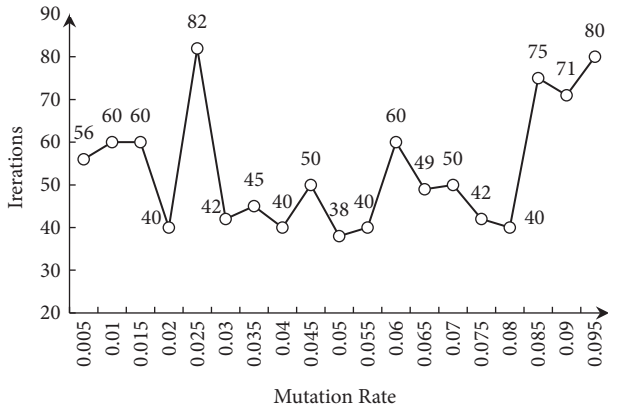


Figure 8: The relationship between the mutation rate and number of iterations.

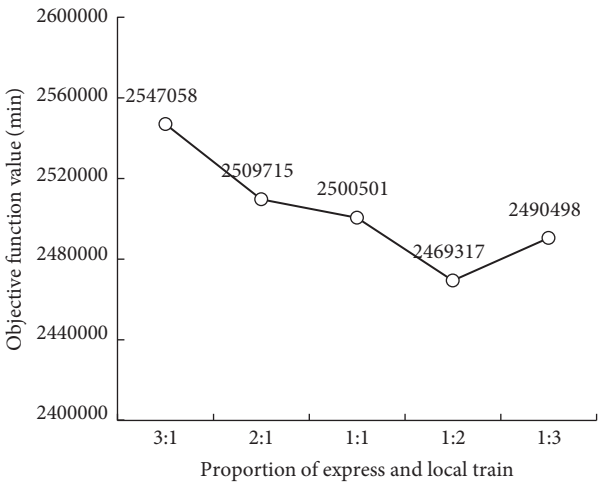


Figure 9: Variation of total travel time.

iterations changes, iterations have little difference from each other and the final fitness value is invariable as Figures 7 and 8 show. Therefore, the final algorithm parameter values are shown in Table 7.

According to the case data, the maximum cross-section passenger flow appears at the BJ station in the downward direction, with 49333 people, so that the number of departure times during the optimization period is 20 pairs/h. As the downward direction of the cross-section of the passenger flow is much larger than the upward direction, it is designed to operate the express and local train mode in the down direction. And in this situation, the optimal result is different with the different proportion of the express/local train. Based on the OD in Table 5 and formula (11), the ratio is approximately 1:2. However, in order not to leave out all the plans, the paper takes 3:1, 2:1, 1:1, 1:2, 1:3 as alternatives.

The case is calculated and analyzed according to the algorithm parameter in Table 7 and fitness function set as (15). Figure 9 shows the changing of travel time value. When the proportion is 1:2, that is to say, operating 6 express trains and 14 local trains, the passenger travel time is optimal. Corresponding express station stop scheme is shown in Table 8, in which 1 stands for stopping at station and 0 for no stopping, and express train overtakes the local at station ST and PSW without stopping.

Table 8: The best express train stop scheme.

Station	HZN	DH	NHL	ST	KZ	ZYK	PSW	SH	BL	NY	DY	AB	SL	SYL	BJ	QSH	HMG	GXB
Plan	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	1

Table 9: Total travel time comparison.

Compared content (unit)	General mode	Combination mode	Time benefit	Per capita benefit
Total travel time (min)	3 638 656	2 469 317	1 169 339	10.34 min

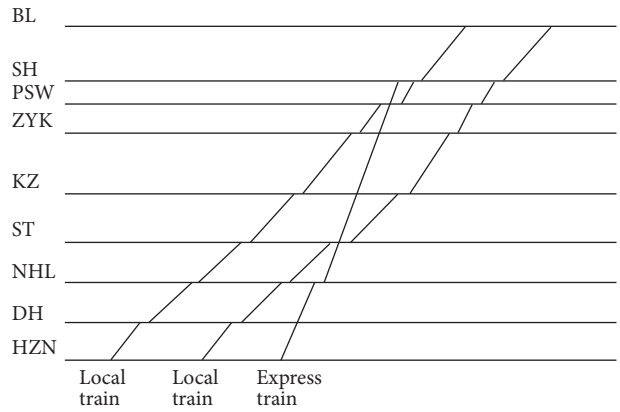


Figure 10: The sketch of train operation.

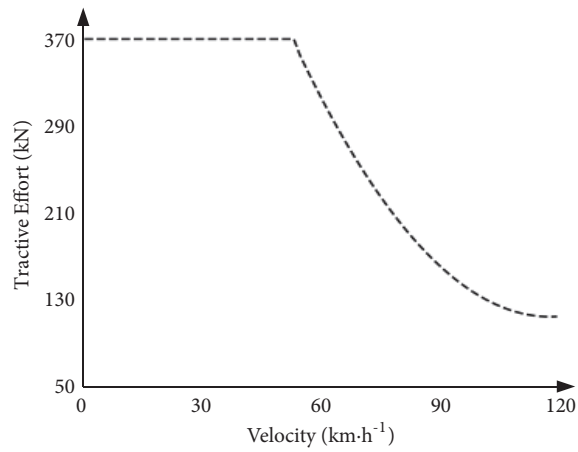


Figure 11: The sketch of traction curve.

4.3. Results Analysis

4.3.1. Time Benefit Analysis. Compared with the general mode which only operates the local trains, the combination saves the passenger travel time with taking the express trains, whereas part of the passengers' time is increased because of slow trains. In the condition of the same train frequency, comparing the total travel time before and after the plan, like Table 9, we can see that in the specific passenger flow conditions, the time benefit is quite obvious.

4.3.2. Capacity Analysis. It has an effect on carrying capacity due to operating the express/local train. According to the parameters in the foregoing case, when the minimum headway is 120s and the ratio of express and local train is 6:14 (roughly 1:2), where the capacity of general mode is 30 pairs/h, while in condition of the combination mode, the capacity depresses to 24 pairs/h.

From the stop scheme, we draw a train graph like Figure 10. With not considering the train turnaround temporarily, here we only draw the sketch of two sets of express/local train combination which run from the station HZN to the BL, and the start order is the local train, local train, and the express train, respectively. In a triple of the combination, the first local train is crossed at station PSW, and the second is overtaken at ST without others.

4.3.3. Energy Consumption Comparison. The difference between the energy consumption of express trains and slow trains is mainly caused by the express train running directly at some stations without stopping and experiencing a kind of uniform speed at a longer distance in the interval. The train traction curve, brake curve, and speed curve diagrammatic drawing are as shown in Figures 11, 12, and 13, respectively. Traction energy consumption and braking energy consumption are calculated by (4) and (5), and (6) represents the regenerative braking energy. Above all, the total consumption is as formula (3). In this case, there are three substations along the rail line. Under the above optimal stopping plan and the series calculation of train energy consumption, the results are compared in Table 10. It shows that the express/slow train mode has great advantages in energy saving and ensures the efficiency of the operating enterprises.

5. Conclusion

This paper studied the problem of the train stopping schedule in the condition of express and local train operation mode. With the analysis of the passengers' requirement to minimize the travel time, travel distance, and mainly the stations, the paper established a passenger train-choice model and used the entropy method to calculate the weight coefficient scientifically. On this basis, a 0-1 integer programming mathematical model aimed at minimizing the total travel time was set up, and the genetic algorithm was designed to solve the model. Finally, the algorithm result was insensitivity with the parameters. Besides, the travel time and energy consumption were both saved effectively through the case analysis, resulting in a considerable social benefits. However, the combination of express and local train operation reduced the line capacity, and the actual operation can change the ratio of express/local train to get relatively large capacity.

Table 10: Train energy consumption comparison.

Compared content (unit)	General mode	Combination mode	benefit	Benefit rate
Energy consumption (KJ)	1.41×10^8	1.19×10^8	2.26×10^7	16%

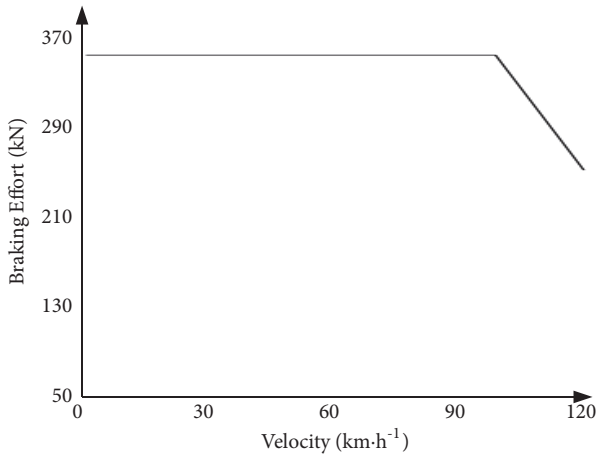


Figure 12: The sketch of brake curve.

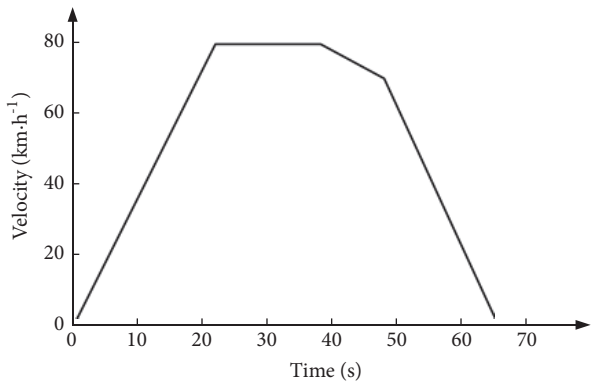


Figure 13: The sketch of speed curve.

Conflicts of Interest

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

Acknowledgments

This paper is supported by the Research Projects of the Social Science and Humanity on Young Fund of the Ministry of Education under Grant no. 15YJCZH108, the Research Projects of Natural Science Foundation of Guangdong Province under Grant no. 2015A030310341, and the Research Project of Shenzhen Technology University.

References

[1] Z. Rashedi, M. Mahmoud, S. Hasnine, and K. N. Habib, "On the factors affecting the choice of regional transit for commuting in Greater Toronto and Hamilton Area: Application of an advanced RP-SP choice model," *Transportation Research Part A: Policy and Practice*, vol. 105, pp. 1–13, 2017.

[2] C. Hirai, T. Kunimatsu, S. Kondou, N. Tomii, and M. Takaba, "Modelling and solving the train stop deployment planning problem," in *Proceedings of the 12th IASTED International Conference on Artificial Intelligence and Soft Computing, ASC 2008*, pp. 67–72, esp, September 2008.

[3] L. Wang and G. J. Wei, "Stopping Schedules of the Express /Local Train for Urban rail Transit Based on GASA Algorithm," *Urban Mass Transit*, vol. 18, no. 3, pp. 96–100, 2015.

[4] R. Tang and Y. Gao, "A Model of Customer Satisfaction-based for Express/Local Train Operation," *Modern Urban Transit*, vol. no. 5, pp. 59–64, 2015.

[5] E. Hassannayebi, S. H. Zegordi, M. R. Amin-Naseri, and M. Yaghini, "Train timetabling at rapid rail transit lines: a robust multi-objective stochastic programming approach," *Operational Research*, vol. 17, no. 2, pp. 435–477, 2017.

[6] 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 14th IEEE International Intelligent Transportation Systems Conference, ITSC 2011*, pp. 1586–1591, usa, October 2011.

[7] Y.-C. Lai, M.-C. Shih, and G.-H. Chen, "Development of efficient stop planning optimization process for high-speed rail systems," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 1802–1819, 2016.

[8] 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 & Environment*, 2016.

[9] D. Canca, E. Barrena, E. Algaba, and A. Zarzo, "Design and analysis of demand-adapted railway timetables," *Journal of Advanced Transportation*, vol. 48, no. 2, pp. 119–137, 2014.

[10] Q. Zhen and S. Jing, "Train rescheduling model with train delay and passenger impatience time in urban subway network," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 1990–2014, 2016.

[11] D. Di and D. Yang, "Passenger flow analysis model about express/slow train in urban rail transportation corridor," *Tongji Daxue Xuebao/Journal of Tongji University*, vol. 42, no. 1, pp. 78–83, 2014.

[12] X. M. Zhao and Q. X. Sun, "Passenger Choice Behavior for Regional Rail Transit under Express/Local Operation with Overtaking," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 5, pp. 104–109, 2016.

[13] R. Xu and X. Li, "Optimization of Routing Mode on Regional Express Rail," *Urban Mass Transit*, vol. 9, no. 5, pp. 36–39, 2006.

[14] J. Q. Liao, *Research of Regional Rail Transit Express Train Plan*, Beijing Jiaotong university, 2016.

[15] Q. Hu, "Study on Mixed-routing in Railway Transit for Disequilibrium Passenger Flow," *Value Engineering*, vol. 31, no. 32, pp. 81–83, 2012.

[16] Y. Zhu, B. Mao, Y. Bai, and S. Chen, "A bi-level model for single-line rail timetable design with consideration of demand and capacity," *Transportation Research Part C: Emerging Technologies*, vol. 85, pp. 211–233, 2017.

[17] M. Y. Qu and S. M. Huang, "Study on express/local train plan of urban rail transit," *Railway Transport and Economy*, vol. 34, pp. 77–82, 2012.

[18] Q. Peng, W. Li, Y. Wang, Q. Zhong, and J. Sun, "Study on Operation Strategies for Metro Trains Under Regenerative Braking," *Tiedao Xuebao/Journal of the China Railway Society*, vol. 39, no. 3, pp. 7–13, 2017.

[19] K. Nomura and M. Miyatake, "Scheduling method for minimum energy consumption considering constraints of time intervals between local and express trains," in *Proceedings of the 2016 IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, pp. 469–475, gbr, August 2016.

[20] X. M. Zhao, Q. X. Sun, Y. Ding, R. J. Shi, and Y. Chen, "Passenger choice behavior for regional rail transit under express/local operation with overtaking," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 5, pp. 104–109, 2016.

[21] Y. Wang, T. Tang, B. Ning, T. J. J. van den Boom, and B. De Schutter, "Passenger-demands-oriented train scheduling for an urban rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 1–23, 2015.

[22] X. Wang, H. Wang, and X. Zhang, "Stochastic seat allocation models for passenger rail transportation under customer choice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 96, pp. 95–112, 2016.

[23] L. R. Huang, *Research on operation organization of urban rail transit of express/local train*, Southwest Jiaotong University, 2017.

[24] H. Niu and X. Zhou, "Optimizing urban rail timetable under time-dependent demand and oversaturated conditions," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 212–230, 2013.

[25] W. Li, R. Xu, and W. Zhu, "Multi-line cooperation method for passenger flow disposal in metro transfer station under train delay," *Tongji Daxue Xuebao/Journal of Tongji University*, vol. 43, no. 2, pp. 239–244, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 4264528, 18 pages
https://doi.org/10.1155/2018/4264528

Research Article

On Individual Repositioning Distance along Platform during Train Waiting

Fabien Leurent and Xiaoyan Xie

LVMT, UMR-T 9403, Ecole des Ponts, IFSTTAR, UPEM, UPE, 77455, Champs-sur-Marne, France

Correspondence should be addressed to Xiaoyan Xie; xiaoyan.xie@enpc.fr

Received 24 November 2017; Revised 17 April 2018; Accepted 3 May 2018; Published 21 June 2018

Academic Editor: Taku Fujiyama

Out of waiting times spent in rail stations on boarding platforms, some part can be reinvested by the trip-makers to optimize their positions of boarding and save on travel time for the rest of their trips. This paper provides a stochastic model, in which user's journey is decomposed into phases of, successively, walking in the access station, platform positioning, waiting for boarding, train riding, and walking in the egress station. Walking speed and target position are modeled as individual factors, and in-station distances as random variables. Service timetable is exogenous. This makes egress times and exit instants random variables that are characterized by distribution and mass probability functions under closed-forms, for both single and distributed walking speeds. Specific statistical distributions are shown to ease computation. The resulting PDF formulae make likelihood functions of the model parameters. Maximum likelihood estimation is proposed and applied to a case study of commuter rail line in Paris: journeys between stations Vincennes and La Défense along line A of the Regional Express Railways. Based on data from Automated Fare Collection and Automatic Vehicle Location systems and pertaining to an individual user, satisfactory results were obtained.

1. Introduction

Passenger waiting time is one of the most crucial factors for transit design and planning, related to passenger satisfaction and to measure public transport (PT) quality of service [1–4]. However, it was significant at the trip level in urban PT, and the ratio of passenger actual waiting time to actual journey time was between 10 and 30% [5–7]. Furthermore, waiting time is most often spent at standing, often so in a crowded place. That makes passenger less comfortable. In the literature, social sciences tried to reduce perceived or actual waiting time, or emotional and psychological costs, whereas operations research investigated waiting time reduction strategies [8–12], as waiting time could be considered scarce, economic resource and unavoidable [13, 14]. In order to maximize time utility, some passengers allocate waiting time to activities/preoccupations. However, beyond the waiting time investigations, existing research does not sufficiently explore how to reuse the actual waiting time to perform an efficient travel, especially in high-frequency urban rail PT system. That has received only some attention in cognitive aspect [10, 15–18].

Despite the attention paid so far to individual waiting time reuse in PT station, a related issue seemed to remain unexplored: the reuse of waiting time for passenger repositioning along the boarding platform. Of course, this is of interest for railway-based transit submodes only, since the platform lengths of these modes extend from some dozen meters up to some hundred meters, e.g., up to 200 m in Paris. Assuming that individual walking speeds range from 0.8 to 1.7 m/s, repositioning time from one platform end to another may reach 0.5 to 3.5 min, indeed a valuable gain for passengers in their daily schedules. The influence of waiting time on individual walk on boarding platform can be significant. Such extent makes the "longitudinal" positioning of individual passengers along their trains a matter of significance for their journeys, since it involves walking times along the platform at both access and egress stations, together with waiting time and comfort on the boarding platform as well as in-vehicle in relation to the distribution of passengers along the train. This

pushed frequent passengers (e.g., commuters) to optimize their longitudinal position along the boarding platform with respect to their egress station [19], and also "travel assistant" applications (such as CityMapper, RATP, Paris ci la sortie du Métro) to deliver positioning advice to every passengers, including occasional ones (e.g., tourists).

Railway operators, on their side, consider the longitudinal positioning of candidate riders in relation to train dwelling times: higher passenger densities at some spots along the platform will not only require more boarding time but also slow down the egress of alighting passengers. Dynamic information via travel assistants or specific signage such as with variable color panels has recently been implemented on an experimental basis [20]. From a recent survey in the Paris area of passengers boarding in a suburban train station about their eagerness to reposition themselves in relation to platform crowding, it was shown that the acceptable shift amounts to 2.4 cars on average, about 50 m [19].

The objective of this paper is to estimate passenger longitudinal repositioning distance on boarding platform during train waiting and underlying distributions of individual walking speed and distances. Building upon our previous stochastic models [21–23] based on smartcard data, a stochastic modeling approach about passenger individual repositioning distance with posterior analysis is proposed in this paper, by extending our primary model about passenger longitudinal repositioning [24].

Concerning the proposed passenger flow stochastic model, over the last decade, a new branch of knowledge was emerged, the modern data measurement and data-driven statistical analysis in PT field. Modern data, Automated Fare Collection (AFC) data and Automatic Vehicle Location (AVL) data, among others have become available either for PT service evaluation and improvement, or for passenger flow analysis. Passenger flow study has given rise to the stochastic modeling and statistical estimation of fine individual passenger travel phenomena by trip leg or in station in rail transit system, a closed black-box. Several relevant studies are reviewed in greater details as below; for other applications about AFC data see [25, 26].

The stochastic features of passenger journeys were modeled by [27], who decomposed every journey into separate stages, walking-in, wait, walking-out, and transfer between Origin-Destination (O-D) pairs. A regression model was proposed to generate distributions of walking-in and walking-out times in access and egress stations (gamma distributions) and distributions of waiting times at access and transfer stations (uniform distributions) including the additional waiting time caused by fail-to-board. Based on Poisson arrival assumption, a mixed-weight inference of passenger path choice behavior, gate-to-gate journey time distribution, and crowding metrics on platform were derived from AFC data between two O-D pairs in Beijing. Based on this study, Sun and Schonfeld [28] further developed a schedule-based assignment model by extending the fail-to-board (or left behind) model of [29], and then the transfer penalty. Since simplified assumptions were taken, that constrained the model application.

A key element in the transit model is the probability of matching between a given train run serving the boarding station and a given passenger journey. To generalize the previous works in [27, 28], Zhu et al. [30–33] devised a probabilistic individual Passenger-to-Train Assignment Model (PTAM), in which the matching probability between train runs and individual journeys was emphasized, for rail transit service estimation. A regression model of passenger walking speeds was considered. The PTAM was developed so as to provide estimators of the main state variables of passenger flows both in-station (waiting loads) and in-vehicle (loading snakes) times as well as in passenger in access and egress stations. Within this wide scope, the issue of passenger positioning was mentioned: Zhu [30] suggested modeling it as a Random Variable (RV), yet apparently with no further development. The model was applied on a five-station metro line segment using AFC data and AVL data of HongKong and validated by synthetic data. The authors proposed estimation for several factors, the distributions of passenger walking-in and walking-out speeds by trip leg, the distribution of fail-to-board (left behind) on boarding platform (the number of trains/times a passenger unable to board due to capacity constraint), and in-station crowding level. Nevertheless, this study considered explicitly only distributions of walking-in and walking-out speeds (as normal distributions) and a two-itinerary choice scenario without transfer. To integrate transfer in crowding discomfort analysis, Hörcher et al. [34] combined the probabilistic assignment model of [31–33] with the standing and sitting choice models [35, 36] and applied to a PT network of HongKong. A group of transfers and their choices are modeled in a Bayesian framework at the journey level to model complex trips that could involve transfers hence combine legs (journeys), in order to study route choice between station pairs and reveal the influence of in-vehicle congestion on route disutility to passengers.

In the meanwhile, in more recent studies, econometric methods were also applied for modeling and estimating individual passenger in-station underlying statistical distributions. Inspired from the econometric approach using highway individual vehicle toll data in [37, 38], Leurent and Xie [21] modeled explicitly constant or uniform-distributed passenger walking speed with shifted exponential-distributed distances by trip leg. Individual passenger movement was integrated to train run. Parameters were estimated by the Maximum Likelihood Estimation (MLE) by using AFC data and timetable of the line RER A in Paris area. These two models in [21] were further developed by adding a shifted exponential-distributed waiting time in [22], which made computation tractable and yielded consistent estimates with formers. Xie and Leurent [23] demonstrated that normal-distributed walking speed reduced likelihood function computation complexity. Our stochastic models and related MLE were much analogous to those in [30–33]. Their independent development led to respective assets: the authors had a wide scope of traffic variables and considers fail-to-board an arriving train, which we did not take into account. Nevertheless, on our side, there was a linkage between the tap-in and tap-out (TITO) times based on the individual walking speed,

in addition to the explicit consideration of individual positioning. Moreover, the consideration of passenger individual distances in the model resolved the conservative assumption that the minimum access and egress time were zero (infinite passenger speed) in [27, 28, 31–33]. Simultaneously, [39] provided a stochastic frontier model by extending the model in [40] for estimating Erlang-distributed waiting time and service reliability. Walking times and in-vehicle travel time were normal distributions. Parameters were also estimated by the maximum likelihood method using AFC data and timetable of the Piccadilly line of London's underground network.

Since the study on passenger longitudinal repositioning distance along boarding platform during train waiting was scarce, passenger longitudinal repositioning distance was primarily modeled in our most recent work [24]. By extending this model, a novel approach about passenger individual repositioning distance with posterior analysis is proposed in this paper. To simplify computation complexity and save model optimization time, the passenger longitudinal repositioning distance is modeled explicitly, but implicit for waiting times. It brings about an explicit model that is both physical and probabilistic: longitudinal repositioning distance is modeled in relation to walking distances on the boarding and alighting platforms and to the individual passenger walking speed and the time available to the passenger for walking along the platform up to the train arrival hence prior to boarding it. The influence of longitudinal position is traced out on the passenger journey time between selected points in the access and egress stations, such as tap gates. Analytical formulae are obtained for related Probability Density Functions (PDFs) of tap-out times conditioning on tap-in time and walking speed, or on tap-in time only. They are analytical closed-form PDFs that provide likelihood functions, which are tractable by MLE under a simple specification of the basic statistical distributions. To accomplish that, computation and optimization are detailed. Posterior analysis imitated Bayesian analysis is applied. Train occurrence and passenger trips are reanalyzed.

This paper is organized as follows. Section 2 provides the physical and stochastic model in a bottom up way from elementary bricks by journey phase to individual journey time. The analytical formulae of the probability to take a given train, the distribution of repositioning distances and the distributions of egress times are built. In Section 3, particular statistical distributions are selected as inputs in order to obtain closed-form formulae for the PDFs of interest, thus easing the computation of the model outcomes. Section 4 puts forward the statistical estimation scheme. MLE with optimization is involved on the basis of AFC and AVL datasets in order to estimate the parameters of the statistical distributions. The resulting estimates can then be used for ex-post analysis of individual journeys so as to infer their most likely features. A real case study on the commuter line RER A in Paris area, France, is provided in Section 5. Estimation results are provided for boarding positions, walking speeds and in-station walking distances, and applied to posterior analysis for inferring passenger individual journey details. Lastly, Section 6 assesses the outreach and limitations of our approach and points to directions for further research.

2. Physical and Stochastic Models of Passenger Repositioning along the Platform

Stochastic models on passenger repositioning between a pair of O-D stations (called also a trip leg) are depicted in this section. This study is built upon the time-space diagram of traffic flow theory, the kinematic theory, and basic econometric theory to understand passenger individual in-station movement related to train run choice by trip leg along an urban rail transit line.

We first state the physical model of one individual passenger making a train journey by availing oneself of train runs and waiting time prior to boarding (Section 2.1). The in-station walking distances are salient features, which are modeled as RVs (Section 2.2). This makes the passenger access to train runs a stochastic process. We derive the probability distributions of passenger-to-train assignment, of the walking (positioning) distance along boarding platform, and of the residual waiting time up to boarding. The egress phase is also a stochastic process, in which the distance that remains to walk along the alighting platform is a RV, the total walking-out time from train alighting to tap-out gate and the tap-out instant from the egress station as well. Lastly, we model the walking speed as a RV and derive the consequences on access and egress processes (Section 2.3). The main notations used in this study are introduced in Table 1.

2.1. Physical Model of a Passenger Journey. Consider here a noncyclic urban rail transit line with J stations. There are several itineraries between station platform and tap gates with different kinds of pedestrian facilities (flat road, stair, escalator, lift, etc.).

Let us consider a transit passenger (user), denoted by index u, with individual walking speed v that is a "cruising speed." The passenger makes a simple journey by train along a line from access station a to egress station e. The instants of passage at selected points in each station at TITO gates are denoted as h_u^+ and h_u^- (Figure 1), respectively, and called TITO instants.

In access station, the walking-in distance from tap-in gate O to boarding platform entrance A is denoted as s^+, whereas the walking distance is denoted as s on boarding platform from A to boarding (also waiting) point M. In egress station, the walking-out distance from alighting platform exit E to tap-out gate D is denoted as s^-, whereas the walking distance is denoted as s' on alighting platform from alighting points N to E. Passenger walking paths in stations are called "walking links," green arrows in Figure 1. The total walking-in and total walking-out times in access and egress stations are $t^a = t^+ + s/v$ and $t^e = t^- + s/v$, respectively, where $t^+ = s^+/v$ and $t^- = s^-/v$. Passenger's journey time on this trip leg is equal to $h_u^- - h_u^+$.

Let us now consider the longitudinal dimension of platforms. In urban rail transit system, station platforms are long objects of relatively modest width (e.g., several meters),

Table 1: Notations.

Notations	Descriptions
	Passenger
u and U	Passenger number and passenger set by station leg
v	Passengers' walking speeds on walking links in access and egress stations
L	Passenger longitudinal repositioning distance
h_u^+ and h_u^-	Passenger u's tap-in instant at a access gate and tap-out instant at a egress gate on a trip leg
s^+ and t^+	Passenger walking-in distance and time from tap-in gate to platform access point in access station
s	Passenger walking-in distance from platform entrance to platform boarding point in access station
s'	Passenger walking-out distance from platform alighting point to platform exit in egress station
s^- and t^-	Passenger walking-out distance and time from platform egress point to tap-out gate in egress station
t^a and t^e	Total walking-in and total walking-out times in access and egress stations
w_u^i	Passenger apparent waiting time
$\widehat{w}_u^i$	Passenger residual waiting time on waiting (also boarding) point, $\widehat{w}_u^i = w_u^i - s/v$
$\Delta h_{u,i}^+$	Passenger total time cost in access station for boarding train run i, $\Delta h_{u,i}^+ = t_a + \widehat{w}_u^i$
	Train
i and I	Train run number and train run set along a line
n_u^0 and I_u^0	Number and subset of train runs for passenger u
n_u and I_u	Feasible number and subset of train runs for passenger u
h_i^+ and h_i^-	Train run i's departure time at access station and arrival time at egress station
v_i	Train run speeds between access and egress stations
	Network
j and J	Station number and station set along a line
a and e	Access station and egress station
	Auxiliary variables or functions
X	Cumulative Distribution Function (CDF) of variable x
$\dot{\mathrm{X}}$	Probability Density Function (PDF) of variable x
$p_i^{h,v}$	Probability to take train run i, conditioning on h and v
$\tilde{s}_i^{h,v}$	Threshold value of repositioning distance
$\psi_{i,h}(x, v)$	Kernel function of repositioning on platform
t_p	Studied period, $t_p = [t_1, t_2]$
$\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$	Parameter vector and its estimated value
$\boldsymbol{\Theta}$	Space of parameter vector $\boldsymbol{\theta}$

whereas their lengths range from some dozen meters in tramway stations to a couple of hundred meters in stations of metro, urban or suburban train. In the latter case, passengers are expected to walk up to a given position for waiting and boarding, boarding point M with abscissa s_M along the platform from a "starting point" A (platform entrance) with abscissa s_A. The walking-in distance s on boarding platform amounts to $|s_M - s_A|$. If a passenger does not make a significant longitudinal move on the train, passenger's boarding point (door) M and alighting point (door) N is the same door of the train. Along egress platform, the walking-out distance s' amounts to $|s_E - s_N|$ by using the same scale to measure abscissas along platforms of a given line. Thus, the total distance walked by the passenger on both the access and the egress station platforms amounts to $s+s' = |s_M - s_A| + |s_E - s_N|$. Given boarding platform entrance A and alighting platform exit E, this is a function of the boarding abscissa s_M only, made up of two parts.

Let us denote by $L \equiv |s_E - s_A|$ the maximum distance for passenger longitudinal repositioning, called longitudinal repositioning distance. Postulating that the passenger is well-aware of positions, rational and unimpeded by crowding, then it holds that $L = s + s'$ and the function does not vary with s_M. However, s_M influences the walking times prior to and posterior to running aboard train: passenger u can decide to invest time s/v in order to save on $s'/v = (L - s)/v$ just after alighting. For example, assuming that the platform entrance is at one end of the station platform (that of train head or tail) and the platform exit is at the other end of the train (that of train tail or head).

The time available to u for positioning oneself may be limited by the instant of train departure. Indexing by i the train run of interest to the passenger, let us denote as h_i^+ its departure time from access station and h_i^- its arrival time at egress station. Train travel time from access station to egress station along the train trajectory (red arrows in Figure 1)

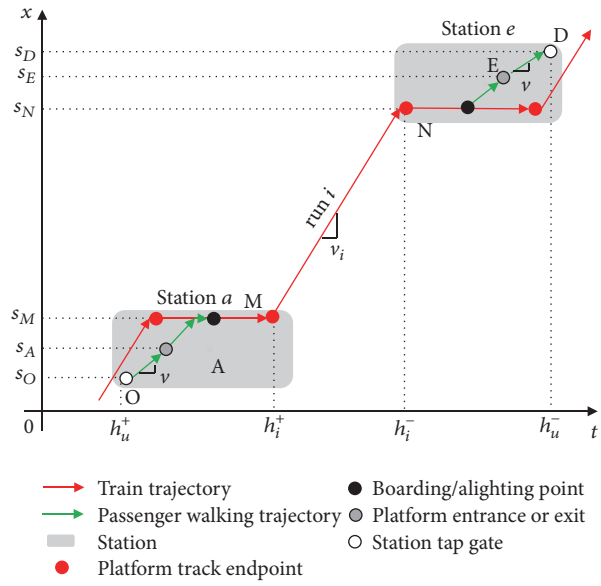


Figure 1: Time-space diagram: train and passenger trajectories.

is equal to $h_i^- - h_i^+$. Train travel time is considered as in-vehicle travel time for all passengers who take the same train. The frequency of each service is heterogeneous: Peak Hour (PH) is different from Off-Peak Hour (OPH). Trains along passengers' travel direction consist in a set I for all train runs i during the studied period $t_p = t_2 - t_1$. The traffic conditions and line service quality (punctuality, regularity, etc.) depend on passengers' conditions in all served stations.

Then, on boarding platform, the apparent waiting time of passenger u amounts to $w_u^i = (h_i^+ - h_u^+) - t^+$. By selecting a particular location along the platform, the passenger succeeds to turn part of that apparent waiting time w_u^i into "useful" time for the rest of his trip. His "distance investment" s is limited by both L and the available time before the train departure. Letting $\Delta h_{u,i}^+ \equiv h_i^+ - h_u^+ > 0$ passenger total time cost in access station for boarding train run i, the time available to the passenger for positioning himself along the platform is limited by $t_{u,i} \equiv \Delta h_{u,i}^+ - t^+$. So, the maximum distance for repositioning on boarding platform amounts to $s_{u,i} \equiv v t_{u,i} = v\Delta h_{u,i}^+ - s^+$.

Assume that the passenger is a rational decision-maker willing to minimize his or her exit time and walks along boarding platform as much as possible, yielding $s = \min\{L, v\Delta h_{u,i}^+ - s^+\}$. This optimizing behavior induces an adjustment of the passenger to the temporal prism of opportunities that are opened by the respective times h_i^+ and $h_u^+ + s^+/v$. The waiting time until train departure on boarding point M is called residual waiting time $\widehat{w}_u^i$ and amounts to $(h_i^+ - h_u^+) - (s^+ + s)/v = w_u^i - s/v$.

Thus, the journey time $h_u^- - h_u^+$ is derived by $(s^+ + s)/v + \widehat{w}_u^i + (h_i^+ - h_u^+) + (s' + s^-)/v$. The time-space diagram (Figure 1) depicts both passenger trajectory and train trajectory between the access and egress stations.

2.2. Models Conditioning on Distributed Walking Distances. We model firstly passenger journey with distributed walking distances and constant walking speed; the stochastic model is integrated with respect to walking (positioning) distance s on boarding platform.

2.2.1. In-Station Walking Distances Probabilistic Models. As a general notation, for a Random Variable (RV) x, its Cumulative Distribution Function (CDF) is denoted as X, and its Probability Density Function (PDF) as $\dot{X}$. Recall that $X(x) = \int_x \dot{X}(z)dz$, where $\dot{X}$ is regular enough.

Urban rail transit stations vary from simple stations at grade providing access to one line only, to complex transit hubs connecting several lines and equipped with several platforms on several floors. Most of them are underground. Whatever the case, the distances between station tap gates and platform entrances or exits extend to some dozen meters at least and up to some hundred meters. Since the walking-in distance s^+ is variable, it is modeled as a RV with CDF S^+ and PDF $\dot{S}^+$. The same hypothesis is applied to walking-out distance s^-, as a RV with CDF S^- and PDF $\dot{S}^-$.

2.2.2. Access Model. A passenger u is characterized by the (h, v) pair, where $h = h_u^+$, that is taken as exogenous.

Train run i is taken if and only if $(h + s^+/v) \in]h_{i-1}^+, h_i^+]$. Then, the probability to take train run i is as follows:

$$p_i^{h,v} \equiv \Pr\left\{\left(h + \frac{s^+}{v}\right) \in]h_{i-1}^+, h_i^+]\right\} = S^+\left(v\Delta h_{u,i}^+\right) - S^+\left(v\Delta h_{u,i-1}^+\right). \quad (1)$$

The positioning distance on boarding platform depends on the train run that is taken and the walking-in distance

$$s_i^{h,v} = \min\left\{L, v\Delta h_{u,i}^+ - s^+\right\}. \quad (2)$$

Two cases must be distinguished

(i) either $s_i^{h,v} = L$, that holds iff $s^+ \le v\Delta h_{u,i}^+ - L$;

(ii) or $s_i^{h,v} < L$, that holds iff $s^+ > v\Delta h_{u,i}^+ - L$.

The former case of total positioning can happen only if $v\Delta h_{u,i}^+ - L \ge v\Delta h_{u,i-1}^+$. Denoting $\hat{s}_i^{h,v} = \max\{v\Delta h_{u,i-1}^+, v\Delta h_{u,i}^+ - L\}$, total positioning happens for $s^+ \in]v\Delta h_{u,i-1}^+, \hat{s}_i^{h,v}]$, with probability

$$\Pr\{s = L \mid h, v, i\} = \frac{S^+\left(\hat{s}_i^{h,v}\right) - S^+\left(v\Delta h_{u,i-1}^+\right)}{p_i^{h,v}}. \quad (3)$$

In the alternative case of partial positioning $s = x < L$, the associated PDF is

$$\Pr\{s = x \mid h, v, i\} = \frac{\dot{S}^+\left(v\Delta h_{u,i}^+ - x\right)}{p_i^{h,v}}. \quad (4)$$

Bringing together the two cases, a CDF $S_i^{h,v}$ for the positioning distance s is built up, as $S_i^{h,v}(s) = \int_0^{\min\{s, L, v(h_i^+ - h_{i-1}^+)\}} (\dot{S}^+(v\Delta h_{u,i}^+ - x)/p_i^{h,v})dx + 1_{\{s \ge L\}}((S^+(\hat{s}_i^{h,v}) - S^+(v\Delta h_{u,i-1}^+))/p_i^{h,v})$. This CDF satisfies the fact that

$$S_i^{h,v}(s) = \frac{\left[S^+\left(v\Delta h_{u,i}^+\right) - S^+\left(v\Delta h_{u,i}^+ - \min\left\{s, L, v\left(h_i^+ - h_{i-1}^+\right)\right\}\right) + 1_{\{s\geq L\}}\left(S^+\left(\hat{s}_i^{h,v}\right) - S^+\left(v\Delta h_{u,i-1}^+\right)\right)\right]}{p_i^{h,v}}. \quad (5)$$

Thus the positioning distance s is an RV conditioning on v and h on the passenger side and on i on the train side It is endowed with a closed-form CDF that involves the tap-in instant h_u^+, the train departure time h_i^+, the CDF S^+ of walking-out distance from tap-in gate to the boarding platform entrance, and the target distance of repositioning L.

Of course, the RV does only exist when the matching probability is strictly positive, i.e., iff $\Delta h_{u,i}^+ > 0$. We denote as the subset I_u of train runs i that are feasible for u, i.e., such that $\Delta h_{u,i}^+ > 0$ and $h_u^+ < h_i^+$.

2.2.3. Egress Model. The positioning distance s that stems from the access model makes an input to the egress model. The distance to walk in the egress station includes $s' = L - s$ on the alighting platform plus s^-. The conditional egress time $t_{i,h,v,s}^e = (L - s + s^-)/v$ satisfies the fact that

$$\Pr\left\{t_{i,h,v,s}^e \leq x\right\} = \Pr\left\{L - s + s^- \leq vx\right\}. \quad (6)$$

So $t_{i,h,v,s}^e$ has CDF $T_{i,h,v,s}^e(x) = S^-(vx+s-L)$. It is integrated with respect to s; then

$$\begin{aligned} T_{i,h,v}^e(x) &= \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} T_{i,h,v,s}^e(x)\, dS_i^{h,v}(s) \\ &= \frac{\psi_{i,h}(x,v)}{p_i^{h,v}}. \end{aligned} \quad (7)$$

The resulting value $T_{i,h,v}^e(x)$ is a probability conditioning on i, h, v. Let us define $\psi_{i,h}(x,v) \equiv \Pr\{t_{v,h}^e \leq x \bigcap (u \text{ takes } i)\}$. From the definition of conditional probability, we have that

$$\psi_{i,h}(x,v) \equiv \Pr\left\{t_{v,h}^e \mid (u \text{ takes } i)\right\} p_i^{h,v} = p_i^{h,v} T_{i,h,v}^e(x). \quad (8)$$

On replacing $S_i^{h,v}(s)$ with its expression in (5), we get

$$\begin{aligned} \psi_{i,h}(x,v) \equiv{}& \left(S^+\left(\hat{s}_i^{h,v}\right) - S^+\left(v\Delta h_{u,i-1}^+\right)\right) S^-(vx) \\ &+ \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} S^-(vx - L + s)\, \dot{S}^+\left(v\Delta h_{u,i}^+ - s\right) ds. \end{aligned} \quad (9)$$

The first term is zero if $\hat{s}_i^{h,v} = vh_{u,i-1}^+$; i.e., $L \geq v(h_i^+ - h_{i-1}^+)$.

Thus the total walking-out time t^e as an RV $t_{i,v,h}^e$ conditioning on v, h, and admissible run $i \in I_u$ is endowed with a closed-form CDF that involves the tap-in instant h_u^+, the train departure time h_i^+, the target distance of repositioning L, and both distributions of in-station walk distances S^+ and S^- in access and egress stations, respectively. The product between S^- and $\dot{S}^+$ in the formula exhibits the influence of v on both the egress and access times. In other words, the individual walk speed links together the egress and access times.

To integrate with respect to i, the CDF $T_{h,v}^e(x) = \sum_{i\in I} p_i^{h,v} T_{i,h,v}^e(x)$ is considered. From (5), it satisfies that

$$T_{h,v}^e(x) = \sum_{i\in I} \psi_{i,h}(x,v). \quad (10)$$

Concerning the tap-out instant h_u^-, conditioning on i, we have $h_u^- = t^e + h_i^-$ and $H_{i,h,v}^-(x) = T_{i,h,v}^e(x - h_i^-)$.

By integrating with respect to i, we get the unconditional CDF $H_{i,h,v}^-(x) = \sum_{i\in I} p_i^{h,v} T_{i,h,v}^e(x - h_i^-)$, which satisfies the fact that

$$H_{h,v}^-(x) = \sum_{i\in I} \psi_{i,h}\left(x - h_i^-, v\right). \quad (11)$$

Thus the tap-out instant h_u^- is an RV conditioning on v and h, which is endowed with a closed-form CDF that involves not only the tap-in instant h_u^+, the train departure time h_i^+, the target distance of repositioning L, and the distributions of in-station walk distances S^+ and S^- in access and egress stations but also the instants of train arrival at the egress station h_i^-.

From the CDF of the total walking-out time and tap-out instant, either conditioning on i or not, it is easy to derive the associated PDF by straightforward differentiation with respect to argument x. That is,

$$\begin{aligned} \dot{\psi}_{i,h}(x,v) \equiv \frac{\partial}{\partial x}\psi_{i,h}(x,v) = {}& \left(S^+\left(\hat{s}_i^{h,v}\right)\right. \\ &\left. - S^+\left(v\Delta h_{u,i-1}^+\right)\right) v\dot{S}^-(vx) \\ &+ v\int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} \dot{S}^-(vx - L + s) \\ &\cdot \dot{S}^+\left(v\Delta h_{u,i}^+ - s\right) ds. \end{aligned} \quad (12)$$

The associated PDFs are

$$\dot{T}_{i,h,v}^e(x) = p_i^{h,v}\dot{\psi}_{i,h}(x,v); \quad (13a)$$

$$\dot{T}_{h,v}^e(x) = \sum_{i\in I}\dot{\psi}_{i,h}(x,v); \quad (13b)$$

$$\dot{H}_{i,h,v}^-(x) = p_i^{h,v}\dot{\psi}_{i,h}\left(x - h_i^-, v\right); \quad (13c)$$

$$\dot{H}_{h,v}^-(x) = \sum_{i\in I}\dot{\psi}_{i,h}\left(x - h_i^-, v\right). \quad (13d)$$

It should be noted that the above stochastic model assumes that a passenger has constant velocity motion in access and egress stations. From a practical point of view, this assumption may not be realistic and thus too restrictive. Extensions will be considered in next subsection to consider a distributed walking speed.

2.3. Models Conditioning on Distributed Speed and Walking Distances. When passenger walking speed v is a distributed variable, the stochastic model is integrated with respect to walking speed as well.

2.3.1. In-Station Walking Speed Probabilistic Model. The walking speed v of one passenger u is a notional speed, averaged for that passenger on that journey over a range of travel situations. It may be called an individual "cruising speed" during walking. Although personal walking habits may be consistent, as a fact, the cruising walking speed fluctuates from one occasion to another, for instance, between journeys reiterated on a given access and egress pair from day to day.

In addition to this intraindividual diversity, there is an even larger diversity between individual passengers, since people differ in their respective walking abilities. Young adults can walk faster than elderly people and are likely to be more hurried. People with luggage or young child either walking or in a stroller walks more slowly than the average adult. A rough indication about the statistical distribution of walk speeds for a typical population of transit users in the urban setting was close to a normal distribution with mean of about 0.90 m/s and standard deviation of about 0.20 m/s or to a uniform distribution ranging from 0.58 to 1.24 m/s in the Appendix, the cases with waiting time integrated in walking time.

Assuming that passengers' walking speeds on walking links in access and egress stations obey the same statistical distribution, it is easy to extend the stochastic model of passenger repositioning to a diversity of walking speeds, by considering walking speed in a given population U of passengers as a RV, with CDF V and PDF $\dot{\mathrm{V}}$.

We still denote S^+ and S^- the distribution functions of in-station walking distances. In fact, they can be expected to have different average values and wider spreads as compared to their individual counterparts. Such differences between intra-individual and inter-individual distributions were classical in the stochastic modeling of socioeconomic behaviors of highway individuals in Chapter 6 of [38].

2.3.2. Access Model. To integrate with respect to v, the probability to take train run i is still conditioning on the tap-in time $h = h_u^+$ and is

$$\begin{aligned} p_i^h &\equiv \int p_i^{h,v} \mathrm{dV}(v) \\ &= \int \left(\mathrm{S}^+\left(v\Delta h_{u,i}^+\right) - \mathrm{S}^+\left(v\Delta h_{u,i-1}^+\right)\right) \mathrm{dV}(v). \end{aligned} \tag{14}$$

Then, conditioning on i and h, the distribution of walking speeds has PDF as follows:

$$\dot{\mathrm{V}}_i^h(v) = \frac{\dot{\mathrm{V}}(v)\, p_i^{h,v}}{p_i^h}. \tag{15}$$

2.3.3. Egress Model. Conditioning on i and h, the egress time $t_{i,h}^e$ has CDF $\mathrm{T}_{i,h}^e(x) = \int \mathrm{T}_{i,h,v}^e(x)\dot{\mathrm{V}}_i^h(v)\mathrm{d}v = \int(\psi_{i,h}(x,v)/p_i^{h,v})\dot{\mathrm{V}}_i^h(v)\mathrm{d}v$, so

$$\mathrm{T}_{i,h}^e(x) = \frac{1}{p_i^h}\int \psi_{i,h}(x,v)\,\mathrm{dV}(v). \tag{16}$$

Denoting $\widetilde{\psi}_{i,h}(x) \equiv \int \psi_{i,h}(x,v)\mathrm{dV}(v)$, it is then easy to integrate with respect to i. As $\{t_h^e \le x\} = \bigcup_{i\in I}\{t_{i,h}^e \le x \bigcap i\}$, there is $\Pr\{t_h^e \le x\} = \sum_{i\in I}\Pr\{t_{i,h}^e \le x\}\Pr\{i\}$ from the composition of conditional probabilities. Thus,

$$\mathrm{T}_h^e(x) = \sum_{i\in I}\mathrm{T}_{i,h}^e(x)\, p_i^h = \sum_{i\in I}\widetilde{\psi}_{i,h}(x). \tag{17}$$

The exit instant h_u^- has CDF as follows:

$$\mathrm{H}_h^-(x) = \sum_{i\in I}\widetilde{\psi}_{i,h}\left(x - h_i^-\right). \tag{18}$$

Thus, both the total walking-out time and the tap-out instant are endowed with closed-form CDF that involve the tap-in instant h and the statistical distribution of s^+, s^-, and v. This enables us to derive their PDF by differentiation with respect to their argument x. Denoting $\dot{\widetilde{\psi}}_{i,h}(x) \equiv (\partial/\partial x)\widetilde{\psi}_{i,h}(x)$, it holds that

$$\dot{\mathrm{T}}_h^e(x) = \sum_{i\in I}\dot{\widetilde{\psi}}_{i,h}(x); \tag{19a}$$

$$\dot{\mathrm{H}}_h^-(x) = \sum_{i\in I}\dot{\widetilde{\psi}}_{i,h}\left(x - h_i^-\right). \tag{19b}$$

3. Distribution Specification for Tractable Computation

The analytical formulae obtained so far involve the integration of specific functions along one or two scalar dimensions, namely one dimension of space (with respect to distance s) and eventually another dimension of speed (with respect to walking speed v). The outcomes can be obtained by numerical integration along the said axes. This can be circumvented by availing ourselves of ad hoc distributions that yield straightforward formulae for the respective PDFs of total walking-out time and tap-out instant, though it is not so simple.

In this section, we firstly put forward specific distributions that are suitable to our purpose (Section 3.1). Then, we provide a lemma (Section 3.2) for the core computation to deal with distributed speed by providing formulae for $x \mapsto \dot{\psi}_{i,h}(x,v)$ and $x \mapsto \dot{\widetilde{\psi}}_{i,h}(x)$ functions (Section 3.3). Finally, the variations of those two functions are illustrated (Section 3.4).

3.1. Ad Hoc Specification of the Distributions. The obtained general PDF functions constitute the likelihood functions of all assumed parameters. To prepare for further work on the MLE of those parameters, some hints of PDF computation under ad hoc selection of distributions are provided.

For a given pair of access and egress stations, we take v as either a variable or a normal-distributed walking speed, together with shifted exponential-distributed walking distances and the variable L. The distribution functions of v, s^+, and s^- are specified as follows:

(i) Individual speed follows a normal (Gaussian) distribution with mean μ and variance σ^2, since a normal distribution reduces likelihood function complexity [23].

(ii) Walking-in distance s^+ obeys a shifted exponential distribution with main parameter λ^+ and shift B^+; thus CDF $\mathrm{S}^+(x) = 1_{\{x\geq B^+\}}[1-\exp(-\lambda^+(x-B^+))]$ and PDF $\dot{\mathrm{S}}^+(x) = 1_{\{x\geq B^+\}}\lambda^+\exp(-\lambda^+(x-B^+))$.

(iii) Walking-out distance s^- is also a shifted exponential distribution, with main parameter λ^- and shift B^-; thus CDF $\mathrm{S}^-(x) = 1_{\{x\geq B^-\}}[1-\exp(-\lambda^-(x-B^-))]$ and PDF $\dot{\mathrm{S}}^-(x) = 1_{\{x\geq B^-\}}\lambda^-\exp(-\lambda^-(x-B^-))$.

3.2. Lemma. Let us establish a property for a normal RV $N(\mu,\sigma^2)$ combined with a consumption function that is the product of Kv and an exponential function e^{-lv+k}. The aim is to obtain a straightforward formula for integral function $F(x) \equiv \int_{-\infty}^{x} Kve^{-lv+k}\mathrm{d}\Phi((v-\mu)/\sigma)$.

As $e^{-lv}e^{-(1/2)((v-\mu)/\sigma)^2} = e^{-(1/2)((v-\mu')/\sigma)^2}e^{l\mu-(1/2)l^2\sigma^2}$ for $\mu' = \mu - l\sigma^2$, we can put aside the constant factor $Ke^{k+l\mu-(1/2)l^2\sigma^2}$ and focus on

$$\begin{aligned}\int_{-\infty}^{x} v\mathrm{d}\Phi\left(\frac{v-\mu'}{\sigma}\right) &= \int_{-\infty}^{x}\frac{v-\mu'}{\sqrt{2\pi}\sigma}e^{-(1/2)((v-\mu')/\sigma)^2}\mathrm{d}x \\ &\quad + \int_{-\infty}^{x}\mu'\mathrm{d}\Phi\left(\frac{v-\mu'}{\sigma}\right) \\ &= \frac{\sigma}{\sqrt{2\pi}}\left[-e^{-(1/2)((v-\mu')/\sigma)^2}\right]_{-\infty}^{x} \\ &\quad + \mu'\Phi\left(\frac{x-\mu'}{\sigma}\right) \\ &= \mu'\Phi\left(\frac{x-\mu'}{\sigma}\right) \\ &\quad - \frac{\sigma}{\sqrt{2\pi}}e^{-(1/2)((x-\mu')/\sigma)^2}.\end{aligned} \tag{20}$$

So the final result is $\int_{-\infty}^{x} Kve^{-lv+k}\mathrm{d}\Phi((v-\mu)/\sigma) = Ke^{k+l\mu-(1/2)l^2\sigma^2}[\mu'\Phi((x-\mu')/\sigma) - (\sigma/\sqrt{2\pi})e^{-(1/2)((x-\mu')/\sigma)^2}]$.

3.3. Core Computation. The PDF formulae (13d), (13b), and (13c) for a constant speed and (19a) and (19b) for normal-distributed speed involve basic bricks of the form $x \mapsto \dot{\psi}_{i,h}(x,v)$ and $x \mapsto \dot{\tilde{\psi}}_{i,h}(x)$, respectively. These are obtained by straightforward differentiation of function $\psi_{i,h}(x,v)$ with respect to x.

As $\psi_{i,h}(x,v) \equiv (\mathrm{S}^+(\tilde{s}_i^{h,v}) - \mathrm{S}^+(v\Delta h_{u,i-1}^+))\mathrm{S}^-(vx) + \int_0^{\min\{L,v(h_i^+-h_{i-1}^+)\}}\mathrm{S}^-(vx-L+s)\dot{\mathrm{S}}^+(v\Delta h_{u,i}^+-s)\mathrm{d}s$, then $\dot{\psi}_{i,h}(x,v) \equiv (\mathrm{S}^+(\tilde{s}_i^{h,v}) - \mathrm{S}^+(v\Delta h_{u,i-1}^+))v\dot{\mathrm{S}}^-(vx) + v\int_0^{\min\{L,v(h_i^+-h_{i-1}^+)\}}\dot{\mathrm{S}}^-(vx-L+s)\dot{\mathrm{S}}^+(v\Delta h_{u,i}^+-s)\mathrm{d}s$. This function is in two parts, left and right, respectively.

The left part is computed firstly and produces the two terms: $(\mathrm{S}^+(\tilde{s}_i^{h,v}) - \mathrm{S}^+(v\Delta h_{u,i-1}^+)) = e^{-\lambda^+(v\Delta h_{u,i-1}^+-B^+)^+}1_{\{v\Delta h_{u,i-1}^+\geq B^+\}} - e^{-\lambda^+(\max\{v\Delta h_{u,i-1}^+,v\Delta h_{u,i}^+-L\}-B^+)^+}1_{\{\max\geq B^+\}}$ and $v\dot{\mathrm{S}}^-(vx) = v\lambda^-e^{-\lambda^-(vx-B^-)}1_{\{vx\geq B^-\}}$.

So the product of the two terms is ready to compute the left part in $\dot{\psi}_{i,h}(x,v)$.

To obtain $x \mapsto \dot{\tilde{\psi}}_{i,h}(x)$, it is necessary to integrate the product over v. The product is nonzero only if $vx \geq B^-$; i.e., $v \geq B^-/x$. Then, it gives rise to two terms with respective formula as follows:

(L-a) $v\lambda^-e^{-\lambda^-(vx-B^-)-\lambda^+(v\Delta h_{u,i-1}^+-B^+)^+}1_{\{v\Delta h_{u,i-1}^+\geq B^+\}}$ and

(L-b) $v\lambda^-e^{-\lambda^-(vx-B^-)-\lambda^+(\max\{v\Delta h_{u,i-1}^+,\Delta h_{u,i}^+-L\}-B^+)^+}1_{\{\max\{v\Delta h_{u,i-1}^+,\Delta h_{u,i}^+-L\}\geq B^+\}}$.

The first one (L-a) must be integrated for $v \geq B^+/\Delta h_{u,i-1}^+$ only. Denoting $= \max\{B^-/x, B^-/\Delta h_{u,i-1}^+\}$, there is $\int_z^{+\infty} v\lambda^-e^{-\lambda^-(vx-B^-)-\lambda^+(v\Delta h_{u,i-1}^+-B^+)}\mathrm{dV}(v) = \lambda^-e^{\lambda^+B^++\lambda^-B^-}\int_z^{+\infty} ve^{-(\lambda^-x+\lambda^+\Delta h_{u,i-1}^+)v}\mathrm{dV}(v)$.

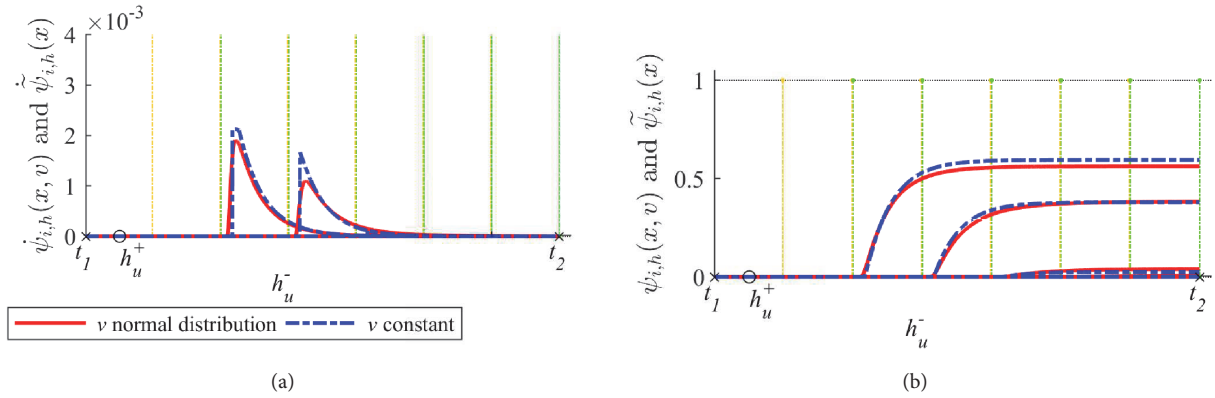
The second term (L-b) is dealt with similarly, yet with distinction between two subdomains depending on which function is greater between $v \mapsto v\Delta h_{u,i-1}^+$ and $v \mapsto v\Delta h_{u,i}^+ - L$.

If $v\Delta h_{u,i-1}^+ \leq v\Delta h_{u,i}^+ - L$, i.e., $L/(h_i^+ - h_{i-1}^+) \leq v$, then, denoting $y = \max\{B^-/x, L/(h_i^+ - h_{i-1}^+)\}$, the second term reduces to $\int_y^{+\infty} v\lambda^-e^{-\lambda^-(vx-B^-)-\lambda^+(v\Delta h_{u,i-1}^+-L-B^+)}\mathrm{dV}(v) = \lambda^-e^{\lambda^+(L+B^+)+\lambda^-B^-}\int_y^{+\infty} ve^{-(\lambda^-x+\lambda^+\Delta h_{u,i-1}^+)v}\mathrm{dV}(v)$.

In the other case where $v < L/(h_i^+ - h_{i-1}^+)$, the subdomain of integration is empty if $L/(h_i^+ - h_{i-1}^+) < B^-/x$ or $[B^-/x, L/(h_i^+ - h_{i-1}^+)]$ otherwise. The formula is the same as for (L-a).

Concerning the right part, there is

$$\begin{aligned}v\int_0^{\min\{L,v(h_i^+-h_{i-1}^+)\}}\dot{\mathrm{S}}^-(vx-L+s)\dot{\mathrm{S}}^+\left(v\Delta h_{u,i}^+-s\right)\mathrm{d}s &= v\lambda^-\lambda^+\int_0^{\min\{L,v(h_i^+-h_{i-1}^+)\}}e^{-\lambda^-(vx-L+s-B^-)-\lambda^+(v\Delta h_{u,i-1}^+-s-B^+)}1_{\{vx-L+s\geq B^-\}} \\ 1_{\{v\Delta h_{u,i-1}^+-s\geq B^+\}}\mathrm{d}s &= \sigma v\lambda^-\lambda^+\int_{B^-+L-vx}^{\min\{L,v(h_i^+-h_{i-1}^+),v\Delta h_{u,i-1}^+-B^+\}}e^{-\lambda^-(vx-L+s-B^-)-\lambda^+(v\Delta h_{u,i-1}^+-s-B^+)}\mathrm{d}s \\ &= \sigma\frac{\lambda^-\lambda^+}{\lambda^+-\lambda^-}ve^{\lambda^+(B^+-v\Delta h_{u,i-1}^+)+\lambda^-(L+B^+)}\left[e^{(\lambda^+-\lambda^-)s}\right]_{(B^-+L-vx)^+}^{\min\{L,v(h_i^+-h_{i-1}^+),v\Delta h_{u,i-1}^+-B^+\}},\end{aligned} \tag{21}$$



(a) (b)

Figure 2: With respect to tap-out instant: (a) bricks of PDFs and (b) bricks of CDFs.

where $\sigma \equiv 1_{\min\{L, v(h_i^+ - h_{i-1}^+), v\Delta h_{u,i-1}^+ - B^+\} \geq (B^- + L - vx)^+}$. This gives the right part in $\dot{\psi}_{i,h}(x, v)$ as a tractable analytical formula. It is thus easy to compute function $\dot{\psi}_{i,h}(x, v)$.

As for integration over the distribution of walking speeds, the right part breaks into two bricks to which the lemma applies.

3.4. Illustration. By taking into account the previous distributions of v, s^+, s^-, the parameter vector is a six- or seven-fold vector $\boldsymbol{\theta} = \{\mu, \sigma, B^+, \lambda^+, B^-, \lambda^-, L\}$ (without σ for a constant speed $v = \mu$). During a studied period $[t_1, t_2]$ and a given h_u^+ (black circle), Figure 2(a) depicts the variations of functions $x \mapsto \dot{\psi}_{i,h}(x, v)$ and $x \mapsto \tilde{\dot{\psi}}_{i,h}(x)$ with respect to tap-out instant h_u^- starting from that of a given train arrival time h_i^- at the egress station (green dashed lines), either conditioning on constant (blue dashed line) or normal-distributed (red solid line) walking speed v. The brown dashed lines represent train departure times in access station. Figure 2(b) provides the corresponding integral functions $x \mapsto \psi_{i,h}(x, v)$ and $x \mapsto \tilde{\psi}_{i,h}(x)$. The difference between conditional and unconditional looks minor though discernible. For a given tap-in moment h_u^+, the functions relate to each train run i which the passenger could take. The most feasible train is the first arrival train at egress station: the latter the train arrival, the smaller the probability.

4. Statistical Estimation

Previous analytical formulae constitute the cores of our stochastic models. They derive analytical closed-form formulae that provide likelihood functions, which are tractable under a specification of basic statistical distributions. The stochastic models are theoretical constructs that involve human behaviors of trip-making in relation to the dynamic process of train runs. Such a theoretical model can be applied to particular cases, notably so by estimating the values of its parameters so as to make its outcomes replicate observed values well.

As reported by [25, 26], AFC data provide ample information on users' pairs of tap instants in stations, while AVL data provide both departure and arrival times in stations for each train run along its route [41].

In this section, we put forward an approach of MLE (Section 4.1). Its implementation involves a scheme for practical application (Section 4.1). Then, we build upon model estimation by proposing an inference method to enrich the observed data by adding "most likely predictions" of unobserved items (Section 4.1).

4.1. Maximum Likelihood Estimator. Let us assume here that TITO (h_u^+, h_u^-) pairs are observed over a journey between an O-D pair of access and egress stations. The model parameters consist in a six- or seven-fold vector $\boldsymbol{\theta} = \{\mu, \sigma, B^+, \lambda^+, B^-, \lambda^-, L\}$ (without σ for a constant speed $v = \mu$). Knowing $\boldsymbol{\theta}$, $h = h_u^+$, and train departure and arrival time pairs $\{(h_i^+, h_i^-) : i \in I\}$, the PDF $x \mapsto \dot{\mathrm{H}}_{h_u^+}^-(x = h_u^-)$ is a function of tap-out instant h_u^-. Conversely, given (h_u^+, h_u^-) and $\{(h_i^+, h_i^-) : i \in I\}$, the same formula can be interpreted as a function of $\boldsymbol{\theta}$, called the individual likelihood function for one trip of a passenger. It is then denoted as

$$\mathrm{L}_u(\boldsymbol{\theta}) = \dot{\mathrm{H}}_{h_u^+}^-(h_u^+). \tag{22}$$

The joint observation of a sample $\{(h_u^+, h_u^-) : u \in U\}$ of journeys provides a joint/total likelihood function, denoted $\mathrm{L}_U(\boldsymbol{\theta})$. If the observations are independent, then

$$\mathrm{L}_U(\boldsymbol{\theta}) = \prod_{u \in U} \mathrm{L}_u(\boldsymbol{\theta}). \tag{23}$$

The MLE consists to set up the value of parameter vector $\boldsymbol{\theta}$ so as to maximize the likelihood function of the observed sample, or equivalently to maximize the log-likelihood function as follows:

$$\begin{aligned} \Lambda_U(\boldsymbol{\theta}) &\equiv \ln \mathrm{L}_U(\boldsymbol{\theta}) = \sum_{u \in U} \Lambda_u(\boldsymbol{\theta}) \quad \text{and} \quad \Lambda_u(\boldsymbol{\theta}) \\ &= \ln \mathrm{L}_u(\boldsymbol{\theta}). \end{aligned} \tag{24}$$

That simplifies the computation of MLE.

The estimator of MLE can be applied to our stochastic models for either an individual passenger observed on several journeys or a set of passengers to differentiate 'intra-' versus 'inter-' individual cases. In the former case [24],

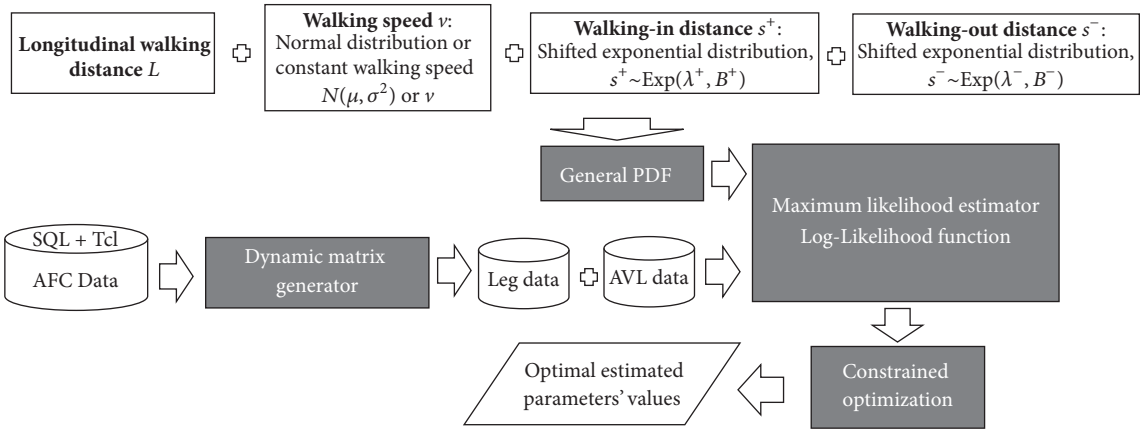


Figure 3: Modeling and optimization schema based on MLE.

the underlying statistical distributions are related to the particular passenger: his walking speed as either a constant or a distribution to allow for fluctuations and his own conditions for in-station walking distances and repositioning target L. In the latter case [21–24, 42], the condition is aimed to ensure statistical independence within the sample. The O-D stations must be shared by the passengers so as to give consistency to the L, S^+, and S^- notions. The former will be further investigated in this paper.

The estimator of MLE is endowed with powerful statistical properties that are well known in econometrics [43]. Although it has often some bias, this bias vanishes when the sample size tends to infinity; and the estimator variance is minimal. An especially valuable property is that the Hessian matrix of the log-likelihood function evaluated at the global optimum, up to a minus sign, contains the estimated covariance by pair of scalar parameters $\Lambda_U(\boldsymbol{\theta}) \equiv \ln \mathrm{L}_U(\boldsymbol{\theta}) = \sum_{u \in U} \Lambda_u(\boldsymbol{\theta})$. As these properties pertain to points of global maximization, a suitable optimization algorithm must be used. Furthermore, the properties rely upon a requirement of parameter-free domain for the RVs in the model. The requirement is satisfied by the normal distribution of speeds, but not by the shift in a shifted exponential distribution. So the application of MLE to our model is somewhat heuristic. However, under given B^+ and B^-, the estimation of the remaining parameters meets the domain condition.

4.2. Optimization. In practice, the estimator searches for the estimate $\widehat{\boldsymbol{\theta}}$ within an admissible space $\boldsymbol{\Theta}$ with bounds of vector components. To improve the optimization approach, a global pre-estimation of the space is proposed to find the optimal initial point. The pseudocode by nonlinear constrained optimization is shown in Figure 3.

The available AFC dataset was exploited by a specific dynamic O-D matrix inference scheme devised in [44], which extended previous works of [45, 46]. The scheme involves three principal steps as in [22]: (i) extracting the data of a given line from the dataset of the transit network, (ii) data filtering to exclude one-tap individual records and data inferring process for other oddities, and (iii) generating O-D pairs by scanning individual records to select appropriate trips.

4.3. Guidelines for Posterior Analysis. Based on parameter estimates, we can model each journey in the observed sample. The outcomes fall into three categories: (i) user's attributes of walking speed v and maximum longitudinal repositioning distance L; (ii) matching probabilities between train runs and passengers; and (iii) the level of probability associated with a pair of TITO times (h_u^+, h_u^-).

Matching probabilities associated with a given journey may be analyzed in three steps: (i) to identify the number n_u of feasible train runs I_u per trip by the model below from the number n_u^0 of hypothetical train runs compatible with $h_u^+ < h_i^+ < h_i^- < h_u^-$; (ii) among the feasible train run probabilities, to identify the biggest one; (iii) to evaluate the "dominance ratio" of the most likely run, the ratio between the second biggest and the biggest probabilities when the feasible train runs is more than 1.

Furthermore, based on the parameter estimates, we can make inference about some journey items that are not observed per se. Such ex-post analysis in a given journey of a passenger can be performed along the passenger's trajectory in the following way. In this model, assume that in each trip k of passenger, the train run with the biggest probability is taken, but conditioning on estimated average speed $\mathrm{E}(v)$. Main terms of passenger trajectory are derived directly by the following calculations:

(i) mean of truncated exponential distribution, $\mathrm{E}(s^+(k))$, between B^+ and B^+ plus maximum value compatible with times and speed $s_{\max}^+ = (h_i^+ - h_u^+)\mathrm{E}(v)$

(ii) repositioning distance, $s(k) = \min(s1, \mathrm{E}(L))$, where $s1 = s_{\max}^+ - \mathrm{E}(s^+(k))$, based on estimated L

(iii) residual waiting time, the waiting time in excess of repositioning time, $\widehat{w} = \max(0, s1 - \mathrm{E}(L))/\mathrm{E}(v)$

(iv) mean of truncated exponential distribution, $\mathrm{E}(s^-(k))$, between B^- and B^- plus maximum value compatible with times and speed $s_{\max}^- = (h_i^- - h_u^-)\mathrm{E}(v)$.

Hence, tap-out time h_u^- is equal to $h_u^+ + t^+ + \widehat{w}_u^i + L/\mathrm{E}(v) + (h_i^+ - h_u^+) + t^-$, where $t^+ = \mathrm{E}(s^+(k))/\mathrm{E}(v)$ and $t^- = \mathrm{E}(s^-(k))/\mathrm{E}(v)$. Once the tap-in times are given, the tap-out times are calculated.

5. Case Study of a Commuter Rail Line in Paris Area

The models are applied to a real case study, the busiest urban rail transit line RER A in Paris area, France, on the basis of AFC data provided by IdFM (ex STIF) and AVL data provided by RATP.

After introducing the case, observations of trip-making, and train traffic (Section 5.1), we estimate the model parameters with distributed walking speed for two samples of journeys (Section 5.2). Then, building upon the data and the estimated parameters, we infer the distance and time components for every sampled journey (Section 5.3).

5.1. Case Presentation: Navigo System, Line RER A, and Related Datasets. There are two main systems of urban rail transit in Paris area [47]: (i) the semiclosed metro system including 14 lines, equipped only tap-in gates, (ii) the heavier train system, including the "Transilien" and RER (Réseau Express Regional, the Regional Express Network) systems, and the closed RER system including 5 lines equipped both TITO gates, except for connected transfer stations between RER lines. The Paris transit system is integrated as concerns fares, and there is a unique smartcard called Navigo. Thus, the AFC system of transit in Paris area is called Navigo system. The Navigo system records anonymous passenger information including the smartcard number which is anonymized (with anonymous number that is maintained for only 3 months), the date, the validation instant at tap-in or tap-out gate, the gate ID, and the access or egress station name. During the PHs on workdays, more than 90% of the trips taken PT were home-work or home-study trips using network subscription hence the smartcard.

The line RER A is the busiest urban rail transit line in Paris area and maybe Europe, carrying more than one million passengers every workday [44]. It contains 46 stations in total 109 km and is structured around a central trunk into which are grafted five branches [48]: two eastward branches, northeast terminal Chessy and southeast terminal Boissy; and three westward branches, northwest terminal Cergy, central-west terminal Poissy, and southwest terminal Saint-Germain. The central trunk between stations Vincennes and La Défense passes through the largest underground mass transit hub, Châtelet-Les Halles, and serves the major business district La Défense in France.

The train time headways on the central trunk range from 2 min at peak to 10 min off peak. Our study focused on the O-D pair between Vincennes and La Défense on the central trunk. Each station has a number of entrances in relation to its importance, from 2 at Vincennes to 6 at La Défense. The topological structures of O-D pair Vincennes and La Défense are detailed in Figure 4, in which depicts the line platform at either station and indicates some of the passenger walking paths (green arrows) between tap gates and train doors. The 'Copy' nodes connect to the same destination nodes as the 'Copied' one to form in-station itineraries. There are more route choices in egress station. The complexity of rout choice in egress station comes from the choices between alighting doors and the platform exit points. There is another source of route choice in egress station, the choices between tap gates and station exits, which could not be considered only by the AFC data. The diversity of passenger paths is a source about the variability of in-station walking distance.

AVL and AFC datasets were made available to us by the line operator RATP and the mobility authority IdFM, respectively, for a period in March 2015 from the 16th to the 29th, excluding the 21st, 22nd, and 23rd. AVL data of RATP includes trains' arrival and departure times in stations. Out of the AFC data pertaining to the O-D pair Vincennes and La Défense in either direction, we selected one sample per direction, both for a given passenger with maximum number of such journeys during the period. As it turns out, the two busiest cards are identical, with 15 trips in Case 1 from La Défense to Vincennes and 16 trips in Case 2 from Vincennes to La Défense. Figure 5(a) exhibits the TITO times of the 31 trips (5 trips in 1 day, 4 trips in 4 days, 3 trips in 3 days, 2 trips in 0 day, 1 trip in 1 day, and 0 trips in 2 days). All trips are during workdays, except one trip on Sunday March 29th 2015. Indeed, the trip pattern is peculiar since every trip from Vincennes to La Défense is followed by a return trip after a short period of about two hours in the morning or one hour in the evening; those trips would be home-work trips. Figure 5(b) evaluates the related journey times: on average, the journey takes more time from Vincennes to LaDéfense during PHs than that from La Défense to Vincennes during OPHs.

5.2. Estimations. The admissible spaces of scalar parameters were specified on the basis of field measurement or literature on urban mobility. As for in-station walking distances, a range of $[4, 325]$ m applies to s^+ in Vincennes, and a range of $[4, 525]$ m applies to s^- in La Défense. Platform lengths amount to 225 m in each station: yet we allowed a feasible space of $[0, 600]$ m for L, as it may be bigger than the sum of access and egress station platform lengths 450 m. A feasible range of $(0, 4]$ m/s is imposed on average walking speed v or μ, based on the last household travel survey in Paris area [49]. This walking speed v includes the relative speed of escalator or lift. Based on previous given bounds of variables, reasonable ranges for all estimated parameters are defined, as reported in Table 2.

While the log-likelihood function of the model gets its maximum value, parameters' estimates and their Standard Deviations (SDs, approximations) are obtained in Table 2. The results show averaged day-to-day dynamic variation of this passenger. It appears that the estimates of the target repositioning distances, $\widehat{L}_1$ and $\widehat{L}_2$, are relatively consistent and fairly precise. The same applies to the distance shifts $\widehat{B}_1^+$, $\widehat{B}_2^+$, and $\widehat{B}_1^-$, but $\widehat{B}_2^-$ is less precise. About the main parameters of in-station distances, the estimated values are small for $\widehat{\lambda}_1^+$, $\widehat{\lambda}_1^-$, $\widehat{\lambda}_2^+$, and $\widehat{\lambda}_2^-$: the first of them is not significantly different

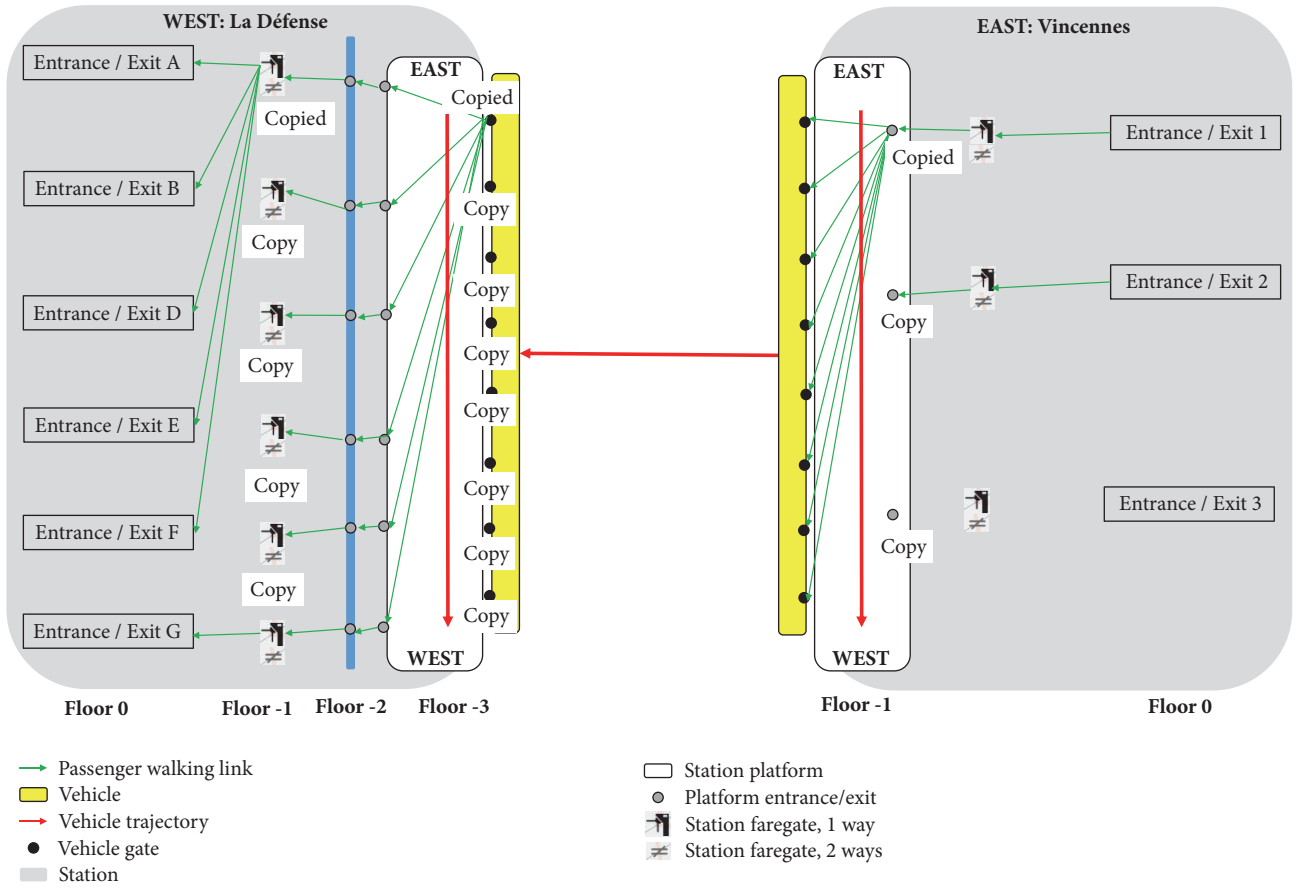


Figure 4: Sketch of stations' topological structures: Vincennes and La Défense.

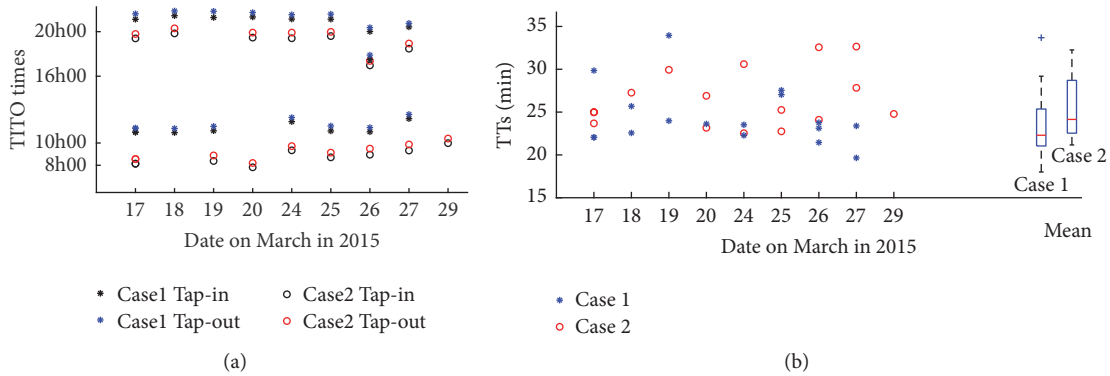


Figure 5: Trips made by the busiest passenger during nine days in March 2015.

from zero. As for speed estimates, the average values are fair in Case 2 and tolerable in Case 1. However, the discrepancy between the two cases calls for further investigation, since the data pertain to the same user for whom an identical walking speed is expected in both directions. The parameter of speed dispersion, σ, has a low estimate in Case 2 but a bigger one in Case 1.

In all, the consideration of an individual passenger enabled us to recover meaningful information about his or her trip-making behavior. The target repositioning distances are significant, at about 101 or 76 m depending on the direction. Combined to the respective estimates of average speeds, these distances correspond to repositioning time of about 60 s in both cases. This value is close to half of a time headway at peak hours. The in-station distances exhibit a mirror effect at Vincennes station (similarity between s_2^+ and s_1^-), but there is a discrepancy at La Défense station between s_1^+ and s_2^-. A potential reason may pertain to high passenger crowding that occurs at La Défense during PHs, which makes individual walking more dependent on crowd dynamics,

Table 2: Parameter estimation for a single passenger by MLE.

Indicator (Unit)	Parameter (Unit)	Range	Case 1		Range	Case 2	
			Estimated value	Standard deviation		Estimated value	Standard deviation
v (m/s)	μ (m/s)	(0, 4]	1.70	0.054	(0, 4]	1.19	0.21
	σ	(0, 4]	0.33	0.059	(0, 4]	0.07	0.01
s^+ (m)	λ^+ (1/m)	(0, 2]	0.017	0.015	(0, 2]	0.73	0.03
	B^+ (m)	(0, 550]	126.86	14.72	(0, 150]	106.29	10.14
s^- (m)	λ^- (1/m)	(0, 2]	0.91	0.042	(0, 2]	0.10	0.01
	B^- (m)	(0, 150]	93.49	2.78	(0, 550]	70.52	28.33
L (m)	L (m)	(0, 600]	101.46	8.28	(0, 600]	76.14	6.35
Log-likelihood function			-47.48			-49.01	

Table 3: Indicators' mean values.

Mean value (Unit)	Range	Case 1		Range	Case 2	
		Estimated mean value	Standard deviation		Estimated mean value	Standard deviation
E(v) (m/s)	(0, 4]	1.70	0.33	(0, 4]	1.19	0.07
E(s^+) (m)	[4, 525]	186.75	59.89	[4, 325]	107.66	1.37
E(s^-) (m)	[4, 325]	94.59	1.10	[4, 525]	80.52	10.00
E(L) (m)	[3, 600]	101.46	8.28	[3, 600]	76.14	6.35
E(t^+) (min)	NA	1.83	NA	NA	1.51	NA
E(t^-) (min)	NA	0.93	NA	NA	1.28	NA

∗NA: not available.

especially so at station egress. The difference between the parameter values in the same station indicates the difference of walking-in and walking-out routes. The difference of values between different stations depicts the station topological structures.

Mean walking speed and mean walking distances are calculated based on parameter estimates and illustrated in Table 3. Passenger's mean walking speed with normal distribution is equal to $\hat{\mu}$ with SD $\hat{\sigma}$. As regards the walking distances with shifted distributions, the mean values are recovered as $\hat{B} + 1/\hat{\lambda}$ with SD $1/\hat{\lambda}$. The mean value of longitudinal walking distance is estimated directly by the model. Table 3 provides also some more indications about in-station mean walking times, derived directly by E(s^+)/E(v) and E(s^-)/E(v). The results are consistence with the above comments. It suggests that, for a given passenger, walking times may be more reliably estimated than the pair of walking distances and walking speed.

In the Appendix, estimation results for a former model neglecting the repositioning behavior are recalled, for the same O-D pair but a single day of observation and a population set of passengers. The journey is from Vincennes to La Défense, which corresponds to Case 2 here. There is much agreement between the intraindividual estimates of Case 2 and the interindividual estimates of model M1 (normal-distributed walking speed) in the Appendix, except for the shift parameter of in-station distances on the egress side. In fact, $\hat{L}_2 + \hat{B}_2^- \approx 140$ m are closer to $\hat{B}_A^- \approx 204$ m, which seems to confirm that repositioning strategy on the boarding platform constitutes a distance investment by the users for the rest of the journey.

5.3. Posterior Analysis to Infer Journey Details. Similar to Bayesian analysis, parameter estimates are used for reanalyzing train run occurrences and passenger trip attributes. It will check the rationality of previous results.

Based on the sampled data and parameter estimates, the journey elements are inferred following the lines given in Section 4.3. About the matching probabilities, Figures 6(a) and 6(b) depict the numbers of hypothetical n_u^0 and feasible n_u train runs per trip, respectively. Figure 6(b) confirms that the postulate in the Bayesian analysis will be a good approximation. Figure 6(c) shows the maximum probability to take a train run among all n_u feasible train runs. The maximum probability for a tip to take a train is equal to 1 when there are only one feasible train. For the rest, the results show that most of the maximum probabilities to take a train are bigger than 70.66%, but one is 1×10^{-16}. on March 17th. This exception can be caused by the bias of AVL data measurement. Figure 6(d) calculates the dominance ratios among n_u while $n_u > 1$. Since the ratios are so small, smaller than 1×10^{-3}, it shows that there is only one train that is feasible while $n_u > 1$ as well.

A disaggregate analysis of passenger individual trip attributes is proposed by using the passenger trip model proposed in Section 4.3. Passenger trajectories are reproduced by the simple inference model based on previous estimated values. Figure 7 gives each observed trip and its referred result, Case 1 in the left column and Case 2 in the right column. To confront the real tap-out times, Figures 7(a1) and 7(a2) compare the inferred tap-out times (red circle) to the observed ones (blue asterisk). Since the differences (black cross) between them are very small, the agreement is very good, which indicates that the stochastic model has good

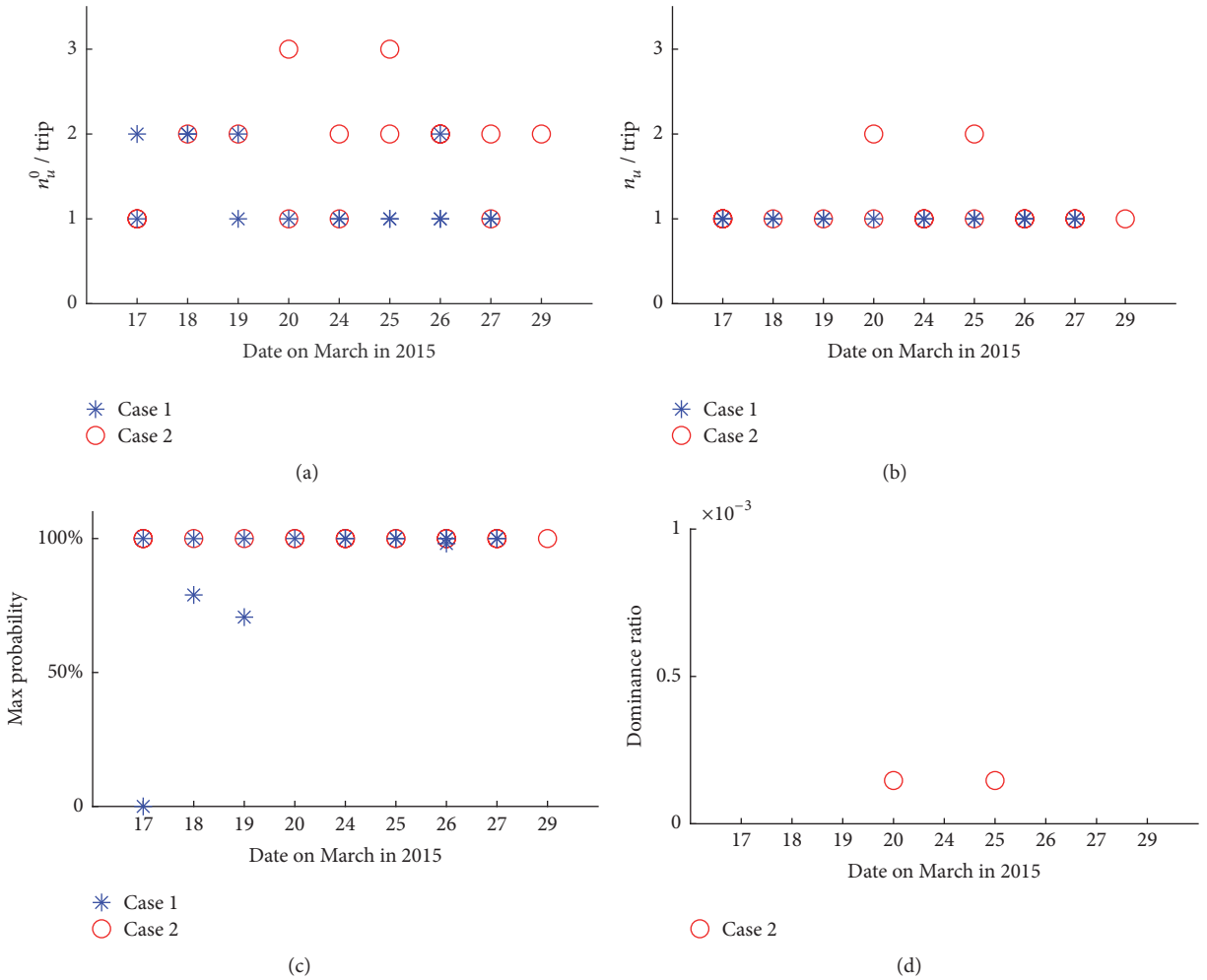


Figure 6: Run number and run probabilities per trip.

predictive ability. Figures 7(b1) and 7(b2) about positioning distances show that the total repositioning could be achieved in all but one occurrence. Lastly, Figures 7(c1) and 7(c2) illustrate the residual waiting times obtained by inference: the values are higher at La Défense (Case 1) than those at Vincennes (Case 2), certainly because Case 1 occurs mostly off peak while Case 2 occurs mainly at peak.

6. Conclusions and Perspectives

This section assesses the outreach and limitations of our model and points to directions for further research.

6.1. Summary and Outreach. This paper provided a stochastic model of passenger trip-making along a transit journey by urban rail line, with explicit representation of individual positioning along the boarding platform and the optimizing behavior to save on travel time for the rest of the trip.

The behavioral postulate was appropriate for passengers well aware of the trip conditions at their egress station. This fits well commuters—hence the vast majority of transit users at peak periods—and also customers availing themselves of "travel assistant" applications on their smartphones.

The stochastic model was easy to use in the perspective of simulation, as it followed the physical sequence of phases in a journey path (walking in, platform positioning and waiting, train riding, and walking out). It could readily be applied as a submodel in the frame of a traffic assignment model to a transit network.

While the simulation ability was demonstrated in the case study, the paper was primarily oriented to the estimation perspective: analytical formulae were given to characterize the statistical distributions of egress times and exit instants that stem from the set of modeling assumptions. The CDF and PDF formulae conveyed the influences of individual attributes and behavior (speed and target relocation distance), along with those of local conditions, i.e., in-station distances for the walking phases.

In the estimation perspective, we used the PDF formulae as likelihood functions for the model parameters. We put forward particular yet realistic enough specifications for the statistical distributions, so as to make numerical computation more tractable.

An application was carried out to an O-D pair of stations along a busy rail line in Paris. AFC and AVL data were extracted for the trips of an individual user over a two-week

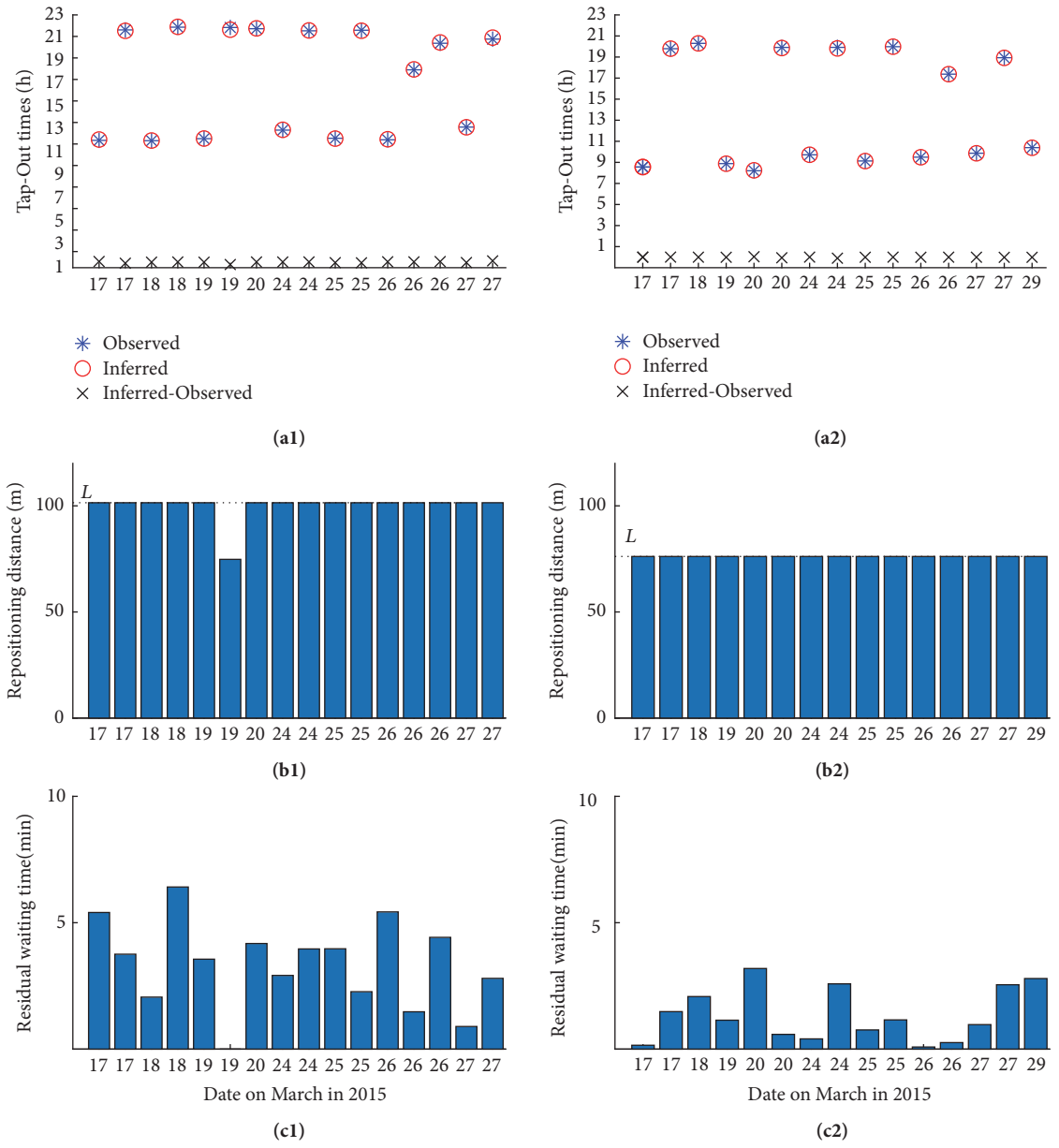


FIGURE 7: Inferred results about: (a) tap-out time; (b) repositioning distance; and (c) residual waiting time.

period. Valuable information was recovered from statistical estimation and posterior inference, notably a time saving of about 1 min owing to platform repositioning along 70 or 100 m, depending on the journey direction. This indicated that the estimation scheme was able to capture fine phenomena and also that the repositioning phenomenon had a limited importance on the journey travel time of this individual user.

This also demonstrated once again the positioning behavior of train users along boarding platforms, which was of interest to railway operators for passenger traffic management under severe crowding. The consequences for station layout and flow orientation ware traced out in related work [19].

6.2. Potential Applications and Limitations. Where AFC and AVL datasets are available, our model can easily be applied to estimate the distance and time components of users' journeys in a gate-to-gate setting which represents quality of service better than just the service quality of the train runs. The identification of positioning will make the estimation of the remaining time components more realistic and reliable. The individual behavior, as postulated and estimated on the basis of empirical data, can easily be simulated for users' journeys whatever the availability status of observations is, because the behavioral structure is endowed with replicability.

Despite its finesse, the stochastic model in its current version does not capture congestion phenomena: neither in-vehicle crowding or the potential restrictions (i.e., the probability of fail-to-board and the related issue of passengers "left behind" by trains), the crowding of platforms and its potential influence on individual positioning, nor the

Table 4: Estimated parameters' values by MLE in M1 and M2.

Parameter (Unit)	Range	M1		M2	
		Estimated value	Standard deviation	Estimated value	Standard deviation
M1: μ (m/s); M2: a (m/s)	$(0, 4]$ and $a < b$	0.90	0.0023	0.58	0.07
M1: σ; M2: b (m/s)	$(0, 4]$	0.20	0.0007	1.24	0.13
λ^+ (1/m)	$(0, 2]$	0.78	0.0048	0.32	0.18
B^+ (m)	$(0, 150]$	83.49	0.7314	72.32	6.81
λ^- (1/m)	$(0, 2]$	0.13	0.0002	0.18	0.02
B^- (m)	$(0, 550]$	204.30	0.7998	264.46	26.66
Log-likelihood function		-109.85		-138.18	

Table 5: Indicators' mean values in M1 and M2.

Parameter (Unit)	Range	M1		M2	
		Estimated mean value	Standard deviation	Estimated mean value	Standard deviation
E(v) (m/s)	$(0, 4]$	0.90	0.04	0.91	0.19
E(d^+) (m)	$[4, 325]$	84.77	1.29	75.41	3.09
E(d^-) (m)	$[4, 525]$	212.22	7.93	270.32	5.66

crowding of platform access points and especially egress points, which may entail queuing and delay among exiting passengers.

6.3. Further Developments. So the consideration of crowding phenomena makes up a first direction for further research on passenger behavior along a rail journey.

A second direction is to devote more attention to the in-station phases, especially so for vertical pedestrian elements that influence individual speed under congestion as well as free-flow conditions. While these issues are well known in the micro-simulation of pedestrian traffic (cf. the Legion and Viswalk modeling software, among others), their estimation on the basis of AFC and AVL data is an open issue.

A third direction for research is to extend the stochastic model with platform repositioning to more complex trip patterns that involve transfers: this is our next objective.

Lastly, more detailed data of users' trajectories are available from smartphones owing to applications that monitor geolocation data from one or several sources–GPS, GSM, or beacons Wifi or Bluetooth. Indeed, location data collected every second say and with fine GPS or Galileo accuracy constitute ideal material for the refined analysis of passenger trip-making. Such research remains to be done for large underground transit stations, where satellite or beacon signals are impeded or modified by local layout—corridors, walls, floors, and ceilings.

Appendix

The results of previous models without passenger longitudinal walking distance in [23] are presented. The two stochastic models M1 and M2 are normal-distributed and uniform-distributed walking speeds models. The tested datasets are all passengers' trips from Vincennes to La Défense on March 16th 2015. Table 4 shows the parameters' values and Table 5 introduces the indicators' mean values.

Conflicts of Interest

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

Acknowledgments

This research was supported by the Research and Education Chair on "the Socio-Economics and Modeling of Urban Transit," operated by Ecole des Ponts ParisTech (ENPC) in partnership with the Mobility Authority in the Paris area (IdFM, Île-de-France Mobilités, ex STIF), to whom the authors are grateful. The authors also thank the Autonomous Operator of Parisian Transit (RATP) who provided the train data.

References

[1] TCQSM. Transit Capacity and Quality of Service Manual (TCQSM). 3rd edition. TRB's Transit Cooperative Highway Research Program (TCRP) Report 165, 2013.

[2] V. Benezech and N. Coulombel, "The value of service reliability," *Transportation Research Part B: Methodological*, vol. 58, pp. 1–15, 2013.

[3] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.

[4] F. Leurent, E. Chandakas, and A. Poulhès, "A traffic assignment model for passenger transit on a capacitated network: Bi-layer framework, line sub-models and large-scale application," *Transportation Research Part C: Emerging Technologies*, vol. 47, no. 1, pp. 3–27, 2014.

[5] J. Chan, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data [M, thesis]*, Department of Civil and Environmental Engineering, 2007.

[6] M. Friman, "Affective dimensions of the waiting experience," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 13, no. 3, pp. 197–205, 2010.

[7] Y. Fan, A. Guthrie, and D. Levinson, "Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security," *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 251–264, 2016.

[8] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Users' views on current and future real-time bus information systems," *Journal of Advanced Transportation*, vol. 47, no. 3, pp. 336–354, 2013.

[9] C. Brakewood, G. S. Macfarlane, and K. Watkins, "The impact of real-time information on bus ridership in New York City," *Transportation Research Part C: Emerging Technologies*, vol. 53, pp. 59–75, 2015.

[10] G. S. Mishra, P. L. Mokhtarian, and K. F. Widaman, "An empirical investigation of attitudes toward waiting on the part of Northern California commuters," *Travel Behaviour and Society*, vol. 2, no. 2, pp. 78–87, 2015.

[11] M. A. Ramli, V. Jayaraman, K. H. Chee et al., "Improved estimation of commuter waiting times using headway and commuter boarding information," *Physica A: Statistical Mechanics and its Applications*, vol. 501, pp. 217–226, 2018.

[12] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions depending on train service frequency and station characteristics: A smart card data analysis," *Transportation Research Part C*, 2018.

[13] G. S. Becker, "A theory of the allocation of time," *The Economic Journal*, vol. 75, no. 299, p. 493, 1965.

[14] D. Lois, A. Monzón, and S. Hernández, "Analysis of satisfaction factors at urban transport interchanges: Measuring travellers' attitudes to information, security and waiting," *Transport Policy*, 2017.

[15] A. Bouzir, H. Bouguerra, S. Rjiba, and S. Benammou, "Passengers' activities in multimodal transportation stations," in *Proceedings of the International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*, pp. 76–81, 2017.

[16] A. Millonig, M. Sleszynski, and M. Ulm, "Sitting, waiting, wishing: Waiting time perception in public transport," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation*, 2012.

[17] N. Ohmori, T. Hirano, and N. Harata, "Passengers' waiting behavior at bus and tram stops," *Traffic and Transportation Studies*, pp. 520–531, 2008.

[18] M. van Hagen, *Waiting Experience at Train Station [Doctoral, thesis]*, Eburon Uitgeverij, 2011.

[19] Z. Christoforou, P. Collet, B. Kabalan et al., "Influencing Longitudinal Passenger Distribution on Railway Platforms to Shorten and Regularize Train Dwell Times," *Transportation Research Record*, vol. 2648, pp. 117–125, 2017.

[20] Z. Yizhou, *Real Time Crowding Information (RTCI) Provision [Ph.D. thesis]*, KTH, 2015.

[21] F. Leurent and X. Xie, "Exploiting smartcard data to estimate distributions of passengers' walking speed and distances along an urban rail transit line," *Transportation Research Procedia*, vol. 22, pp. 45–54, 2017.

[22] X. Xie and F. Leurent, "Estimating distributions of walking speed, walking distance, and waiting time with automated fare collection data for rail transit," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2648, no. 2, pp. 134–141, 2017.

[23] X. Xie and F. Leurent, "Comparison of passenger walking speed distribution models in mass transit stations," *Transportation Research Procedia*, vol. 27, pp. 696–703, 2017.

[24] F. Leurent and X. Xie, "On passenger repositioning along station platform during train waiting," *Transportation Research Procedia*, vol. 27, pp. 688–695, 2017.

[25] 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.

[26] T. Kusakabe and Y. Asakura, "Combination of smart card data with person trip survey data," in *Public Transport Planning with Smart Card Data*, pp. 73–92, CRC Press, 2016.

[27] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior," *Transportation Research Record*, no. 2275, pp. 58–67, 2012.

[28] 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. 10, Article ID 04015037, pp. 36–51, 2016.

[29] J. D. Schmöcker, *Dynamic capacity constrained transit assignment [Ph.D. thesis]*, Imperial College London, London, UK, 2006.

[30] Y. Zhu, *Passenger-to-train assignment model based on automated data [M. Thesis]*, S.M. Massachusetts Institute of Technology, 2014.

[31] 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.

[32] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transportation Research Procedia, Transportation Research Part C: Emerging Technologies*, vol. 23, Supplement C, pp. 362–379, 2017.

[33] Y. Zhu, *Passenger-to-Itinerary Assignment Model Based on Automated Data [Ph.D. thesis]*, 2017.

[34] D. Hörcher, D. J. Graham, and R. J. Anderson, "Crowding cost estimation with large scale smart card and vehicle location data," *Transportation Research Part B: Methodological*, vol. 95, pp. 105–125, 2017.

[35] A. Sumalee, Z. Tan, and W. H. K. Lam, "Dynamic stochastic transit assignment with explicit seat allocation model," *Transportation Research Part B: Methodological*, vol. 43, no. 8-9, pp. 895–912, 2009.

[36] F. Leurent, "Modelling seat congestion in transit assignment," in *Proceedings of the 17th ISTTT conference*, 2006.

[37] F. Leurent, "A model of travel and stop times, with application to the estimation of travel time functions," in *Proceedings of the 7th Meeting of the EURO Working Group on Transportation*, Roma, Italy, 2000.

[38] F. Leurent, *Modèle désagrégés du trafic. Book "Les collections de l'inrets, Outils et Méthodes N10"*, 2001.

[39] A. M. Wahaballa, F. Kurauchi, T. Yamamoto, and J. Schmöcker, "Estimation of Platform Waiting Time Distribution Considering Service Reliability Based on Smart Card Data and Performance Reports," *Transportation Research Record*, vol. 2652, pp. 30–38, 2017.

[40] D. Aigner, C. A. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.

[41] B. Almadani, S. Khan, R. Tarek et al., "Automatic Vehicle Location and Monitoring System based on Data Distribution Service," *Procedia Computer Science*, vol. 37, pp. 127–134, 2014.

[42] Xie. Xiaoyan and Fabien. Leurent, "Spatial and temporal analyses about distributions of walking speed and distances," in *Proceedings of the 44th European Transport Conference*, Barcelona, Spain, 2016.

[43] G. G. Judge, R. C. Hill, W. E. Griffiths et al., *Introduction to The Theory And Practice of Econometrics*, Wiley, 1st edition, 1987.

[44] X. Xie, F. Leurent, and V. Aguiléra, "Commuter's individual travel time estimation: using AFC data of Paris transit system," in *COST Action TU1004 Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Paris, France, 2015.

[45] J. Zhao, M. Frumin, N. Wilson, and Z. Zhao, "Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 70–88, 2013.

[46] 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.

[47] V. Aguiléra, S. Allio, V. Benezech, F. Combes, and C. Milion, "Using cell phone data to measure quality of service and passenger flows of Paris transit system," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 198–211, 2014.

[48] RATP, 2015, http://www.ratp.fr.

[49] Omnil, *Enquête globale de transport, La mobilité en Île-de-France: la marche*, 2013, http://www.omnil.fr.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 3096190, 7 pages
https://doi.org/10.1155/2018/3096190

Research Article

Fuzzy Approach in Rail Track Degradation Prediction

Mostafa Karimpour,[1] **Lalith Hitihamillage**,[1] **Najwa Elkhoury**,[1] **Sara Moridpour**,[1] **and Reyhaneh Hesami**[2]

[1]*Civil and Infrastructure Engineering, RMIT University, Melbourne, Australia*
[2]*Asset planning and Visualisation, Yarra Trams, Melbourne, Australia*

Correspondence should be addressed to Mostafa Karimpour; mostafa.karimpour1@gmail.com

Received 9 October 2017; Revised 13 February 2018; Accepted 15 April 2018; Published 24 May 2018

Academic Editor: Lingyun Meng

Rail transport authorities around the world have been facing a significant challenge when predicting rail infrastructure maintenance work. With the restrictions on financial support, the rail transport authorities are in pursuit of improved modern methods, which can provide a precise prediction of rail maintenance timeframe. The expectation from such a method is to develop models to minimise the human error that is strongly related to manual prediction. Such models will help rail transport authorities in understanding how the track degradation occurs at different conditions (e.g., rail type, rail profile) over time. They need a well-structured technique to identify the precise time when rail tracks fail to minimise the maintenance cost/time. The rail track characteristics that have been collected over the years will be used in developing a degradation prediction model for rail tracks. Since these data have been collected in large volumes and the data collection is done both electronically and manually, it is possible to have some errors. Sometimes these errors make it impossible to use the data in prediction model development. An accurate model can play a key role in the estimation of the long-term behaviour of rail tracks. Accurate models can increase the efficiency of maintenance activities and decrease the cost of maintenance in long-term. In this research, a short review of rail track degradation prediction models has been discussed before estimating rail track degradation for the curves and straight sections of Melbourne tram track system using Adaptive Network-based Fuzzy Inference System (ANFIS) model. The results from the developed model show that it is capable of predicting the gauge values with R^2 of 0.6 and 0.78 for curves and straights, respectively.

1. Introduction

Modern transport organizations have shifted their focus from construction and expansion of the transport infrastructure towards how to intelligently maintaining them. This has taken place due to many reasons such as budget restrictions and running out of land space. Transport organizations currently focus on exploring the solutions for developing a maintenance management system that will help them accurately predict the time and location where maintenance should be carried out. This will assist transport infrastructure authorities in optimizing the cost and time of maintenance.

Different types of degradation prediction models have been presented in the literature, and these models have been mainly developed for the heavy rail system. Since there are differences in the structure and performance of heavy and light rail systems, it is impossible to use such degradation prediction models to predict the degradation of the light rail system. Consequently, it is needed to develop models which are capable of predicting the degradation of light rail tracks. This particular research will focus on developing a degradation prediction model for light rail network with the focus on tram network of Melbourne, Australia. The map of the current Melbourne tram network is shown in Figure 1.

The maintenance data of the Melbourne tram network has been collected through on-sight inspection and stocked in a nondigitized way for a long time. Traditionally, the rail maintenance used to be planned based on the experience of experts in the field. This procedure has changed since the introduction of new rail track inspection vehicles. These vehicles run through rail tracks and detect a large amount of data related to gauge and twist values from infrastructure condition. Based on this data, prediction models will be developed to predict the degradation of rail tracks

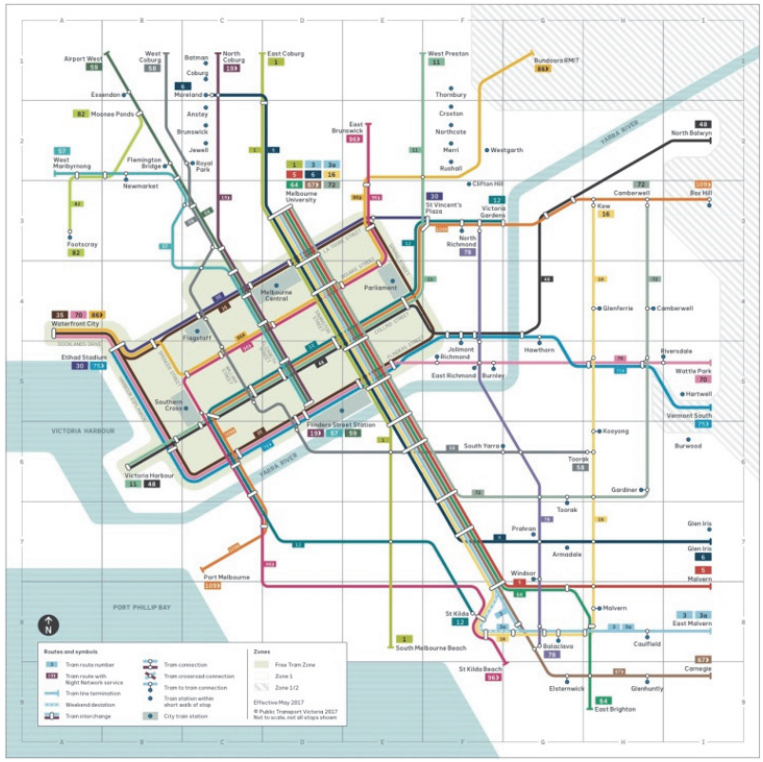


Figure 1: Melbourne tram network.

and estimate the maintenance procedures needed in the future.

In this paper, a review of the previous models on rail degradation prediction is presented. Afterwards, an ANFIS model is proposed to predict the Melbourne tram track degradation. Results from the models for both curves and straight sections are presented then. Finally, conclusions of this research and directions for further work on this topic are presented.

2. Literature Review

According to the literature, the models which have been used to predict rail track degradation could be categorized into three main categories including mechanistic, statistical, and artificial intelligence (AI) models. Out of these three main types, some of them are further categorized into subcategories as in Figure 2.

Previous studies on rail track degradation have represented some models that are capable of predicting degradation using a common set of parameters such as the age of the rail, axle load in Million Gross Tone (MGT), speed, and track curvature.

The mechanistic approach is the oldest model type in predicting the rail track degradation. Mechanistic models are based on the knowledge and understanding of the behaviour of the mechanical components. Mechanistic models involve establishing the mechanical properties either by theory or by testing. These types of models could mainly develop based on laboratory experiment data or collecting data from the field by observing the sections that will be used for modeling for a long period of time. Since such type of models has the capability of predicting rail track settlement and degradation with greater accuracy, according to the literature that is one of the key strengths of this type of models. So this is a type of model more suitable for utilizing in a situation where just one particular section of the track needed to be repaired. However, with this positive, there are few drawbacks that limit the possibility of using this type of models for predicting the future degradation and settlement values. The main drawback of this type of models is that it will not provide predictions with greater accuracy when you try to apply the model to the different sections of the rail network. The reason behind this is because the mechanical properties of the track and external factors they face may vary from place to place. Since it requires an extensive amount of data to develop a model small section of the rail track, developing a model for different sections of the network could be challenging and very time-consuming. Some of the models developed by the Japanese researchers [1–3] and German researchers [4] are among the most famous mechanistic models which have been pioneered, developing degradation prediction models based on mechanistic models around the globe.

The statistical models are based on the data collected from monitoring the track performances and the variables affecting such performances (e.g., traffic, rail type, and maintenance data). Those variables are used as inputs to develop a model which predicts track degradation. Statistical models provide the ability to deal with the considerably large amount of datasets when developing degradation models and they

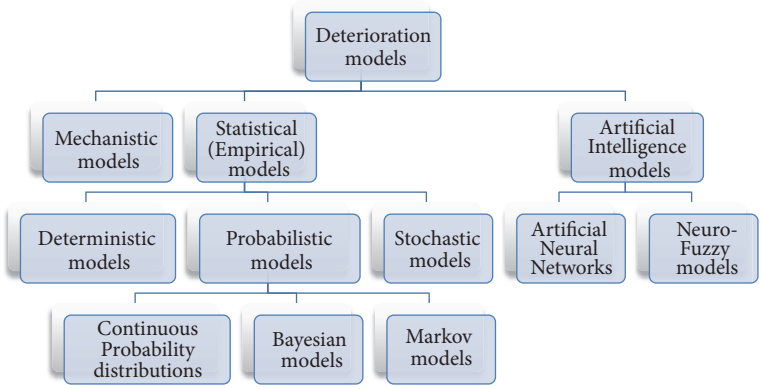


Figure 2: Rail track degradation prediction models.

provide more accurate results when applied to an entire rail network compared to a mechanistic model. However, it also has some downsides such as being able to create a large database which these types of models require due to the lack of availability of historical track data. Some of these statistical models could be further categorized into deterministic [5], probabilistic [6], and stochastic models [7].

The artificial intelligence models are a group of machine learning models which primarily could be categories into two groups' including artificial neural networks (ANNs) and neurofuzzy models. Artificial neural networks were originally inspired by the biological neural networks which are present in the human brain. These models are both simple computational devices which are highly interconnected, and in both models the connections between neurons determine the function of the network. Although ANNs are quite new to tram track degradation predictions, they have been extensively used in other fields of engineering [8]. ANNs are currently on the verge of becoming popular in predicting the degradation of railways due to their capability of providing prediction results with high accuracy [9, 10]. They have also been utilized to evaluate the rail track condition [9]. The work of [9] proposed an ANN Model for track degradation prediction and used parameters such as Track Record Index (TRI) which defines the track quality. Moreover, traffic volume (i.e., light, heavy), speed, geographic location curves radius, and gradient are other factors used to predict the rail degradation [11]. Another study presented an artificial neural network model to predict the degradation of tram tracks using maintenance data in Melbourne railway system. Data was collected from Melbourne tram network from 2009 to 2013, covering different types of segments of four routes such as straights, curves, H-crossings, and crossovers [12]. Out of these segments, curves were the focus since they have a higher failure rate than some other segments [13, 14]; this section's data was utilized in case of light rail degradation modeling using a nonlinear method [15]. However, the results presented in previous researchs were far less accurate than what this paper has gained due to the higher nonlinear capability of ANFIS model.

In this research, at first the data are preprocessed, and all the outliers are eliminated. Following that, a fuzzy model is proposed due to the nonlinear and noisy nature of the data according to the data which were available on the Melbourne tram network. In the next stage, the model is validated on the test samples. The final year gauge values have been considered as the measured value for both training and testing experiments. Model validation/testing was done by comparing the final year gauge values as observed values and the gauge values for the final year estimated by the model.

Since gauge value represents the physical shape of railways, it can be considered as a factor which comprises various important features of rail degradation in its nature. Thus, rail degradation maintenance scheduling could be done by considering gauge degradation as a paramount factor. In this paper, gauge value degradation prediction modeling is done which could be a breakthrough into degradation hotspot detection which leads to lower amount of investment in both rail monitoring and preventive maintenance.

3. Dataset

Melbourne tram network is the world's largest urban operating tram network and covers 250 km of double tracks which includes 25 routes. There are 1700 stops across the network spread out with more than 400 level access stops. Seventy-five percent of the Melbourne's tram network operates on shared roads with other vehicles and manages to provide a service with a punctuality percentage around the high seventies to mid-eighties. To meet the current demands of the rapid growth of Melbourne city, its iconic tram network also needs to evolve accordingly. Along with the increase in Melbourne tram patronage, it has been understood that the expenses related to rail infrastructure maintenance have grown gradually and constantly.

Rail track degradation occurs due to many reasons. Rail vehicles travel at various speeds while carrying various loads. This will cause a wide range of stresses on the rail structure resulting in its decay. When it comes to the light rail tracks, the degradation of the track occurs due to few more other reasons such as the damage occurring by road sharing and weathering due to climate change. Since all these changes influence the decay and are embedded in gauge value, it is considered as the key value for the degradation. Gauge value is defined as the distance between the inner sides of two rails on a railway system.

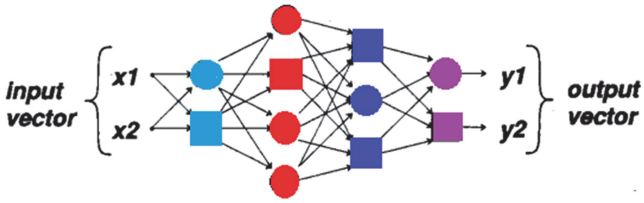


Figure 3: Simple adaptive network.

In this particular study, data are provided by Yarra Trams (which is the operator of the Melbourne tram network and is responsible authority for its maintenance). This data is utilized to develop a degradation model which will provide the ability to predict the future gauge values. The dataset which was provided by them for this research comprises different section types including curves, straight sections, H-crossing, and crossovers. The gauge values have been collected by Yarra Trams for the whole tram network at different years. In addition, curve radius, annual tonnage in Million Gross Tone (MGT), track surface (asphalt and concrete surfaces), rail profile (the cross-sectional shape of a rail which is represented by kilogram per metre), rail type (grooved and T-shapes), rail support (or rail ties categorized into concrete and steel sleepers), location of routes, and track installation date have also been provided by Yarra Trams. These data were gathered over a period of 6 years from 2009 to 2015. Curve and straight sections where the two major groups of the track sections used in this study to analyze and develop the degradation prediction model since they represent the majority of the tram network.

4. Model Development

In this paper, ANFIS is used to estimate the gauge value for $t + 1$ if the data for the past are available. The dataset is consisting of 3,860 different samples of gauge values between 2010 and 2015. Numerous sections of the railway had minor or major maintenance through these years. Thus, linear models are not able to consider the influence of different maintenance types over the years. Therefore, in this paper, an ANFIS model is proposed, and the results show the superiority of nonlinear models in the maintenance modeling. The ANFIS model presented in this paper consists of three important inputs including the two previous values of gauge (the previous two readings of the gauge from the past two years) and the MGT value. The dataset is divided into training and test set. 70% of the data is used for training purpose while the rest is used for testing.

An adaptive network can be considered as Figure 3 and is a feedforward multilayer network in which each node plays a particular role in the input with a set of parameters relating to the node [16]. Circle nodes have no parameter while the square nodes which are adaptive have different parameters that need to be estimated. If the network has L layers and the jth layer has $\#(j)$ nodes, the node in the ith position of jth layer can be written as

$$O_i^j = O_i^j\left(O_1^{j-1}, \ldots, O_{\#(j-1)}^{j-1}, m, n, \ldots\right), \tag{1}$$

where m, n, etc. are the parameters related to this node. O_i^j represents the node output and the function. Considering a set of training data has q entries the sum squared error could be measured as

$$E_q = \sum_{m=1}^{\#L}\left(T_{m,q} - O_{m,q}^L\right)^2, \tag{2}$$

where $T_{m,q}$ is the mth component of qth target. The rate of error from the output node at L, i, can be derived from

$$\frac{\partial E_q}{\partial O_{i,q}^j} = \sum_{m=1}^{\#(j+1)} \frac{\partial E_q}{\partial O_{m,q}^{j+1}} \frac{\partial O_{m,q}^{j+1}}{\partial O_{i,q}^j}. \tag{3}$$

If θ is a parameter of the network $\partial E_q/\partial\theta$ can be written as

$$\frac{\partial E_q}{\partial \theta} = \sum_{O^* \in S} \frac{\partial E_q}{\partial O^*} \frac{\partial O^*}{\partial \theta} \tag{4}$$

S is the node whose output depends on θ. $\partial E/\partial\theta$ can be written as

$$\frac{\partial E}{\partial \theta} = \sum_{q=1}^{q} \frac{\partial E_q}{\partial \theta} \tag{5}$$

The update formula for θ can be written as

$$\Delta\theta = -\gamma \frac{\partial E}{\partial \theta} \tag{6}$$

in which γ is the learning rate which can be defined as

$$\gamma = \frac{j}{\sqrt{\sum_\theta (\partial E/\partial\theta)^2}} \tag{7}$$

in which j is the step size. The change in j results in the convergence speed.

5. Modeling and Results

In this paper, 2,700 gauge observations were used to train the system. The input data are gauge values for year_{s-2}, year_{s-1}, and MGT which are the most important factors in the one step ahead prediction for gauge values, and the output of the system is the gauge values for year_s. The trained system antecedent membership functions are presented in Figure 4.

In this article firstly the datasets are randomly divided into training and testing datasets; following that ANFIS model is utilized using 5 membership functions for trio inputs. The number of membership functions and their shape are selected to have the Least Mean Squared Error (MSE) and higher R^2 in training datasets. These training datasets are then trained using ANFIS algorithm. Ultimately, the trained system is then tested on the test data and the observed gauge values and the predicted ones are compared as shown in Figure 5.

As shown in Figure 5, straight sections are responsible for more frequent and violent fluctuations in comparison to

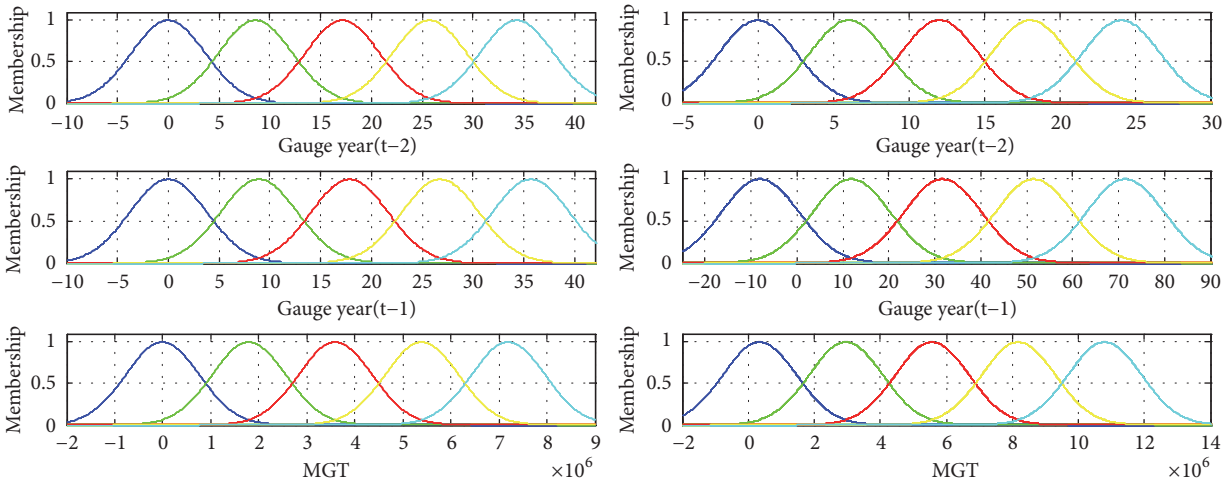
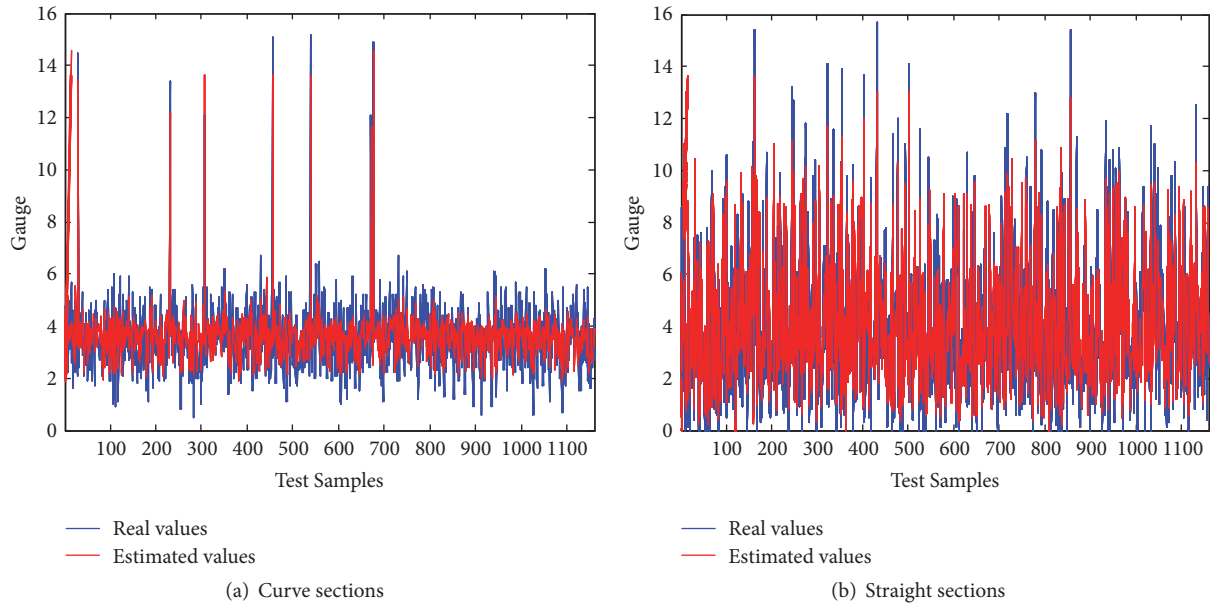


FIGURE 4: Membership function of the antecedents, i.e., the gauge values for year_{s-2} and year_{s-1} and MGT for curves and straight sections shown from left to right, respectively.



(a) Curve sections (b) Straight sections

FIGURE 5: The observed and predicted age values for the test data (30% of the data).

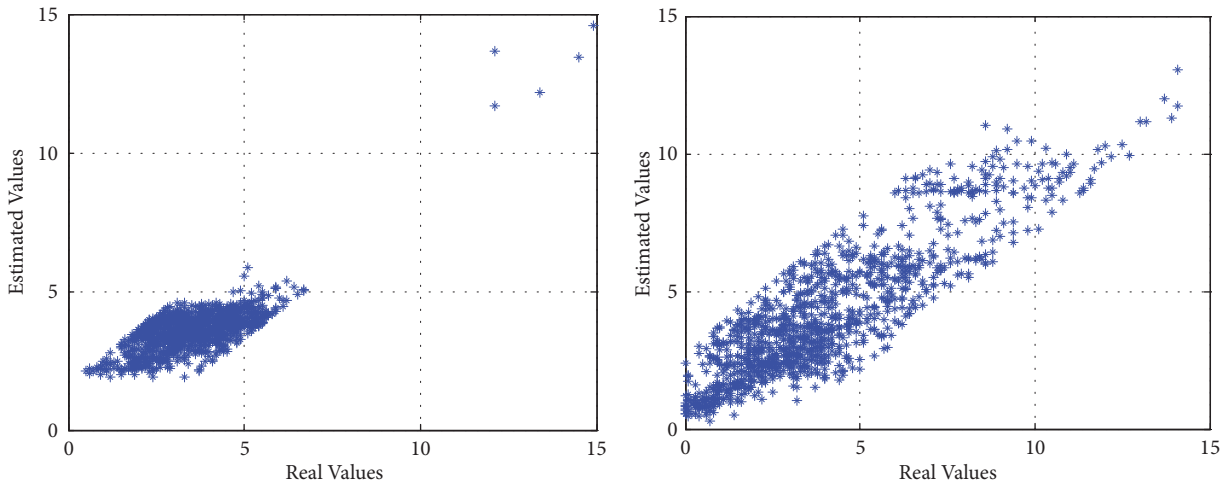
curves. To unveil the key reason, it should be considered that switches and bridges which are more in danger of degradation are mostly quantified as straight sections. The spikes in both graphs reach a height of 14 mm in some samples. This occurrence of less frequent spikes could be justified by stating that it could be the result of noisy data in the curve sections. However, it would not be a valid argument when it comes to explaining the data of the straight section since the spikes in that graph are more frequent. This high decay in the gauge of the straight sections could be due to few reasons. It could be due to the subclasses of the straight sections eventually categorized as the straight sections. These sections tend to suffer greater deterioration in gauge than the other sections of the track. As a result, it will increase the overall gauge deterioration in the straight sections. To get more insight into the nature of the data in different sections Table 1 is presented.

TABLE 1: Mean and standard deviation for curves and straights. Real data is abbreviated as rd and estimated data is abbreviated as ed.

	Curves	Straights
σ_{rd}	1.3579	2.8207
σ_{ed}	0.9766	2.4594
μ_{rd}	3.5320	3.9185
μ_{ed}	3.6160	4.0932

Table 1 shows the mean and standard deviation of the two different sections.

Table 1 indicates the mean and the standard deviation for the observed and predicted gauge values at curves and straights, and it shows a massive difference in standard deviations between curve data and straight data in both



(a) Curve sections (b) Straight sections

Figure 6: The observed versus predicted gauge values.

observed and predicted values. The numbers demonstrate the nature of curves and straight sections. For instance, the values for the standard deviation in curves and straights which are 1.36 and 2.82, respectively, demonstrate that the deviation in the values of gauge value is higher in the straight section. Also, their mean values which are 3.53 and 3.92 for curves and straights demonstrate that the straight sections face higher gauge values. From both mean and standard deviation, it could be understood that higher gauge values are more common in the straight section which means this section type needs more common inspections due to the fact that straight section covers most frequent stop-starts and some other factors such as bridges which are enclosed by a straight section.

To show the accuracy of the model, observed gauge values versus the predicted gauge values are plotted in Figure 6. As demonstrated in this figure, the curve data may contain few outliers which could be seen in the top right corner of the curve graph. Through the graphical presentation provided in Figure 6, it is visible that the model has predicted those outliers accurately, given that they appear quite close to the regression line.

Moreover, Figure 6 depicts that the concentration of the gauge values lies within 0 to 5 in both graphs while on the straight sections there is a considerable spread in gauge values from 5 to 10. This indicates a higher probability of degradation on the straight sections over curves.

R^2 which indicates the proportionate amount of variation in the response variable y explained by the independent variables X in the linear regression model is presented in (8). R^2 of the curve degradation prediction model is 0.60 while that of the straight model is around 0.78. By considering the above figures and the values of R^2, it is clear that the system can predict the gauge values with high accuracy considering the noisy nature of the data.

$$R^2 = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}, \tag{8}$$

Table 2: Statistical parameters of model.

Criteria	Curves	Straights
R square (R^2)	0.6001	0.7808
MSE	0.7350	1.7335
Total Samples	3860	3860
Training Samples	2700	2700
Testing Samples	1160	1160
Number of inputs	3	3

where SSE is the sum of squared errors; SSR is the sum of squared regressions; SST is the squared total sum.

As Table 2 indicates, 3,860 observations have been used to train and test ANFIS models for predicting the gauge values for curve and straight sections. Despite the fact that the model for straight sections shows a better R^2 value, the MSE also has gone up from curve sections to straight sections. This could be because straight section data is spread more than the curve data as shown in Figure 6.

6. Conclusions

Predicting rail track degradation on time and carrying out maintenance accordingly are a plan that all the major authorities that are responsible for maintaining and managing the rail networks try to implement right across the world. The intention behind this exercise is to improve the cost efficiency by reducing the unnecessary costs related to carrying out maintenance work too early or too late. There are many types of research that have been done on developing such models for heavy rail tracks. However, there is a need for a proper model which can predict the degradation in light rail tracks. In this paper, firstly the raw data was captured by connecting the MATLAB to the database, then it is preprocessed, and the outliers are eliminated due to the messy structure of the data environment. Following that to cover all noisy environment ANFIS model was utilized; the model has five membership

functions in the antecedent for all three inputs and is put forward to model rail track degrading using the data for Melbourne tram network between 2010 and 2015. The data is consisting of gauge values for two previous years and the MGT value. For modeling the system 70% of the data is used for the training purpose, and the remaining was used for testing the system. Results show that the model can predict the gauge values for the coming year by the R^2 value of 0.60 and 0.78 for the curves and straight sections, respectively, and the MSE error for the section are 0.7350 and 1.7335, respectively. These models could be utilized to predict the degradation for the year 2016 onwards, and maintenance work could be scheduled according to the results from degradation prediction models.

Conflicts of Interest

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

Acknowledgments

The authors would like to acknowledge the support and collaboration of Yarra Trams for providing this research with the required data.

References

[1] Y. Sato, "Japanese Studies on Deterioration of Ballasted Track," *Vehicle System Dynamics*, vol. 24, no. 1, pp. 197–208, 1995.

[2] Y. Satoh, "Experiment on Ballast Settlement due to Repeated Loading," Railway Technical Research Report, 1959.

[3] Y. Satoh, A. Onishi, and S. Tanaka, "Experiment on grading of crushed stone ballast," Railway Technical Research Report, 1961.

[4] K. Demharter, *Setzungsverhalten des Gleisrostes unter vertikaler Lasteinwirkung*, 1982.

[5] A. Premathilaka, S. Costello, and R. Dunn, "Development of a deterministic rail-wear prediction model," *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice*, vol. 19, p. 40, 2010.

[6] A. R. Andrade and P. F. Teixeira, "Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models," *Reliability Engineering & System Safety*, vol. 142, article no. 5318, pp. 169–183, 2015.

[7] R. G. Mishalani and S. M. Madanat, "Computation of infrastructure transition probabilities using stochastic duration models," *Journal of Infrastructure Systems*, vol. 8, no. 4, pp. 139–148, 2002.

[8] M. N. S. Hadi, "Neural networks applications in concrete structures," *Computers & Structures*, vol. 81, no. 6, pp. 373–381, 2003.

[9] J. Sadeghi and H. Askarinejad, "Application of neural networks in evaluation of railway track quality condition," *Journal of Mechanical Science and Technology*, vol. 26, no. 1, pp. 113–122, 2012.

[10] H. Guler, "Prediction of railway track geometry deterioration using artificial neural networks: A case study for Turkish state railways," *Structure and Infrastructure Engineering*, vol. 10, no. 5, pp. 1–13, 2014.

[11] Y. Shafahi, P. Masoudi, and R. Hakhamaneshi, "Track Degradation Prediction Models, Using Markov Chain," in *Proceedings of the 8th World Congress on Railway Research*, Artificial Neural and Neuro-Fuzzy Network, Seoul, Korea, 2008.

[12] S. Moridpour and R. Hesami, "Degradation and performance specification of Melbourne tram tracks," in *Proceedings of the 3rd International Conference on Transportation Information and Safety, ICTIS 2015*, pp. 270–276, chn, June 2015.

[13] M. Yousefikia, S. Moridpour, S. Setunge, and E. Mazloumi, "Modeling degradation of tracks for maintenance planning on a tram line," *Journal of Traffic and Transportation Engineering*, vol. 2, 2014.

[14] S. Moridpour, E. Mazloumi, and R. Hesami, "Application of artificial neural networks in predicting the degradation of tram tracks using maintenance data," *Applied Big Data Analytics in Operations Management*, 2017, IGI Global.

[15] M. Karimpour, L. Hitihamillage, N. Elkhoury, S. Moridpour, and R. Hesami, "Nonlinear estimation model for rail track deterioration," in *19th International Conference on Transportation Economics and Transportation Systems*, 2017.

[16] J. S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 6701484, 10 pages
https://doi.org/10.1155/2018/6701484

Research Article

Efficiency Assessment of Transit-Oriented Development by Data Envelopment Analysis: Case Study on the Den-en Toshi Line in Japan

Jing Guo[1], **Fumihiko Nakamura**[2], **Qiang Li**[3] and **Yuan Zhou**[4]

[1]*Graduate School of Environmental Studies, Nagoya University, D2-1Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan*
[2]*Institute of Urban Innovation, Yokohama National University, 79-5 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan*
[3]*College of Resources Science and Technology, Beijing Normal University, 19 Xinjiekouwai Street, Beijing 100875, China*
[4]*Department of Geographical Sciences, University of Maryland-College Park, College Park, MD 20742, USA*

Correspondence should be addressed to Qiang Li; liqiang@bnu.edu.cn

Received 20 November 2017; Revised 28 March 2018; Accepted 15 April 2018; Published 15 May 2018

Academic Editor: Taku Fujiyama

Transit-Oriented Development (TOD) is an urban planning approach that encourages a modal shift from private to public transportation. This shift can generate additional benefits from a sustainability perspective. This study aims to assess the efficiency of TOD by applying the data envelopment analysis (DEA) method. The ridership of public transportation is considered as the direct output characteristic of TOD efficiency, and nine indicators of ridership are selected as inputs on the basis of the core concepts of TOD. These concepts include density, diversity, and design (3Ds). The Tokyu Den-en Toshi Line in Japan is presented as a typical case of TOD because this line includes TOD and non-TOD stations. Assessing and comparing the results of all railway stations reveal that almost all indicator values of non-TOD stations are higher than those of TOD stations. The results suggest that TOD planning and programs are inefficient in terms of ridership generation. This implication, however, may be attributed to the inadequacy of the selected indicators for TOD assessment. The results obtained after adding operation year as input indicators and removing transfer station show that TOD stations perform efficiently and in accordance with expectations. This response indicates that the inclusion of influential factors is necessary for equitable TOD assessment. Therefore, other influential factors must be considered when evaluating the efficiency of TOD-based stations with different inherent attributes. In addition, the design input with the largest impact on all inefficient units was identified, suggesting that management of bus service and railway system should be well enhanced.

1. Introduction

Transportation infrastructure is one of the most important components of urban development. The efficient development and implementation of a transportation infrastructure plan are necessary to solve traffic congestion, greenhouse gas emissions, and other environmental consequences that accompany rapid urbanization and motorization [1]. From the perspective of sustainable and integrated urban development, the imbalance between land use and transportation systems is the most critical factor that affects the efficiency of transportation infrastructure. Transit-Oriented Development (TOD) strategy has been proposed as an efficient approach to solve this imbalance [2, 3].

TOD aims to encourage the use of public transportation [4] by increasing population density and mixed land use in areas surrounding transportation hubs and by promoting connectivity between stations and trip origins/destinations. Although the formal term of TOD was not introduced until 1993 [5], its underlying principle has been applied and implemented as a solution to urban growth challenges since the late 19th and early 20th centuries, mainly in the United States and Europe [6]. Later, in 1997, the density, diversity, and design (3D) framework of TOD was proposed and explained [7]. In several American cities, the 3D-based TOD plan has increased the ridership and modal shares of public transport and has decreased vehicle mileage by 3%–5% [8, 9]. Accordingly, the TOD concept has been

applied with an emphasis on the 3D framework in urban and transport planning worldwide. However, its efficiency in practical applications may sometimes not meet expectations [10]. The low utilization of public transport cannot achieve the sustainable cycle of land–people–transport [11]. In other words, whether or not the TOD strategy can play its due role in urban planning and management is inconclusive. Therefore, evaluating the efficiency of TOD is a necessary challenge faced by urban administrators and policy makers who aim to successfully implement, enhance, and improve TOD.

To respond to such a demand, this study selects the appropriate indicators that reflect the 3D concept of TOD in practice. This study assesses the efficiency of a railway system that was developed on the basis of the 3D framework of TOD. Furthermore, it provides advice to improve the factors that most significantly influence the efficiency of TOD. In this study, data envelopment analysis (DEA), which was first proposed by Charnes et al. in 1978, is applied [12, 13]. Since its introduction, DEA has been widely used to measure the relative efficiency of various organizations, referred to as decision-making units (DMU), with multiple inputs and outputs [14–17]. DEA can identify the input with the largest impact on an inefficient unit. Such input can provide guidelines to policy makers. One line of the Japanese railway system is presented as a case study. In Japan, the integration of the transport system and land use has been practiced since the 1910s. This practice is consistent with the concept of TOD. After more than a century, the country has recognized advancements in railway operation and management. The efficiency evaluation of TOD cases in Japan can provide essential guidance or experiences for future transport and urban planning in other countries, especially in developing countries experiencing rapid urbanization.

2. Literature Review

The integration of transportation and land use dates back to the post-World War II era, when planners in Europe, most notably in Stockholm [18] and Copenhagen [19], led suburban development into satellite areas along transit corridors. Since the 1990s, a series of new concepts and approaches, such as the Smart Growth concept [20, 21] and New Urbanism approach [22], for the integration of transportation and land use have emerged to combat uncontrolled urban sprawl [6]. TOD is also gaining popularity as an approach to promote smart growth and sustainable development in the metropolitan areas of many developed and developing countries [23–26] where mass transit systems have been implemented to relieve traffic congestion.

A common definition or criterion remains to be established for TOD despite the substantial body of research that has discussed the concept. For example, Cervero et al. stressed the contributions of the 3Ds to the TOD concept; furthermore, they explained the 3D concept as density in the form of residence and jobs, diversity in the form of mixed land-use development, and design in the form of good street connectivity for pedestrians [25]. Other scholars interpreted TOD by borrowing words from its original explanation and providing new empirical additions. The various definitions for TOD generally state that TOD should promote compact and highly mixed-use development around transit hubs and should include accessible and walkable neighborhoods [18, 19].

Existing studies have demonstrated that TOD provides numerous advantages, such as facilitating polycentric cities and regions, mitigating urban sprawl, boosting public transport ridership, increasing bike usage and walkability, accommodating economic growth, and creating sustainable neighborhoods [6]. The most important function of TOD is to reduce traffic congestion and increase public transit use in areas experiencing accelerated urban development [25, 27, 28]. TOD has attracted increasing attention as a solution to the challenges of urban growth given its widely recognized benefits. TOD implementation, however, is not always successful. Thus, several studies have evaluated the efficiency of TOD (e.g., [9, 21]). Most existing research applied single and several indicators to measure the outcomes or benefits that TOD provides in terms of travel behavior, economics, and environment [28, 29]. Ridership is one of most representative indicators of TOD outcome [25]. Renne and Wells monitored the success of TOD through a survey that included and summarized the various indicators of the influences and the most useful indicators of TOD [3]. However, this study quantified TOD in the form of a single indicator and provided a limited explanation for the ridership efficiency of TOD. In addition, it lost meaningful information for the aspects that contribute to the success or failure of TOD. The multi-indicator evaluation method involves weight setting for each indicator. However, weight setting in existing studies is often empirical or less objective and could provide biased results.

In the past decade, DEA has been widely used to measure the efficiency of different models of transportation system operations or management [30]. The most significant advantage of DEA over traditional weight-based methods is that DEA does not require weight parameters to measure efficiency. Researchers have employed DEA to address two main research topics. One topic is the analysis and comparison of the efficiency of firms, networks, or nations in the field of economics and management science [31–33]. Only a small part of related works has focused on the evaluation of the operational performance and management of railways [30]. The other topic is the selection of the optimal solution for actual operations in vehicle dispatching and route planning [34, 35]. Up to now, DEA has not been applied in the assessment of transportation strategy, at least in terms of TOD.

2.1. DEA. As mentioned above, DEA identifies the frontier (envelopment surface), which is determined directly from data, and quantifies the relative efficiency of a set of comparable DMUs with several inputs and outputs. It solves a linear programming model to detect the efficient inputs and outputs that form the frontier and the inefficient inputs and outputs that form nonfrontier units. In addition to providing the efficiency level, DEA can identify the sources of inefficiency and the projection path to the frontier. The projection path to the envelope surface is determined by whether the model

Table 1: Indicators for DEA method.

		Indicators	Unit	Explanation
Density	x_1	Population density	Person/km^2	Population per square kilometer
Diversity	x_2	Land-use diversity	-	Shannon–Wiener Index
	x_3	Functional land area	km^2	Total residential, commercial, official, public service land area per catchment
	x_4	High-building area	km^2	Area with high buildings (over 4 floors) per catchment
Design	x_5	Bus stops	-	Bus stops per unit catchment area
	x_6	Daily bus operation length	km	Bus operation length per day in each catchment
	x_7	Bus routes	-	Number of routes to and from railway station
	x_8	Road length	km	Road length per catchment
	x_9	Number of station exits	-	Number of exits per railway station
	y	Ridership	Person	Average daily commuters per station

is output-oriented or input-oriented, the choice of which depends on the production process that characterizes the units (i.e., the process minimizes the use of inputs to produce a given output level or maximizes output level given an input level). A large output (i.e., ridership) is preferred in the evaluation of TOD efficiency given several certain inputs. Therefore, output-oriented DEA was selected in this study.

The CCR model, which was first proposed by Charnes et al. in 1978, is the basic DEA model. It functions under the assumption of constant returns to scale and reflects the fact that outputs change proportionally with inputs [12]. Assume that n DMUs ($j = 1, \ldots, n$), m inputs x_{ij} ($i = 1, \ldots, m$), and s outputs y_{rj} ($r = 1, \ldots, s$) exist for each DMU$_j$. The output-oriented CCR model can be formulated as follows:

$$\begin{aligned} \max \quad & \theta + \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \\ \text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0} \\ & \sum_{j=1}^{n} \lambda_j y_{rj} - s_i^+ = \theta y_{r0} \\ & s_i^-, s_r^+ \geq 0 \\ & \lambda_j \geq 0, \end{aligned} \tag{1}$$

where θ is the efficiency score limited to less than 1, ε is the non-Archimedean infinitesimal, and x_{i0} and y_{r0} refer to the input and output of one certain unit DMU$_0$. s_i^- and s_r^+ are the input and output slack, which refers to the excess input or missing output that exists after the proportional change in the input or the output has reached the efficiency frontier. λ_j is the weight given to the DMU$_j$ to achieve its optimal performance. It is calculated automatically during the model-solving process.

The improvement target (projection) of each inefficient unit is the result of respective slack values added to proportional reduction amounts (formula (2) and (3) for the output-oriented DEA model). The improvement target provides alternative ways to improve the performance of each inefficient unit. The difference between x_{i0} and x'_{i0}, y_{r0} and y'_{r0} should receive focus when changing the input and output.

$$x'_{i0} = x_{i0} - s_i^- \tag{2}$$

$$y'_{r0} = \theta y_{r0} + s_r^+. \tag{3}$$

2.2. Indicator Selection. As the output variable, ridership (y) is the number of passengers who ride a public transport system and directly reflects the efficiency of transportation policy and urban planning, such as TOD strategy. Populations living around transport hubs provide the trip demand and have the potential to use the transportation infrastructure; mixed land use in the catchment area helps generate trips by providing different services; and convenient design encourages high numbers of people to use public transportation. All of these effects promote modal shift and increase ridership. Therefore, from the core concept of TOD, that is, density, diversity, and design, nine indicators were selected as the input variables of ridership output (Table 1).

First, density is defined as the number of buildings or population per area. Here, population density (x_1) in a catchment, that is, population per unit area, is the primary indicator that directly determines trip demand and economic activity. The residential population tends to generate the boarding ridership, whereas the commercial and office population is related to the alighting ridership via commercial or employment attractiveness.

Second, diversity is reflected by land use and primarily represents the level of mixed land-use development. This indicator is obtained through the Shannon–Wiener index, which is widely applied to calculate species diversity in the field of ecology. It has been successfully applied in land-use research to estimate the number of land-use types and evenness [16]. Given that residential land, commercial land, official land, and public service land are the primary types of urban functional land areas, the functional land area will affect ridership by providing for the different needs or demands of passengers and other interested people. Furthermore, areas with high buildings (over 4 floors) surrounding transportation hubs affect trip demand generation and represent the basic characteristic of regional development level. Therefore, land-use diversity (x_2), functional land-use area

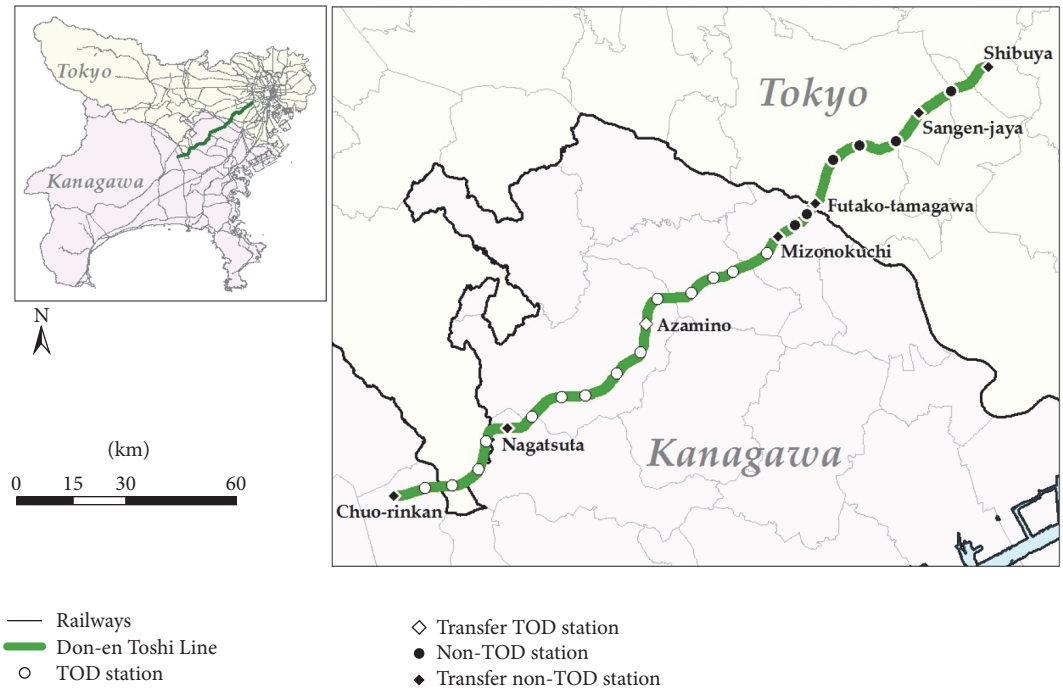


Figure 1: Tokyu Den-en Toshi Line.

(x_3), and high-building area (x_4) were selected as the three indicators of diversity input.

Third, design schemes can influence travel demand by increasing the accessibility of destinations by foot or by other public transportation systems [6]. It contains different implications from the micro to macro level, including the presence of shade trees and overhead lighting, design of on-site stores and services, and connections between work sites and worker residences. In this study, the design points to a basic transit infrastructure around rail stations that shortens the distance or time required to reach a destination. A highly advanced design would have the potential to facilitate public transit use and walking, as well as promote efficient energy use. Therefore, the number of bus stops (x_5), daily bus operation length (x_6), and bus routes (x_7) were selected to represent the connectivity between the rail station and passenger destinations. Connectivity eventually affects ridership choice and mode split. With respect to the total road length (x_8), we hypothesized that long roads in a certain area would increase road density and facilitate passenger access to railway stations. Finally, the number of railway station exits (x_9) also provides passengers with convenient access to aboveground roads or buses and underground stores or supermarkets.

3. Case Study

3.1. Tokyu Den-en Toshi Line. Japan's urban transportation system is famous for its excellent accessibility, efficiency, and punctuality. These characteristics resulted from advanced and reasonable transportation/urban planning. Moreover, in Japan, cities are commonly planned on the basis of TOD because the country has centered its land and city development along railroad construction for more than a century. *Ensenkaihatsu* (or "railway corridor development") was proposed by as early as 1910 and has a longer history than TOD. This urban development concept focused on residential district development along railways.

The Tokyu Den-en Toshi (DT) Line is a major commuter line that connects Shibuya in Tokyo and the Chuo-Rinkanin terminal in the neighboring Kanagawa Prefecture (Figure 1). The earliest interval, which is known as the Mizonokuchi Line, was opened in 1927 and connects the Futako-Tamagawa and Mizonokuchi Stations. In the mid-20th century, economic activity and employment rapidly concentrated in the central area of Tokyo, thus drastically increasing land and real estate prices and promoting ex-urban settlement. In response to this situation, the Japanese government adopted a new TOD plan. The DT Line was born at the right moment, and Tama New Town was subsequently developed. In 1966, the DT Line began operating from the Mizonokuchi Station to the Nagatsuta Station. An expansion was built as an access route to Tama Garden City [36]. The DT Line, which spans Shibuya Station to Futako-Tamagawa Station, was completed in 1977. In 1984, the line was extended from Nagatsuta Station to Chuo-Rinkan Station. Stations between the Mizonokuchi and Chuo-Rinkan Stations, except for Nagatsuta Station (Figure 1), were developed on the basis of TOD. To enable comparison in this study, the earlier/older stations are identified as non-TODs (12 stations), and the latter/newer stations are considered as TODs (15 stations).

The DT Line was selected as a case study in this work for two reasons. First, the DT Line and the Tama New Town

that it traverses provide examples of transportation and land-use integration that are intended to stimulate population concentration and ridership growth. Second, this line has a relatively long history and was constructed under previous and new urban development policies (i.e., some of its stations were originally built with the birth of the line, whereas others are relatively new and were constructed through the TOD project). Therefore, earlier and later/newer stations can be compared to identify the factors that affect ridership.

3.2. Catchment Definition and Data Collection. Station catchment areas, which influence indicator values, need to be determined. In the transportation research field, the catchment area where local residents potentially use the railway is identified through (1) placing a buffer circle around a station (e.g., buffers with a radius of 400, 800, or 1200 m to define walkability to the public transportation facility) [37]; (2) network analysis [38] based on path networks, wherein a catchment area can be an area with a fixed walking or driving distance or travel time; (3) the boundary method, which is based on geo-code trip survey data (in this method, however, only 90% of the trip data are used within the study to remove the outliers) [39]; and (4) statistical modeling based on spatial theory, such as location and allocation method [40]. Among these methods, the boundary method is the most objective method that does not require intensive expert experience.

In this study, the Japanese Nationwide Person Trip Survey was used to identify station catchments through the boundary method. This survey is the most fundamental transportation questionnaire survey and is conducted every 10 years by the Ministry of Land, Infrastructure, Transport, and Tourism to investigate the actual travel behaviors of people in certain regions in Japan. In the Tokyo Metropolitan Area, approximately 730,000 individual questionnaires are collected. Survey items include individual trips, personal attributes, and travel behaviors. The origin collected for the survey was recorded by a small zone geo-code, and the survey results showed the station choices in each zone.

An origin–destination matrix from the zones of passenger residence (origin) to rail stations (destination) was constructed in accordance with the survey results and is shown in the flow map in Figure 2(a). Then, zones that generated few demands were removed on the basis of the rule that a catchment area should cover at least 90% of the ridership around the station (Figure 2(b)). A single catchment area nearly covers the buffer area with a 1,500 m radius around the railway. Given that individuals from the same small zone may select different rail stations, the catchment areas of two or more adjacent rail stations may overlap. Overlapped zones were divided into different stations by the number of station choices and distances to different stations (Figure 2(c)). Finally, the catchment area for each rail station was identified. Following the identification of catchment boundaries, all data for selected indicators (e.g., bus stop data, as an example shown in Figure 2(d)) were collected from the website of the Official Statistics of Japan and National Land Numerical Information. The collected data were then masked, reorganized, and standardized.

4. Results

4.1. Basic Statistical Characteristics. The initial analysis provided the basic statistical characteristics of all selected indicators for the 27 stations. All of the indicator values of non-TODs are higher than those of TOD stations. Therefore, TOD stations do not provide any obvious advantages. In detail, the mean population density of TOD stations is considerably lower than that of older stations. This result may be attributed to the different development histories of non-TOD and TOD stations. The higher land-use diversity of non-TODs than that of TOD stations can be attributed to the main function of the Tama New Town Station along the DT Line. The catchment of the Tama New Town Station mainly functions to relieve housing problems in Tokyo. Thus, this area contains vast tracts of residential land. By contrast, the catchment areas of older stations contain relatively more commercial and office areas than those of the Tama New Town Station. Non-TOD stations have higher functional land areas and numbers of high buildings per unit area than TOD stations, indicating that land use around old stations is better developed than that around new stations. TOD stations consistently have lower values for the five design indicators than non-TOD stations. This result suggested that infrastructure with less accessibility may not serve a high proportion of riders. Compared with non-TOD stations, TOD stations have lower average daily ridership, which is the direct output of transportation operation and management and land use. However, low ridership and indicator values do not necessarily indicate low efficiency and vice versa. Thus, an in-depth assessment is necessary to identify the efficiency of TOD stations.

4.2. TOD Efficiency. To explain whether different 3D patterns can generate correspondingly different ridership patterns, the DEA method was applied to calculate the efficiency of each station pattern. The results are shown in Table 3. Nine stations with E equal to 1 are considered as relatively efficient. However, seven of these stations are non-TOD stations, and only two are TOD stations. The mean efficiency score of TOD stations is lower than that of non-TOD stations, indicating that the 3D-based TOD strategy does not generate sufficient ridership as initially expected.

This finding can be partly explained by the individual attributes of each station. Station attributes include designation as a transfer station, opening date, and operating history. For example, the Shibuya Station is a transfer station with a long operating history in a catchment area with high population density. Accordingly, its corresponding land use is well developed, and its accessibility attracts high numbers of riders. Furthermore, non-TOD stations may achieve TOD goals given their specific attributes. For example, among non-TOD stations, the Azamino Station is highly efficient despite its lack of a long operating history because of its transfer function. The above result suggests that considering only 3D-related indicators cannot provide an adequately comprehensive assessment of TOD efficiency. The inherent attributes of each station cannot be ignored because they constrain the comparability of stations.

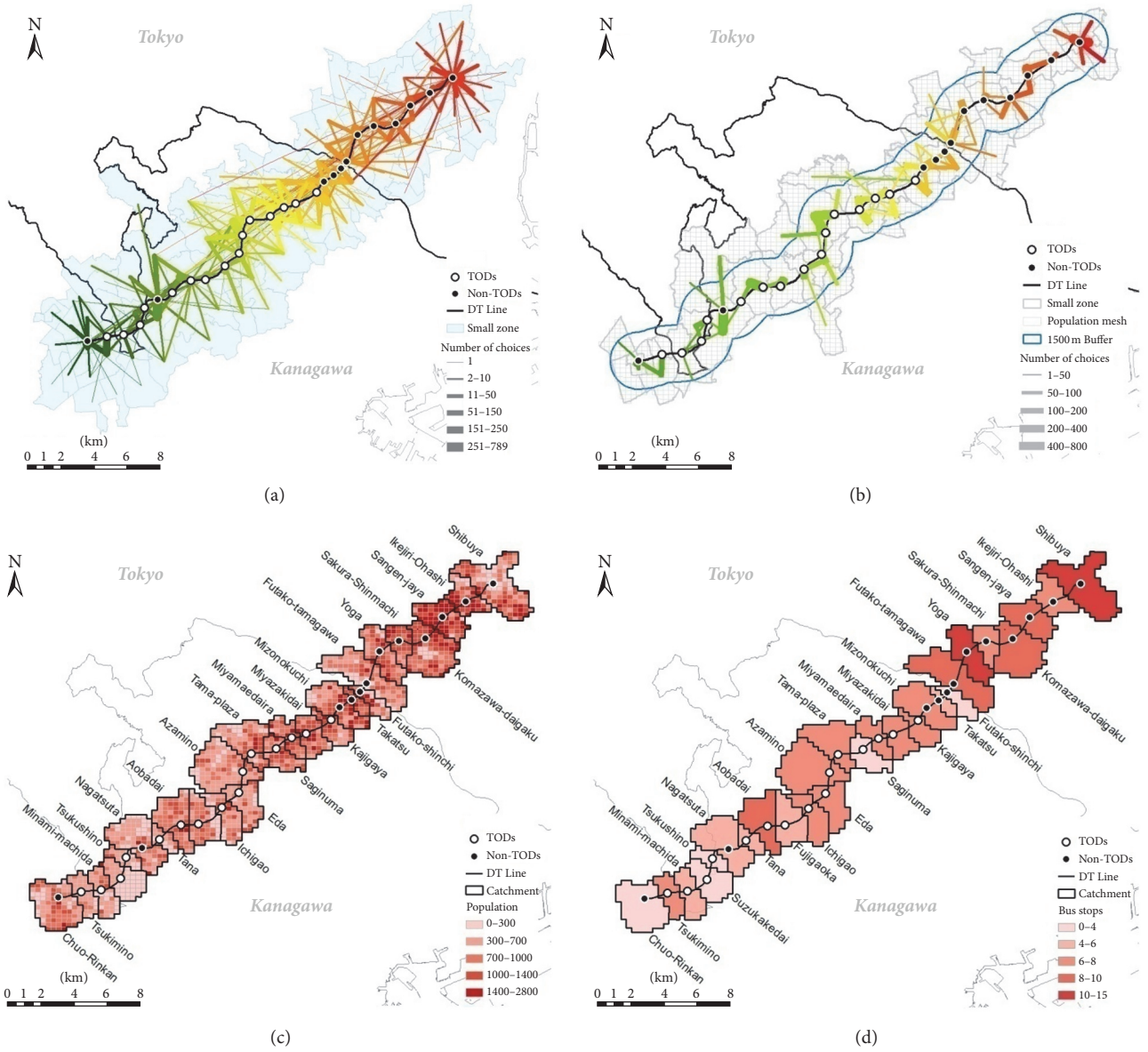


Figure 2: Generation of catchment area and distribution of a selected indicator for each catchment. (a) Original flow map; (b) deletion of extra zones and magnification; (c) finished catchment area; (d) bus stops for each catchment. *Note.* In (a) and (b), the flow color is used to distinguish the populations of different stations. The width of gray bar in legend represents the number of choices of separate stations.

4.3. TOD Efficiency considering Inherent Station Attributes. The inherent attributes of each station may affect its ridership proportion and efficiency. Inherent station attributes include the number of operation years, designation as a transfer station, number of trains running per day, and duration between trips. The former two attributes have been discussed above to account for the relatively inefficient performance of TOD stations. To avoid the biases resulting from operating history and transfer function, operation year was added to the DEA model. At the same time, seven transfer stations were removed from the DMU list to improve the comparability of stations. Hence, the DEA model was recalculated with 20 stations (6 non-TOD and 14 TOD stations) with 10 input indicators, including operation year.

After removing transfer station as an input (Table 4), the differences between the input indicator values of TOD and non-TOD stations have drastically decreased compared with those shown in Table 2. As shown in Table 2, the average ridership of non-TOD stations is almost three times larger than that of TOD stations. However, the average ridership of non-TODs is approximately 20% higher than that of TOD stations when only nontransfer stations are considered. Given that ridership is the output indicator of the DEA model, changes in ridership will inevitably lead to the different efficiency results shown in Table 5. Most TOD stations are relatively efficient, and the mean values of efficiency E for TODs and non-TODs are not considerably different. Yoga Station, a non-TOD station, performed relatively inefficiently

Table 2: Means of the indicators of non-TOD and TOD stations.

	Indicators	Non-TODs	TODs	Difference
Density	Population density	14.21	10.39	3.82
Diversity	Land-use diversity	0.97	0.64	0.33
	Functional land area	0.87	0.86	0.01
	High-building area	0.06	0.03	0.03
Design	Number of bus stops	7.37	6.05	1.32
	Daily bus operation length	575.44	423.51	151.93
	Bus routes	4.50	1.73	2.77
	Road length	19.36	21.37	−2.01
	Number of station exits	12.08	7.87	4.21
Daily ridership		130459	46747	83712

Table 3: Efficiency of each station.

Non-TOD stations			TOD stations		
Station	E	Operation years	Station	E	Operation years
			Kajigaya	0.824	51
Shibuya*	1	132	Miyazakidai	0.536	51
Ikejiri-Ohashi	0.364	40	Miyamaedaira	0.496	51
Sangen-Jaya*	0.739	110	Saginuma	1	51
Komazawa-Daigaku	0.760	40	Tama-Plaza	0.644	51
Sakura-Shimmachi	0.730	110	Azamino*	1	40
Yoga	0.487	110	Eda	0.895	51
Futako-Tamagawa*	1	110	Ichigao	0.507	51
Futako-Shinchi	1	90	Fujigaoka	0.513	51
Takatsu	1	90	Aobadai	0.905	51
Mizonokuchi*	1	90	Tana	0.809	51
Nagatsuta*	1	109	Tsukushino	0.286	49
Chuo-Rinkan*	1	88	Suzukakedai	0.274	45
			Minami-Machida	0.573	41
			Tsukimino	0.257	41
Mean	0.840	93.25	Mean	0.635	48.4

Note. ∗ denotes a transfer station.

Table 4: Means of indicators of nontransfer stations.

	Indicators	Non-TODs	TODs	Difference
Density	Population density	15.74	10.48	5.26
Diversity	Land-use diversity	0.93	0.64	0.29
	Functional land area	0.87	0.86	0.01
	High-building area	0.03	0.02	0.01
Design	Number of bus stops	7.02	6.01	1.01
	Daily bus operation length	542.05	430.90	111.15
	Bus routes	2.83	1.71	1.12
	Road length	22.99	21.71	1.28
	Number of station exits	10.67	7.29	3.38
Operation year		80	49	31
Daily ridership		52478	40566	11912

given its 110-year operation history. The efficiency values of 12 TOD stations with 40–50 years of operation history, except for Ichigao and Tana Stations, are all higher than those of Yoga Station. This result suggested that TOD stations tend to reach their current efficiency more rapidly than non-TOD stations.

4.4. Potential for Improving Efficiency. The frontier project reflects the improvement target that the inputs and outputs of each inefficient station should reach. The distance (difference) between the real point and its projection indicates the volume required for improvement. Then, the percentage

Table 5: Efficiencies of nontransfer stations considering operation year.

Non-TOD stations		TOD stations	
Station	E	Station	E
		Kajigaya	1
		Miyazakidai	1
		Miyamaedaira	1
		Saginuma	1
Ikejiri-Ohashi	0.702	Tama-Plaza	1
Komazawa-Daigaku	1	Eda	1
Sakura-Shimmachi	1	Ichigao	0.708
Yoga	0.881	Fujigaoka	0.958
Futako-Shinchi	1	Aobadai	1
Takatsu	1	Tana	0.809
		Tsukushino	1
		Suzukakedai	0.894
		Minami-Machida	0.928
		Tsukimino	0.999
Mean	0.981	Mean	0.955

Table 6: Difference between the average projected and real values of inefficient stations (%).

	Indicators	Non-TODs	TODs
Density	Population density	6.33	17.89
Diversity	Land-use diversity	20.39	20.14
	Functional land area	6.04	8.67
	High building area	9.46	27.89
Design	Number of bus stops	28.68	26.32
	Daily bus operation length	38.31	46.32
	Bus routes	36.38	28.08
	Road length	19.27	23.17
	Number of station exits	6.34	0.62
Operation year		27.03	13.84

of difference relative to the original input can be calculated to discover the excess input indicator that contributes most significantly to inefficiency. The average difference percentage of each input for non-TOD stations and TOD stations (only including inefficient stations) is listed in Table 6.

All inefficient railway stations have higher numbers of redundant design indicators than the other two types of stations. The results indicated that bus location and services do not increase the corresponding ridership proportion of the railway. One possible reason is that passengers' daily transportation needs cannot be satisfied by the mutual independence of the railway and bus systems. Therefore, bus services and their feeder roles should be enhanced to improve the efficiency of the railway system. Moreover, diversity in the land use of areas surrounding inefficient stations does not generate adequate ridership, suggesting that the mixture of land-use types should be promoted despite requiring long-term effort. Population density is more redundant for inefficient TOD stations than for non-TOD stations. This result can be attributed to the number of serviced people and to age category (i.e., percentage of the elderly, adults, and children) and social level (i.e., percentage of unemployed, professionals, and managers) of the residents of a certain catchment area. For example, high-class workers and the elderly tend to use private transportation. Accordingly, increasing the density and optimizing the age and social category of the target population may help improve TOD efficiency. All aspects listed above should receive additional attention when planning future TOD stations.

5. Discussion and Conclusions

This study focused on measuring the efficiency of the TOD strategy implemented in the DT Line. TOD is a powerful urban planning approach and concept that is applied to resolve various urban problems. This study expounded on TOD assessment on the basis of the 3D framework. In this study, the assessment procedure included the identification of catchment areas and the selection of suitable indicators as input variables and ridership as output variables. The input variables are then subjected to the DEA method to validate the successful performance of the TOD strategy or program and to identify possible solutions to strengthen TOD implementation.

Existing TOD evaluations have failed to provide reliable conclusions regarding TOD performance. For example, a single-indicator assessment cannot adequately characterize the comprehensive efficiency level of a TOD strategy from the perspective of the 3D framework. Meanwhile, a simple weight-based method following the 3D concept is likely to be affected by weights with low objective values and may thus produce a biased result that may contradict reality. These situations suggest the necessity of designing a procedure for the comprehensive assessment of TOD strategies. By considering that density and diversity may generate transportation demand and that design must provide convenient service to satisfy such needs, we employed DEA to reflect the relative efficiency of TOD in accordance with 3D-based inputs and ridership output. Our DEA-derived results reflect the core principle of the TOD concept, which coordinates density, diversity, and design with ridership output.

Many TOD applications have demonstrated that TOD has a significant role in regional development. However, the present results suggested that TOD strategies cannot provide more positive effects than non-TOD strategies because, on the one hand, market power enables non-TOD strategies to achieve effects similar to those achieved by TOD strategies. More importantly, the current indicators selected for TOD assessment do not adequately reflect the 3D concept. In fact, the current indicators for TOD assessment only highlight aspects related to 3D, and some important factors that affect output are excluded from conventional TOD assessment. For example, operating history can also affect the output of TOD stations along the DT Line. Although TOD stations are comparable with non-TOD stations in general, development history results in the incomparable efficiency of new and old stations. Moreover, other factors, such as the age and social class of the catchment population, designation as transfer station, number of trains running per day, and duration

between trips, have been overlooked. Adding operation year as an input indicator and removing transfer station revealed that the performance of TOD stations is relatively efficient and approaches expectations. This result indicated that the inclusion of other influential factors is necessary for equitable TOD assessment. Consequently, taking such factors into account is indispensable when evaluating the efficiency of TOD stations with different inherent attributes. In addition, the design input with the largest impact on all inefficient units was identified. Management of bus service and railway system should be well enhanced and encouragement for public transportation usage should be emphasized.

The applicability of the DEA-based TOD assessment approach is mainly dependent on indicator selection. The framework of the approach includes indicator selection from the perspective of 3Ds, catchment divisions, and the DEA model and is applicable to other public transportation systems (e.g., light rail, metro, bus, and bus rapid transit) and to other regions or countries. Nevertheless, the selected indicators may vary across different applications. Researchers and urban planners may have specific local or specialized priorities and knowledge that differ from those emphasized in the present case. Thus, other researchers should select indicators that are appropriate for their priorities and knowledge.

Nevertheless, the present study presents several critical issues that should be considered to improve the evaluation of TOD approaches. First, additional factors that affect the efficiency of TOD-based systems should be considered during the assessment and planning of a TOD project. These factors include the attributes of the target population (age structure, occupation, and income level) and the transportation system itself. Second, the design scheme should improve the convenience of accessing destinations by foot, bus, or bicycle. In this study, the selected indicators mainly characterized connectivity to railway stations by bus. However, such a characterization may be insufficient for design characterization. Other factors, such as plating strips, overhead lights, and bicycle path lengths, should be considered in future studies. Furthermore, the definition of the analytical unit is crucial for TOD assessment. In this study, the Nationwide Person Trip Survey in Japan was used to identify station catchment areas through the boundary method. When such data are unavailable, the effect of catchment size should be addressed by specifying a different buffer radius, travel distance, or travel time.

Conflicts of Interest

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

References

[1] P. Miller, A. G. de Barros, L. Kattan, and S. C. Wirasinghe, "Public transportation and sustainability: A review," *KSCE Journal of Civil Engineering*, vol. 20, no. 3, pp. 1076–1083, 2016.

[2] R. Cervero, "TCRP Report 102: Transit-Oriented Development in the United States: Experiences. Challenges and Prospects," 2004.

[3] J. L. Renne and J. S. Wells, "Emerging European-style planning in the USA: Transit oriented development," *World Transport Policy & Practice*, vol. 10, no. 2, 2004.

[4] A. Nasri and L. Zhang, "The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas," *Transport Policy*, vol. 32, pp. 172–179, 2014.

[5] P. Calthorpe, *The Next American Metropolis: Ecology, Community, and the American Dream*, Princeton Architectural Press, New York, NY, USA, 1993.

[6] E. Papa and L. Bertolini, "Accessibility and Transit-Oriented Development in European metropolitan areas," *Journal of Transport Geography*, vol. 47, pp. 70–83, 2015.

[7] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: density, diversity, and design," *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp. 199–219, 1997.

[8] R. Ewing and R. Cervero, "Travel and the built environment: a synthesis," *Transportation Research Record*, vol. 1780, pp. 87–114, 2001.

[9] Y. J. Singh, P. Fard, M. Zuidgeest, M. Brussel, and M. V. Maarseveen, "Measuring transit oriented development: A spatial multi criteria assessment approach for the City Region Arnhem and Nijmegen," *Journal of Transport Geography*, vol. 35, pp. 130–143, 2014.

[10] J. E. Evans and R. H. Pratt, "TCRP Report 95 – Chapter 17—Transit Oriented Development," 2007.

[11] Y. Hayashi and J. Roy, *Transport, Land-Use and the Environment*, Springer US, Land-Use and the Environment, 1996.

[12] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[13] A. Charnes, W. Cooper, A. Y. Lewin, and L. M. Seiford, "Data Envelopment Analysis Theory, Methodology and Applications," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 332-333.

[14] S. F. Gomes Júnior, A. P. D. S. Rubem, J. C. C. B. Soares de Mello, and L. Angulo Meza, "Evaluation of Brazilian airlines nonradial efficiencies and targets using an alternative DEA approach," *International Transactions in Operational Research*, vol. 23, no. 4, pp. 669–689, 2016.

[15] M. Samà, C. Meloni, A. D'Ariano, and F. Corman, "A multi-criteria decision support methodology for real-time train scheduling," *Journal of Rail Transport Planning and Management*, vol. 5, no. 3, pp. 146–162, 2015.

[16] T. Yoshida and K. Tanaka, "Land-use diversity index: a new means of detecting diversity at landscape level," *Landscape and Ecological Engineering*, vol. 1, no. 2, pp. 201–206, 2005.

[17] Q. He, J. Han, D. Guan, Z. Mi, H. Zhao, and Q. Zhang, "The comprehensive environmental efficiency of socioeconomic sectors in China: An analysis based on a non-separable bad output SBM," *Journal of Cleaner Production*, 2017.

[18] R. Cervero, "Sustainable new towns. Stockholm's rail-served satellites," *Cities*, vol. 12, no. 1, pp. 41–51, 1995.

[19] R. D. Knowles, "Transit Oriented Development in Copenhagen, Denmark: from the Finger Plan to Ørestad," *Journal of Transport Geography*, vol. 22, pp. 251–261, 2012.

[20] K. Behan, H. Maoh, and P. Kanaroglou, "Smart growth strategies, transportation and urban sprawl: Simulated futures for Hamilton, Ontario," *The Canadian Geographer*, vol. 52, no. 3, pp. 291–308, 2008.

[21] D. L. Winstead, "Smart growth, smart transportation: A new program to manage growth in Maryland," *Urban Lawyer*, vol. 30, no. 3, pp. 537–539, 1998.

[22] H. Dittmar and G. Ohland, *The New Transit Town: Best Practices In Transit-Oriented Development*, Island Press, 2003.

[23] L. Bertolini, C. Curtis, and J. Renne, "Station area projects in Europe and beyond: Towards transit oriented development?" *Built Environment*, vol. 38, no. 1, pp. 31–50, 2012.

[24] M. Givoni and D. Banister, *Integrated Transport: From Policy to Practice*, Routledge, New York, NY, USA, 2010.

[25] A. Galelo, A. Ribeiro, and L. M. Martinez, "Measuring and Evaluating the Impacts of TOD Measures – Searching for Evidence of TOD Characteristics in Azambuja Train Line," *Procedia - Social and Behavioral Sciences*, vol. 111, pp. 899–908, 2014.

[26] N. L. Widyahari and P. N. Indradjati, "The Potential of Transit-Oriented Development (TOD) and its Opportunity in Bandung Metropolitan Area," *Procedia Environmental Sciences*, vol. 28, pp. 474–482, 2015.

[27] B. P. Y. Loo, C. Chen, and E. T. H. Chan, "Rail-based transit-oriented development: lessons from New York City and Hong Kong," *Landscape and Urban Planning*, vol. 97, no. 3, pp. 202–212, 2010.

[28] C. Curtis, J. L. Renne, and L. Bertolini, "Transit Oriented Development. Making it Happen," *Journal of Transport Geography*, vol. 17, no. 6, p. 312, 2009.

[29] C. D. Higgins and P. S. Kanaroglou, "A latent class method for classifying and evaluating the performance of station area transit-oriented development in the Toronto region," *Journal of Transport Geography*, vol. 52, pp. 61–72, 2016.

[30] M. Kamruzzaman, D. Baker, S. Washington, and G. Turrell, "Advance transit oriented development typology: Case study in brisbane, australia," *Journal of Transport Geography*, vol. 34, pp. 54–70, 2014.

[31] D. S. Vale, "Transit-oriented development, integration of land use and transport, and pedestrian accessibility: Combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon," *Journal of Transport Geography*, vol. 45, pp. 70–80, 2015.

[32] J. L. Renne and J. Wells, *Transit-Oriented Development: Developing a Strategy to Measure Success*, Washington, Wash, USA, 2005.

[33] L. Cavaignac and R. Petiot, "A quarter century of Data Envelopment Analysis applied to the transport sector: A bibliometric analysis," *Socio-Economic Planning Sciences*, vol. 57, pp. 84–96, 2017.

[34] X. Li, J. Yu, J. Shaw, and Y. Wang, "Route-Level Transit Operational-Efficiency Assessment with a Bootstrap Super-Data-Envelopment Analysis Model," *Journal of Urban Planning and Development*, vol. 143, no. 3, p. 04017007, 2017.

[35] J. Sanders, "Linking station node- and place functions to traffic flow: a case study of the Tokyu Den-En Toshi line in Tokyo, Japan," 2015.

[36] R. Cervero, A. Round, T. Goldman, and K.-L. Wu, *Rail Access Modes and Catchment Areas for the BART System*, Berkeley, 1995.

[37] J. Lohmann, E. Andersen, and A. Landex, "GIS-based Approaches to Catchment Area Analyses of Mass Transit," in *Esri International User Conference Proceedings*, pp. 1–13, 2009.

[38] T. Lin, R. Palmer, J. Xia, and D. A. Mcmeekin, "Automatic generation of station catchment areas: A comparison of Euclidean distance transform algorithm and location-allocation methods," in *Proceedings of the 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2014*, pp. 963–967, China, August 2014.

[39] O. Harrison and D. O. Connor, "Measuring Rail Station Catchment Areas in The Greater Dublin Area," in *Proceedings of the ITRN 2012*, 2012.

[40] D. Wheeler and A. Wang, "Catchment Area Analysis Using Generalized Additive Models," *Austin Biometrics Biostat*, vol. 2, no. 3, pp. 1–6, 2015.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 7308058, 16 pages
https://doi.org/10.1155/2018/7308058

Research Article

Optimal Operation of High-Speed Trains Using Hybrid Model Predictive Control

Yingze Yang, Zheng Xu, Weirong Liu, Heng Li, Rui Zhang, and Zhiwu Huang

School of Information Science and Engineering, Central South University, Changsha 410083, China

Correspondence should be addressed to Zhiwu Huang; hzw@csu.edu.cn

Received 1 August 2017; Revised 8 January 2018; Accepted 11 April 2018; Published 15 May 2018

Academic Editor: Paola Pellegrini

The high-speed train operation process is highly nonlinear and has multiple constraints and objectives, which lead to a requirement for the automatic train operation (ATO) system. In this paper, a hybrid model predictive control (MPC) framework is proposed for the controller design of the ATO system. Firstly, a piecewise linear system with state and input constraints is constructed through piecewise linearization of the high-speed train's nonlinear dynamics. Secondly, the piecewise linear system is transformed into a mixed logical dynamical (MLD) system by introducing the auxiliary binary variables. For the transformed MLD system, a hybrid MPC controller is designed to realize the precise control under hard constraints. To reduce the online computation complexity, the explicit control law is computed offline by employing the mixed-integer linear programming (MILP) technique. Simulation results validate the effectiveness of the proposed method.

1. Introduction

With rapid development of high-speed railway, the operation safety, punctuality, and energy consumption of train have received more and more attentions. Nowadays, in China, the high-speed railway generally adopts distributed traction power to improve speed and traction efficiency, where locomotives and wagons are marshalled together to form a high-speed train [1]. For distributed traction power trains, there are some critical issues that need to be solved, such as vehicle traction and brake force allocation, operation safety, punctuality, and energy consumption [2].

To achieve the safety and multiobjective optimization, automatic train control system is developed for high-speed railway, and it is one of the key on-board equipment with high safety assurance [3]. The automatic train operation (ATO) system is the essential component that plays a key role in the train operation process [4]. The ATO system controls the train traction/brake force to follow the reference speed. Moreover, ATO system can reduce the energy consumption and improve riding comfort. Now, it is still a challenge to develop an efficient modeling and control method for ATO system [5].

Currently, the researches of automatic train control are mainly based on single-point model and multipoint model [6]. For single-point model, the whole high-speed train is simplified as a single rigid mass point, which ignores the coupling characteristics among the vehicles and the allocation of the distributed traction/brake force. The dynamic properties of single-point model are simple and easy to design. However, high-speed train adopts distributed traction power mode; thus the single-point model is obviously not able to describe the coupling characteristics and hydrodynamic dispersion characteristics of the train. Thus multipoint model is favoured to study the automatic train control strategy.

Yang and Sun [7, 8] analyzed the coupling characteristics among vehicles of high-speed train and established train multipoint dynamical model. For both push-pull driving and distributed driving train types, the hybrid H_2/H_∞ automatic train controller was designed and was compared by H_2 controller and H_∞ controller, respectively. Lin et al. [9] proposed a control strategy based on improved sliding mode control to overcome the effect of the disturbance in the process of train operation, which had good robustness and effectively restrained the high frequency chattering phenomenon. Thus the work reduced the impact of the frequent switching of control input and increased riding comfort. To simplify the controller design, Song et al. [10] utilized the geometric

topology to reduce multiple positions to one position for multipoint model. In their work, to address the nonlinear saturation constraints of traction and brake forces, the robust adaptive controller was designed with low computing load, which had robustness for train disturbance and uncertainty of model parameters.

Based on the multipoint model, Zhuan and Xia [11] have achieved constructive results on the automatic train control strategy of the heavy-duty train in South Africa. Specifically, Zhuan and Xia [12] put forward three kinds of offline open-loop optimal operation strategies, whose goals are to minimize the workshop bonding force, ensure the safety of train driving, and reduce the maintenance costs of the coupler buffer. Then, Chou and Xia [13] proposed a closed-loop cruising linear quadratic regulator controller to minimize the running cost of electric air brake system in heavy-haul train, in which the control objective includes the speed tracking, coupling force, and energy consumption. In order to overcome the communication limitation, the vehicle barrier was introduced and the controller was reconstructed based on the current orbital slope. In addition, Zhuan and Xia [14] adopted the output feedback adjustment method to regulate the speed of the heavy train and verified the application conditions of their proposed method.

As a summary, Scheepmaker et al. [15] presented an extensive review on energy-efficient train control and the related topic of energy-efficient train timetabling, from the first simplest model of a train running on a level track to the advanced models and algorithms of the last decade dealing with varying gradients and speed limits and including regenerative braking. And there have appeared various theories and technologies to realize the efficient and safety train operation, such as the passivity-based cruise control [16], the robust adaptive control [17], and the iterative learning and fault detection approaches [18, 19].

Li et al. [20] considered the optimal guaranteed cost of cruise control for high-speed train movement. The sufficient condition for the existence of guaranteed cost cruise control law is given in terms of linear matrix inequalities. And a convex optimization problem is formulated to determine the optimal guaranteed cost that minimizes the performance upper bound of the cruise control law. Ye and Liu [21] propose a novel approach to solve the complex optimal train control problems in closed loop by introducing some simplifications. The operation sequence consists of maximum traction, speed holding, coasting, and maximum braking on each subsection of the track or a constant force is applied on each subsection. Zhao et al. [22] proposed a new cluster consensus technique to design the distributed control law, by which the trains can track the desired speeds asymptotically and the distance between the neighboring cars can be kept in the ideal range.

However, with the improvement of the requirement for ATO control performance, the above methods lack the ability to address the nonlinear, multiconstraint, and multiobjective problem based on the multipoint model, which leads to their limitations in practical applications. Especially for online operation process of high-speed trains, it raises the nonlinear resistance force which makes many classical control methods based on the linear resistance force model or the single equilibrium point linearizable model difficult to be implemented. And it is necessary to consider the safe speed, the saturation characteristics of traction and braking units, and the work conditions such as maximization of bonding forces, operating punctuality, energy-efficiency, and reduction of coupling wear forces. The model predictive control has the advantages of fully considering the input and state constraints of the system and the multiobjective optimization problems in the design of the controller, which is more suitable for the design of ATO controller [23, 24].

In this paper, an optimal automatic train control strategy is proposed for high-speed trains based on model predictive control. Due to the high computation complexity of nonlinear train model, the common method is to linearize the nonlinear system model at the equilibrium points [25]. However, if the nonlinear system has a wide range of operating condition, there are many deviations on entire running process for singular linearized model with one equilibrium point. Thus, the piecewise linear systems are utilized to approximate the original system, which can describe complex nonlinear systems accurately [26]. In this paper, the nonlinear running resistance is fitted by multiple line segments to construct the piecewise linear system model of train.

In the traditional controller design based on piecewise linear model, the linear submodel will be determined according to the initial state of receding horizon; then the predicted control is solved based on the determined sublinear model, which is fixed in the entire receding horizon [27]. However, for each actual control step, the system state changes and the corresponding submodel switch may also occur. The fixed submodel is difficult to approximate the original nonlinear model, especially near the model switching point. Therefore, the traditional model prediction control just according to current linear submodel can result in a large prediction error.

In order to solve the above problems, the logic variables are introduced to describe the switch of different linear submodels in this paper. By using this method, the piecewise linear system can be transformed into a hybrid logic dynamical system [28, 29]. For the hybrid logic dynamical system, the model switch can be described by changing the logical variables. Then the hybrid model predictive control (MPC) can be utilized to generate the control inputs based on the constructed mixed logic dynamical system.

The hybrid MPC controller uses the mixed logic dynamical model to predict the future evolution of the system at each time step, and a certain performance index is optimized under operating constraints with respect to a sequence of future input moves. The first of such optimal solution is the control action applied to the plant. For the hybrid MPC controller, the optimal problem that needs to be solved is a mixed integer linear program (MILP) due to the hybrid system description. Because the mixed logic dynamical model ensures that the submodel can switch on the entire receding horizon. Thus, more accurate approximation can be obtained by hybrid MPC and the prediction error is reduced.

Compared to the existing work, the proposed hybrid MPC makes a superior trade-off between the model complexity and control performance. The proposed controller can obtain the optimal input sequence without excessive

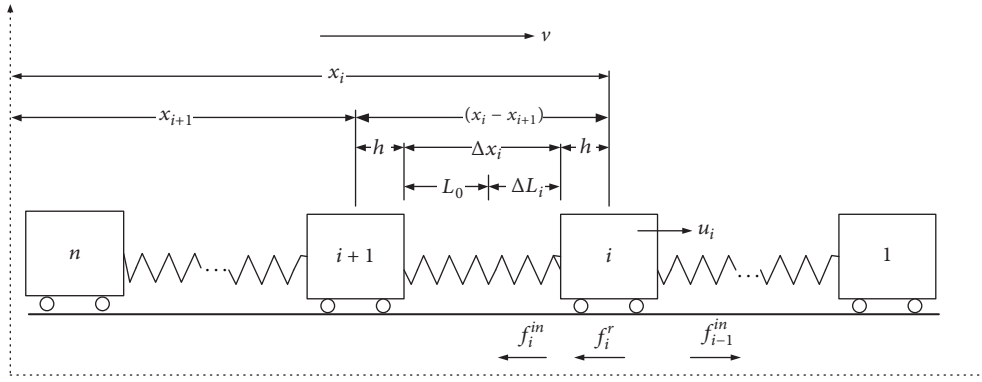


Figure 1: The multiple mass-point model structure of the electric multiple units.

simplifying assumption, while the piecewise linear model reduces the computation complexity under the sufficient precision compared to the classical MPC method. The main contributions in this paper can be summarized as follows.

(1) The nonlinear multiple mass point dynamical model of high-speed trains is constructed. In the model, the complex nonlinear characteristics of running resistance are piecewise linearized and the piecewise linear description is transformed to mixed logic dynamical system.

(2) Based on the MLD system model, a hybrid MPC framework is introduced and the MLD model is used as the prediction model. Based on the MLD prediction model, the control problem can be formulated as a unified mixed integer linear programming (MILP) problem, even the piecewise linear characteristics of train model.

(3) For the formulated MILP problem containing logical variables and continuous variables, the branch and bound method (*B*&*B*) is utilized to transform the original MILP problem into linear programming (LP) by relaxing the logic variable constraints.

The remainder of the paper is organized as follows. In Section 2, we establish the piecewise linear model of the high-speed train and transform the piecewise linear model to MLD model. In Section 3, a mixed logic dynamical system is developed. In Section 4, a hybrid MPC controller is designed by analyzing the performance indices and constraints of high-speed trains. Then, the simulation results are provided to validate the effectiveness of the controller in Section 5. Finally, the major conclusion of this paper is given in Section 6.

2. The Model of High-Speed Train

In this section, we establish the nonlinear multiple mass-point dynamic model of high-speed trains by analyzing their dynamical characteristics. Because the nonlinear model is complex for designing the system controller, the nonlinear curve of the running resistance is piecewise linearized and the piecewise linear model of the train is constructed.

2.1. The Multiple Mass-Point Model of Trains. Figure 1 presents the multiple mass-point model structure for the simplified electric multiple power units. The total number of locomotives and wagons is denoted by n. These vehicles are connected by the vehicle hook couplers. The mechanical properties of the couplers are usually described as the "elastic-damping" system. Thus, the hook coupler force between vehicle i and vehicle $i + 1$ is as follows:

$$f_i^{in} = k_i \Delta L_i + d_i \Delta \dot{L}_i, \tag{1}$$

where $\Delta L_i = x_i - x_{i+1} - h_{i+1} - h_i - L_0$ is the offset value of the relative equilibrium position of the hook coupler between vehicle i and vehicle $i+1$, L_0 is the natural length of the coupler and h_i is the half length of vehicle i, and x_i is the displacement of vehicle i. Since the length of each vehicle is essentially the same and constant (denoted by h) and the hook coupler has the same type, the balance position between two vehicles can be denoted by $L_e = 2h + L_0$. Thus $\Delta L_i = x_i - x_{i+1} - L_e$. And the hook coupler force (1) can be transferred to

$$f_i^{in} = k\left(x_i - x_{i+1} - L_e\right) + d\left(\dot{x}_i - \dot{x}_{i+1}\right). \tag{2}$$

The multiple mass-point dynamic equation of electric multiple power units can be described as

$$\begin{aligned} m_1 \dot{v}_1 &= u_1 - f_1^{in} - f_1^r, \\ m_i \dot{v}_i &= u_i + f_{i-1}^{in} - f_i^{in} - f_i^r, \quad (i = 2, 3, \ldots, n-1) \\ m_n \dot{v}_n &= u_n + f_{n-1}^{in} - f_n^r, \\ \dot{x}_i &= v_i, \quad (i = 1, 2, \ldots, n), \end{aligned} \tag{3}$$

where u_i is the control input of vehicle i. When the train is in traction state or cruise state, $u_i > 0$ is traction force. When the train is in brake state, $u_i < 0$ is the brake force. When the train is in coasting state, $u_i = 0$, implying that there is no traction force and brake force. For locomotive vehicles, they can provide both traction force and brake force, while for wagon vehicles, they can just provide brake force. Assume that f_i^r denotes the running resistance of vehicle i; that is,

$$\begin{aligned} f_i^r &= f_i^b + f_i^a \\ &= m_i c_0 + m_i c_v v_i + m_i c_a v_i^2 + m_i g \cdot \sin\theta_i + 0.004 \\ &\quad \cdot m_i D_i, \end{aligned} \tag{4}$$

where f_i^b and f_i^a are the basic resistance and appended resistance, respectively, and m_i and v_i are the mass and

Table 1: The basic resistance parameters of Chinese electric multiple units high-speed train CRH3.

Parameter	Value	Unit
m_i	47.5	t
c_0	7.75×10^{-3}	Nkg^{-1}
c_v	2.28×10^{-4}	$\mathrm{Ns(mkg)}^{-1}$
c_a	1.66×10^{-5}	$\mathrm{Ns}^2(\mathrm{m}^2\mathrm{kg})^{-1}$

Table 2: The piecewise linear parameters of basic resistance.

j	α_j (s^{-1})	β_j (Nkg^{-1})	Speed ($\mathrm{m{\cdot}s}^{-1}$)
1	0.0008	0.0044	[0, 35]
2	0.0020	−0.0363	[35, 70]
3	0.0031	−0.1100	[70, 100]

velocity of vehicle i, θ_i is the ramp angle, and D_i is the curvature of rail.

2.2. The Multiple Mass-Point Piecewise Linearized Model of Trains. According to the dynamic analysis of high-speed trains, the basic resistance of the train is composed of mechanical resistance and air resistance. When the train is running at low speed, for instance, below 44.44 m/s (160 km/h), the basic resistance is mainly mechanical resistance, and air resistance can be ignored. However, when the train runs at high speed, air resistance will be dominant, and it is related to the square item of the train speed with nonlinear characteristics. As the speed of high-speed trains becomes higher, this nonlinear relationship is becoming stronger. In order to facilitate the controller design of nonlinear model, we consider linearization of the nonlinear model of the train.

For nonlinear systems, the linear model is usually obtained by using the equilibrium linearization method. However, for high-speed trains with varying speeds, it is not appropriate to choose one equilibrium point. Piecewise linear systems approximate the original system by using multiple linear subsystems, which can approximate most complex nonlinear systems accurately. Thus, the least square method is adopted to fit the train running resistance linearly. By using the method, the piecewise linear system model of the train is established.

The train running parameters are selected from Chinese electric multiple units high-speed train CRH3, which are shown in Table 1. Basic resistance after the piecewise linearization is presented in Figure 2, where the train speed range is from 0 m/s to 100 m/s (0–360 km/h). Through lots of experiments, we set $v_a = 35$ m/s and $v_b = 70$ m/s as the boundaries of piecewise linearization. The original nonlinear equations are divided into three linear equations, and the piecewise linear parameters of basic resistance are shown in Table 2.

It can be seen from Figure 2 that the running resistance of the train unit after linearization is already a good approximation to the original nonlinear curve. In fact, using more sublinear model is better for the approximation effect of the original nonlinear model, but it will increase the calculation of the controller. In order to further verify the approximate effect of the linearized train running resistance, the error of

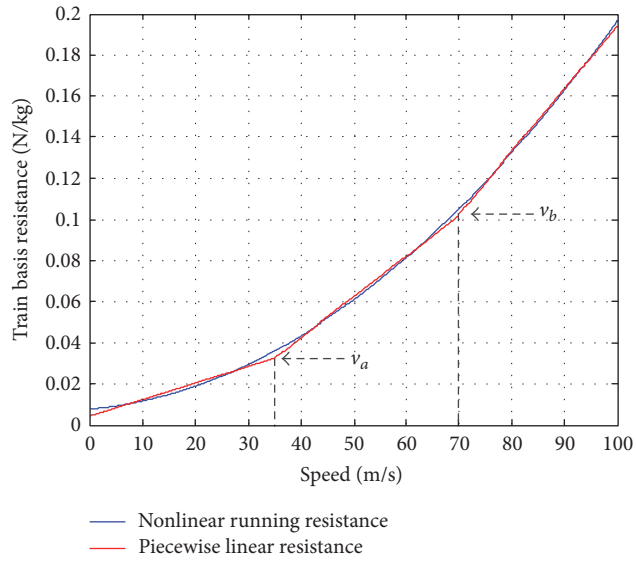


Figure 2: The piecewise linearization of train basic resistance.

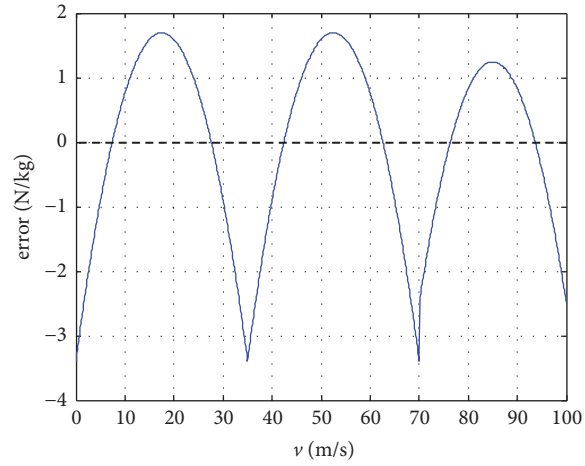


Figure 3: The error between the piecewise linearization of train basic resistance and the original true basic resistance.

the approximate linear curve and the original curve is drawn, which is shown in Figure 3. From Figure 3, the error between the piecewise linearization of running basic resistance and the

practical basic resistance is less than 3.3 N/kg, which is just 0.55% compared to the biggest traction force of single vehicle, 600 N/kg.

Thus, the linearization of train basic running resistance can be approximated as

$$\omega_0 = \alpha_j v + \beta_j, \quad j = 1, 2, 3. \tag{5}$$

In fact, due to the force constraints of hook couplers, there is small difference among the operative velocities of vehicles. Thus, the velocity of vehicle 1 can be taken as the piecewise point of piecewise linear model. Taking it into the train multiple mass-points model, the piecewise linear dynamic model of train can be described as

$$\begin{aligned} \dot{x}(t) &= A_1^c x(t) + B^c u(t) + f_1^c \quad v_i \le v_a \\ \dot{x}(t) &= A_2^c x(t) + B^c u(t) + f_2^c \quad v_a < v_i \le v_b \\ \dot{x}(t) &= A_3^c x(t) + B^c u(t) + f_3^c \quad v_i > v_b, \end{aligned} \tag{6}$$

where the state variable $x = [v_1, v_2, \ldots, v_n, x_1, x_2, \ldots, x_n]^T$, control input $u = [u_1, u_2, \ldots, u_n]^T$, system matrix $A_j^c \in \mathfrak{R}^{2n\times 2n}$, $B^c \in \mathfrak{R}^{2n\times n}$, and $f_j^c \in \mathfrak{R}^{2n\times 1}$ are as follows:

$$\begin{aligned} A_j^c &= \begin{bmatrix} A_{11}^j & A_{12}^j \\ I_{n\times n} & 0_{n\times n} \end{bmatrix}, \\ B^c &= \begin{bmatrix} B_{11} \\ 0_{n\times n} \end{bmatrix}, \\ f_j^c &= \left[f_j^1, f_j^2, \ldots, f_j^n, 0, 0, \ldots, 0\right]^T, \\ A_{11}^j &= \begin{bmatrix} -\frac{d}{m_1} - \alpha_j & \frac{d}{m_1} & 0 & \cdots & 0 & 0 & 0 \\ \frac{d}{m_2} & -\frac{2d}{m_2} - \alpha_j & \frac{d}{m_2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & \frac{d}{m_{n-1}} & -\frac{2d}{m_{n-1}} - \alpha_j & \frac{d}{m_{n-1}} \\ 0 & 0 & 0 & \cdots & 0 & \frac{d}{m_n} & -\frac{d}{m_n} - \alpha_j \end{bmatrix}, \\ A_{12}^j &= \begin{bmatrix} -\frac{k}{m_1} & \frac{k}{m_1} & 0 & \cdots & 0 & 0 & 0 \\ \frac{k}{m_2} & -\frac{2k}{m_2} & \frac{k}{m_2} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{k}{m_{n-1}} & -\frac{2k}{m_{n-1}} & \frac{k}{m_{n-1}} \\ 0 & 0 & 0 & \cdots & 0 & \frac{k}{m_n} & -\frac{k}{m_n} \end{bmatrix}, \end{aligned} \tag{7}$$

and $B_{11} = \mathrm{diag}\{1/m_1, 1/m_2, \ldots, 1/m_n\}$, and $f_j^i = k_1 L_e/m_i - \beta_j - 9.98\sin\theta_i - 0.004D_i$. In practical controller design, L_e can be eliminated by state transformation, so the equilibrium position is set to zero for each subsystem.

Then, the zero-order holder is used to discretize the continuous time model. By setting the sampling period as T_s, the discrete-time piecewise linear model can be obtained as

$$x(k+1) = A_j^d x(k) + B_j^d u(k) + f_j^d, \quad j = 1, 2, 3, \tag{8}$$

where $A_j^d = e^{A_j^c T_s}$, $B_j^d = \int_0^{T_s} e^{A_j^c \tau} d\tau B^c$, and $f_j^d = \int_0^{T_s} e^{A_j^c \tau} d\tau f_j^c$.

3. Hybrid MPC Based ATO Design

A discrete-time piecewise linear model of high-speed trains has been established, which consists of multiple piecewise linear submodels. For the traditional predictive controller design based on the piecewise linear model, the linear submodel based on the current state is determined at the beginning of the prediction and then the predictive control is operated according to the determined linear submodel. However, the predictive submodel has been fixed and does not change throughout the prediction horizon, which cannot

reflect the system nonlinear characteristics by only using a fixed submodel. Actually, the predictive system state may change, and the corresponding submodel may also switch during the solution process. Using a fixed submodel is not enough to express the nonlinearity of the original model, especially near the switching point between submodels. Therefore, the traditional predictive control based on piecewise linear model may produce large prediction errors.

In order to address the above issues, a mixed logic dynamical system is introduced in this paper, which is proposed by Bemporad and Morari in the context of MPC technology [28]. By introducing logical variables to represent different linear submodels, the piecewise linear system can be transformed into a mixed logic dynamical system, which is a global dynamical system covering the whole nonlinear dynamics. Based on the model predictive control of the mixed logic dynamical system, the switching of the model can be expressed by the change of the logic quantity in the model. The model can be switchable throughout the predicted horizon, which ensures more accurate approximation of the original nonlinear model and reduces the prediction error. According to the propositional transformation of the mixed logic dynamic system and mixed integer linear inequality conversion rules, we transform the train piecewise linear model into a mixed integer dynamic model by introducing the binary variables [28].

Two binary variables $\delta_1(k)$ and $\delta_2(k)$ are given by

$$\begin{aligned} &[v_i \le v_a] \Longleftrightarrow [\delta_1(k) = 1] \\ &[v_i \le v_b] \Longleftrightarrow [\delta_2(k) = 1]. \end{aligned} \tag{9}$$

Based on $v_i \in [v_{\min}, v_{\max}]$, (9) can be written as follows:

$$\begin{aligned} &\delta_1(k)(v_a - v_{\min}) \ge v_i(k) - v_{\min} \\ &\delta_1(k)(v_a - v_{\max}) \le v_a - v_i(k) - \varepsilon \\ &\delta_2(k)(v_b - v_{\min}) \ge v_i(k) - v_{\min} \\ &\delta_2(k)(v_b - v_{\max}) \le v_b - v_i(k) - \varepsilon, \end{aligned} \tag{10}$$

where ε is a small positive number and generally takes 10^{-6}. It is used to transform a strict equality into a nonstrict inequality.

By defining $\Phi_i(k) = A_i^d x(k) + B_i^d u(k) + f_i^d$, the relationship between $\delta_1(k)$, $\delta_2(k)$, and $x(k+1)$ can be shown in Table 3.

Then, the piecewise linear system (8) can be described as

$$\begin{aligned} x(k+1) = {} & \delta_1(k)\delta_2(k)\Phi_1(k) \\ & + (1 - \delta_1(k))\delta_2(k)\Phi_2(k) \\ & + (1 - \delta_1(k))(1 - \delta_2(k))\Phi_3(k). \end{aligned} \tag{11}$$

We introduce a new variable $\delta_3(k) = \delta_1(k)\delta_2(k)$, which satisfies

$$\begin{aligned} &-\delta_1(k) + \delta_3(k) \le 0 \\ &-\delta_2(k) + \delta_3(k) \le 0 \\ &\delta_1(k) + \delta_2(k) - \delta_3(k) \le 1. \end{aligned} \tag{12}$$

Table 3: The relationship between $\delta_1(k)$, $\delta_2(k)$, and $x(k+1)$.

$\delta_1(k)$	$\delta_2(k)$	$x(k+1)$
1	1	$\Phi_1(k)$
0	1	$\Phi_2(k)$
0	0	$\Phi_3(k)$

In addition, the auxiliary mixed logical vectors and variables are introduced, that is, $z_1^x(k) = \delta_1(k)x(k)$, $z_2^x(k) = \delta_2(k)x(k)$, $z_3^x(k) = \delta_3(k)x(k)$, $z_1^u(k) = \delta_1(k)u(k)$, $z_2^u(k) = \delta_2(k)u(k)$, and $z_3^u(k) = \delta_3(k)u(k)$, which can be expressed as

$$\begin{aligned} &x_{\min}\delta_i(k) \le z_i^x(k) \le x_{\max}\delta_i(k) \\ &x(k) - x_{\max}(1 - \delta_i(k)) \le z_i^x(k) \\ &\quad \le x(k) - x_{\min}(1 - \delta_i(k)) \\ &u_{\min}\delta_i(k) \le z_i^u(k) \le u_{\max}\delta_i(k) \\ &u(k) - u_{\max}(1 - \delta_i(k)) \le z_i^u(k) \\ &\quad \le u(k) - u_{\min}(1 - \delta_i(k)), \end{aligned} \tag{13}$$

for $i = 1, 2, 3$, where $x \in [x_{\min}, x_{\max}]$ and $u \in [u_{\min}, u_{\max}]$. The maximum tract input $u_{\max}$ is a nonincreasing function of speed, which is approximated by a piecewise linear, quadratic, and/or hyperbolic function of speed [15]. Thus it can be described as a group of hyperbolic or parabolic formulas and each formula approximates the actual traction force for a certain speed interval [30], such as

$$u_{\max}(v) = c_{1,j} + c_{2,j}v + c_{3,j}v^2, \quad v \in [v_j, v_{j+1}] \tag{14}$$

or

$$\begin{aligned} u_{\max}(v) = {} & \frac{c_{h,j}}{v}, \\ & v \in [v_j, v_{j+1}], \text{ for } j = 1, 2, \ldots, M-1, \end{aligned} \tag{15}$$

where the piecewise point v_j and v_{j+1} and the parameters c_{1j}, c_{2j}, c_{3j}, and c_{hj} can be determined by train operation experiments.

Then (8) can be transformed into the following mixed logical dynamical (MLD) system:

$$\begin{aligned} x(k+1) = {} & Ax(k) + B_1u(k) + B_2\delta(k) \\ & + B_3z(k) + B_4 \\ E_2\delta(k) + E_3z(k) \le {} & E_1u(k) + E_4x(k) + E_5, \end{aligned} \tag{16}$$

where $z(k) = [z_1^x(k)^T\ z_2^x(k)^T\ z_3^x(k)^T\ z_1^u(k)\ z_2^u(k)\ z_3^u(k)]^T$, $\delta(k) = [\delta_1(k)\ \delta_2(k)\ \delta_3(k)]^T$, $A = A_3^d$, $B_1 = B_3^d$, $B_4 = f_3^d$, $B_2 = [-f_3^d, f_2^d - f_3^d, f_3^d - f_2^d + f_1^d]$, and $B_3 = [-A_3^d, A_2^d - A_3^d, A_1^d - A_2^d + A_3^d, -B_3^d, B_2^d - B_3^d, B_1^d - B_2^d + B_3^d]$.

Specifically, the resulting MLD model (14) is obtained by using the HYSDEL compiler [31]. Matrices $E_1, \ldots, E_5$ include constraints (16). They can be obtained and analyzed in Matlab using, for example, the Hybrid Toolbox [32].

4. Controller Design

Based on the MLD system model, a hybrid MPC framework is introduced in this section. Within this, the MLD model is used as the prediction model. Then the control problem is formulated as a constrained finite time optimal control problem and converted into mixed integer linear programming (MILP) problem, for which efficient solvers are available.

4.1. Formulation of Optimal Control Problem. Automatic train control needs to track the desired velocity profile quickly such that the timeliness of the interstation travel can be ensured. With the development of high-speed railway, energy-saving driving has become an important goal of automatic train control. As the running distance of the train between stations is fixed, the energy consumption directly related to the traction and braking force. Thus, the control input is used to express energy consumption through this paper. In addition, the in-train force should be kept in the desired range to ensure the safety of the train.

Therefore, the optimal problem can be established as follows, which considers the speed tracking error, control input, and the in-train force.

$$J\left(u_N(k)\right) = \sum_{i=0}^{N-1} \left\|\varepsilon\left((k+i) \mid k\right)\right\|_Q^1 + \left\|u\left((k+i) \mid k\right)\right\|_R^1 + \left\|x_{in}\left((k+i) \mid k\right)\right\|_H^1, \tag{17}$$

where k is the current sampling time; $u_N(k) = [u(k), \ldots, u(k+N-1)]$ is the sequence of N control inputs. $\varepsilon((k+i) \mid k) = v((k+i) \mid k) - v^r(k+i) \in \mathbb{R}^{n\times 1}$ represents the difference between actual speed v and reference speed v^r, $u((k+i) \mid k) = [u_1(k+i), u_2(k+i), \ldots, u_n(k+i)]^T$ represents the control input, and $x_{in}((k+i) \mid k) = (x_{in}^1(((k+i) \mid k), \ldots, x_{in}^{n-1}(((k+i) \mid k))$ represents the relative displacements between neighboring cars. 1 means 1-norm and Q, R, and H are the weight matrixes, respectively.

In order to ensure the safe and efficient operation of the train, some constraints must be satisfied as follows.

(1) Constraint on Control Input. The tractive and braking forces are bounded due to the nature physical characteristics of the traction motor. Thus, the control input is set as $-F_{\max}^B \le u_i(k) \le F_{\max}^T$.

(2) Constraint on Maximum Operational Velocity. The maximum speed of the train depends first on the maximum speed allowed by the train itself. Second, when the line conditions and operating conditions change, the maximum allowable speed of the train may change.

(3) Constraint on In-Train Force. The in-train force should be in the range $[-f_{\max}^{in}, f_{\max}^{in}]$ to keep the operational safety of the trains.

(4) Constraint on Coupler Deformation. The coupler deformation should be in the range $[-\Delta L_{\max}, \Delta L_{\max}]$ to keep the operational safety of the trains.

Then, the constraints considered can be formulated by

$$\begin{aligned} -F_{\max}^B &\le u_i(k) \le F_{\max}^T \\ 0 &\le v_i(k) \le v_{\max} \\ -f_{\max}^{in} &\le f_{i,i+1}^{in} \le f_{\max}^{in} \\ -\Delta L_{\max} &\le \Delta L_i \le \Delta L_{\max}. \end{aligned} \tag{18}$$

4.2. Hybrid MPC Controller Design. Based on (17) and (18), a constrained optimal control problem can be constructed as follows:

$$\begin{aligned} \min \quad & J\left(u_N(k)\right) \\ & = \sum_{i=0}^{N-1} \left\|\varepsilon\left((k+i) \mid k\right)\right\|_Q^1 + \left\|u\left((k+i) \mid k\right)\right\|_R^1 \\ & \quad + \left\|x_{in}\left((k+i) \mid k\right)\right\|_H^1 \\ \text{s.t.} \quad & x(k+1) \\ & = Ax(k) + B_1 u(k) + B_2 \delta(k) + B_3 z(k) + B_4 \\ & E_2\delta(k) + E_3 z(k) \le E_1 u(k) + E_4 x(k) + E_5 \\ & -F_{\max}^B \le u_i(k) \le F_{\max}^T \\ & 0 \le v_i(k) \le v_{\max} \\ & -f_{\max}^{in} \le f_{i,i+1}^{in} \le f_{\max}^{in} \\ & -\Delta L_{\max} \le \Delta L_i \le \Delta L_{\max}. \end{aligned} \tag{19}$$

In order to facilitate the solution of the optimal problem, (19) can be transformed to the linear objective function (20) by introducing the auxiliary variables $w^\varepsilon(k+i \mid k) = \|\varepsilon((k+i) \mid k)\|_Q^1$, $w^u(k+i \mid k) = \|u((k+i) \mid k)\|_R^1$, and $w^{x_{in}}(k+i \mid k) = \|x_{in}((k+i) \mid k)\|_H^1$.

$$J\left(u_N(k)\right) = w^\varepsilon(k+N \mid k) + \sum_{i=0}^{N-1} w^\varepsilon(k+i \mid k) + w^u(k+i \mid k) + w^{x_{in}}(k+i \mid k) \tag{20}$$

with some linear inequalities:

$$\begin{aligned} w^\varepsilon(k+i \mid k) &\ge \pm\varepsilon\left((k+i) \mid k\right) \\ w^u(k+i \mid k) &\ge \pm u\left((k+i) \mid k\right) \\ w^{x_{in}}(k+i \mid k) &\ge \pm x_{in}\left((k+i) \mid k\right). \end{aligned} \tag{21}$$

In addition, the state update equation (22) can be obtained by the MLD model.

$$\begin{aligned} x(k+i) = A^i x(k) + \sum_{j=0}^{i-1} A^j \left(B_1 u\left(k+i-j-1\right)\right. \\ \left. + B_2\delta\left(k+i-j-1\right) + B_3 z\left(k+i-j-1\right)\right) \\ E_2\delta(k) + E_3 z(k) \le E_1 u(k) + E_4 x(k) + E_5. \end{aligned} \tag{22}$$

Table 4: Definition of some notations.

Notation	Definition
P	The MILP problem
J_0^*	The minimum upper bound of optimal performance of train operation
V^*	The optimal solution vector of optimal performance index for train operation
Π	The object table of recording all LP subproblems that need to be solved
$J(\overline{V})$	The optimal performance of the current relaxed LP problem
$\overline{V}$	The optimal solution vector of the relaxed LP subproblem
Q_{num}	Number of LP subproblems that have been solved
Maxnum LP$(j)[P_0]$, LP$(j)[P_1]$	The maximum number of LP problems

Then, based on (20), (21), and (22), the optimal control problem can be converted to a mixed integer linear programming (MILP) problem as follows:

$$J(V,\theta) = \min_{V} \quad C^T V \tag{23a}$$

$$\text{st.} \quad GV \le W + S\theta(k), \tag{23b}$$

where $\theta(k) = [x(k)^T, v^r(k), \ldots, v^r(k+N_p-1)]^T$ is commonly represented as the parameter vector and $V = [u'_{N_p}(k), \widetilde{\delta}(k)^T, \widetilde{z}(k)^T, \widetilde{w}(k)^T]^T$ is the optimization vector. Let

$$\widetilde{\delta}(k) = \begin{bmatrix} \delta(k) \\ \delta(k+1) \\ \vdots \\ \delta\left(k+N_p-1\right) \end{bmatrix},$$

$$\widetilde{z}(k) = \begin{bmatrix} z(k) \\ z(k+1) \\ \vdots \\ z\left(k+N_p-1\right) \end{bmatrix}, \tag{24}$$

$$\widetilde{w}(k) = \left[w^{\varepsilon}(k), w^{u'}(k), \ldots, w^{\varepsilon}\left(k+N_p-1\right), w^{u'}\left(k+N_p-1\right)\right]^T;$$

$C^T = [0,\ldots,0,1,\ldots,1]$. The matrices G, W, and S are obtained according to (16), (18), and (21).

At each sampling step k, the MILP problem is solved based on the current parameter vector $\theta(k)$, and an optimal sequence of future control input $u'_{N_p}(k)$ can be obtained. At time $k+1$, the parameter vector is updated to $\theta(k+1)$, and a new MILP problem is solved to get the new optimal control input. This procedure is repeated at each sampling step, and the problems can be solved in real time. In this paper, only $N_c = N_p$ is considered.

4.3. Solution to MILP Problem. The problem of automatic train control is established as a constrained finite time optimal control problem and further transformed into mixed integer linear programming (MILP) problem by analyzing the operation target and constraints. Since the constructed MILP problem contains logical variables and continuous variables, this section uses the branch and bound method (B&B) to solve this problem so that, in each control cycle, the optimal traction of each vehicle or braking force output can be obtained.

The basic idea of branch and bound method is to relax the logic variable constraints in the optimization problem. $\delta_i(k) \in \{0,1\}$ in the MLD model of formula (16) can be relaxed as $\delta_i(k) \in [0,1]$, which transforms the original MILP problem into linear programming (LP). Then, the traditional linear programming method is used to solve the optimal subproblem. Algorithm 1 is presented to solve the MILP problem and some notations are defined and presented in Table 4.

5. Performance Evaluation

In this section, we will evaluate the performance of the proposed hybrid MPC ATO algorithm using numerical results. We will first introduce the parameter settings of the ATO system. Specifically, the parameters of the train model are from the CRH-3 high-speed train in China, which are given in Table 5. Therein, $F_{\max}^T$ and $F_{\max}^B$ are the maximum tractive force and braking force, respectively.

5.1. Simulation Scenario. In order to evaluate the performance of the controller, the desired velocity curve including accelerating, decelerating, velocity step increase, velocity step decrease, and constant-velocity stages is characterized by the following equation:

$$v_r(t) = \begin{cases} 0.4t & 0 \le t \le 100 \\ 40 & 100 \le t \le 400 \\ 75 & 400 \le t \le 900 \\ 50 & 900 \le t \le 1150 \\ -t + 1200 & 1150 \le t \le 1200. \end{cases} \tag{25}$$

In (25), the acceleration is 0.4 m/s, and the deceleration is 0.5 m/s, both of which are within the threshold of acceleration and velocity. In the operation of trains, the basic resistance is the main resistance for the trains. Thus, we only consider the basic resistance in this paper.

(1) initialize $J_0^* = +\infty$, $Q_{\text{num}} = 0$, Π = Null;
(2) **for** each $i \in [1, n]$ **do**
(3) $\delta_i(k) \in [0, 1] \leftarrow \delta_i(k) \in \{0, 1\}$;
(4) obtain LP-subproblem LP(i);
(5) $\Pi \leftarrow \Pi \cup \text{LP}(i)$
(6) **end for**
(7) **if** there are no LP(i) $\in \Pi$ that need to be solved **then**
(8) obtain optimal performance index J_0^*;
(9) obtain optimal solution vector V^*;
(10) obtain optimal control input $u(k) \leftarrow V^*$;
(11) **else**
(12) **while** $Q_{\text{num}} <$ Maxnum **do**
(13) **for** each $j \in [n, 1]$ **do**
(14) extracting LP(j) from Π;
(15) **if** LP(j) has feasible solution **then**
(16) obtain $J(\overline{V})$ and $\overline{V}$;
(17) **else**
(18) **if** LP(j) satisfies $\delta_i(k) \in \{0, 1\}$ and $J(\overline{V}) < J_0^*$ **then**
(19) $J_0^* \leftarrow J(\overline{V})$ and $V^* \leftarrow \overline{V}$;
(20) $Q_{\text{num}} = Q_{\text{num}} + 1$;
(21) **end if**
(22) **if** $J(\overline{V}) > J_0^*$ **then**
(23) $Q_{\text{num}} = Q_{\text{num}} + 1$;
(24) **end if**
(25) **if** LP(j) dose not satisfy $\delta_i(k) \in \{0, 1\}$ and $J(\overline{V}) < J_0^*$ **then**
(26) choose $\delta_i(k)$;
(27) $\text{LP}(j)[P_0] \leftarrow \delta_i(k) = 1$;
(28) $\text{LP}(j)[P_1] \leftarrow \delta_i(k) = 0$;
(29) $\Pi \leftarrow \Pi$ LP(j) $\cup$ LP(j)$[P_0]$ $\cup$ LP(j)$[P_1]$;
(30) **else**
(31) $Q_{\text{num}} = Q_{\text{num}} + 1$;
(32) **end if**
(33) $Q_{\text{num}} = Q_{\text{num}} + 1$;
(34) **end if**
(35) **end for**
(36) **end while**
(37) **end if**

Algorithm 1: Solution of MILP problem.

The designed MPC based ATO controller considers the train velocity tracking performance, energy consumption, and intertrain force. With the train parameters, we provide two simulation scenarios to evaluate the performance of the proposed algorithm under different operating conditions. Specifically, we evaluate the performance of the controller with different train types: all-locomotive train and partial-locomotive train.

5.1.1. Scenario A. In this scenario, we evaluate the performance of the ATO controller with the all-locomotive train. To simplify the simulation, we choose the number of trains as $n = 4$, and the train is comprised of all locomotives. This implies that each locomotive has the tracking motor and braking unit and thus can provide both the tracking force and the braking force. The parameters of the train are shown in Table 5, where m_i is the mass of locomotive i. Since the numbers of passengers in each locomotive may be different, we choose $m_1 = 48.5t$, $m_2 = 52.5t$, $m_3 = 51.0t$, $m_4 = 47.5t$. For the MPC controller, the prediction horizon N_p and the control horizon N_c also affect the controller performance. Without loss of generality, we choose $N_p = N_c = 3$. The weighting matrix gain is chosen as

$$Q = \text{diag}\left(\underbrace{4, \ldots, 4}_{n}\right),$$
$$R = \text{diag}\left(\underbrace{1, \ldots, 1}_{n}\right), \tag{26}$$
$$H = \text{diag}\left(\underbrace{1, \ldots, 1}_{n-1}\right).$$

5.1.2. Scenario B. In this scenario, we evaluate the effectiveness of the ATO controller in partial-locomotive trains. To

Table 5: Parameters of the CRH-3 high-speed train.

Parameter	Value	Unit
m_i	47.5	t
c_0	7.75×10^{-3}	Nkg^{-1}
c_v	2.28×10^{-4}	$\mathrm{Ns(mkg)}^{-1}$
c_a	1.66×10^{-5}	$\mathrm{Ns}^2(\mathrm{m}^2\mathrm{kg})^{-1}$
$F_{\max}^T$	300	kN
$F_{\max}^B$	300	kN
k	1×10^4	kNm^{-1}
d	1000	kNsm^{-1}
θ	0	–
D	0	–
$\Delta L_{\max}$	0.2	m
$f_{\max}^{in}$	2000	kN

simplify the simulation, we choose the number of trains as $n = 4$, and the train is comprised of 2 locomotives and 2 wagons. The locomotive can provide both the tracking force and the braking force, while the wagons can only supply the brake forces. In order to guarantee the total force the same as Scenario A, the maximum tracking force of each locomotive is set as $F_{\max}^T = 600$ kN, and the other train parameters are the same as in Table 5.

5.2. Simulation Results. We will provide simulation results of the ATO controller in different operating conditions. The simulations are conducted in Matlab/SIMULINK. We fist consider the simulations in Scenario A, which is based on all-locomotive train. Then we will further consider the simulation results in Scenario B, which is based on partial-locomotive trains. And in Scenario C, we focus on more realistic condition where the track force limit is dynamic.

5.2.1. Scenario A. The velocity curve of each locomotive is shown in Figure 4, where x-axis represents the time and y-axis represents the velocity. From the figure, we can see that the velocities of four locomotives almost remain the same during the operation time. Specifically, in the accelerating and decelerating stages, each locomotive can operate correspondingly according to the given acceleration and deceleration. At $t = 400$ s, we notice that the reference velocity has the step increase. From the zoomed-in figure, we can find that the locomotives move with the almost same velocity, where the maximum velocity error is 0.003 m/s. From the zoomed-in figure from $t = 470$ s to $t = 490$ s, the velocities of locomotives reach the steady value at about $t = 490$ s, where the settling time is 90 s and the steady error is 0.5 m/s. In the steady state, the velocities of locomotives remain fixed. Based on the simulation results, we can find that the train can track the time-varying reference with small settling time and steady error, which verifies the effectiveness of the algorithm.

The tracking and braking forces of locomotives are shown in Figure 5. In the accelerating stage, the resistance of the train increases with the velocity. In order to guarantee the constant acceleration, the tracking forces of locomotives increase gradually. In the deceleration stage, the resistance of

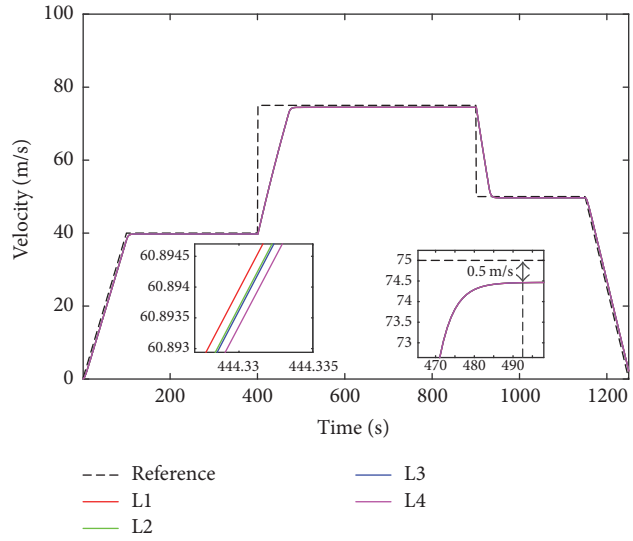


Figure 4: The velocity curve of each locomotive in Scenario A.

the train decelerates with the velocity. In order to guarantee the constant deceleration, the braking forces of locomotives increase gradually. Compared with the acceleration or deceleration stages, the tracking forces of locomotives are small and fixed during the constant velocity stage. This is because the resistance of train keeps having a relatively small value during the constant velocity stage. At $t = 400$ s, the reference has a step increase. In order to accelerate to the desired velocity, the tracking forces of locomotives increase to the maximum rapidly and maintain the maximum tracking force for about 72 s. At $t = 900$ s, the braking forces increase to the maximum to decrease the velocity and maintain about 32 s to decrease to the desired velocity. From the figure, we can find that the control outputs of locomotives are almost the same and thus guarantee the reasonable allocation of braking force and tracking force. When the reference velocity curve changes, each locomotive can regulate the tracking force and braking force based on the difference between the actual velocity and desired velocity. The forces for locomotives are

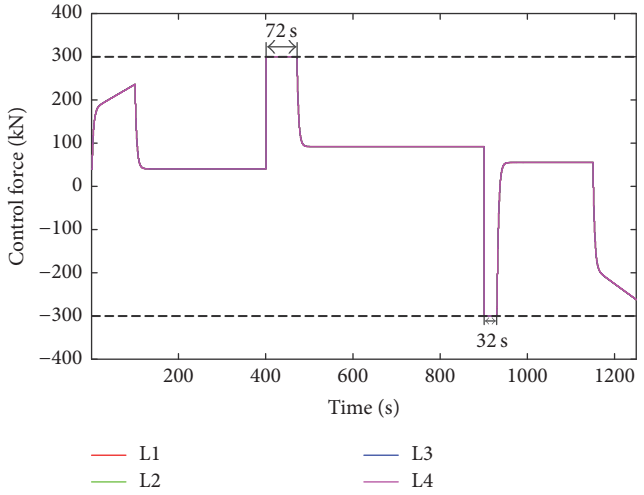


Figure 5: The force output of each locomotive in Scenario A.

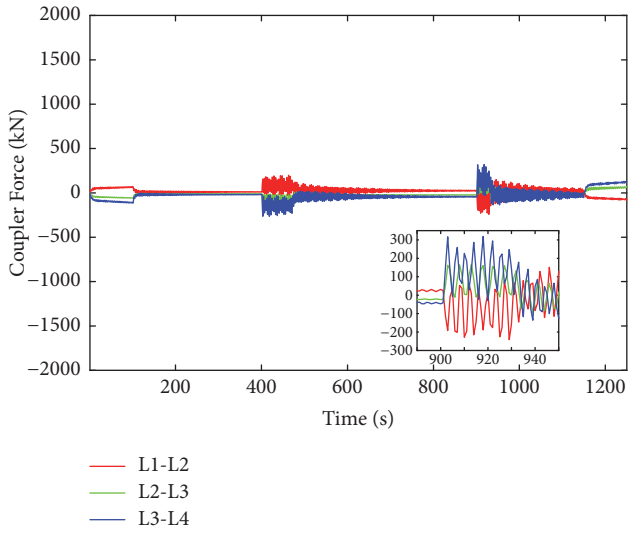


Figure 6: The coupler force among locomotives in Scenario A.

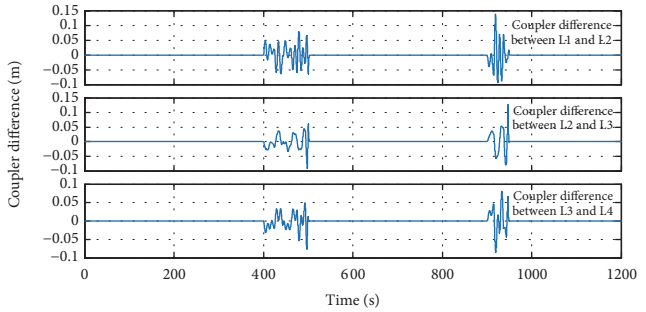


Figure 7: The coupler deformation between two neighbouring locomotives in Scenario A.

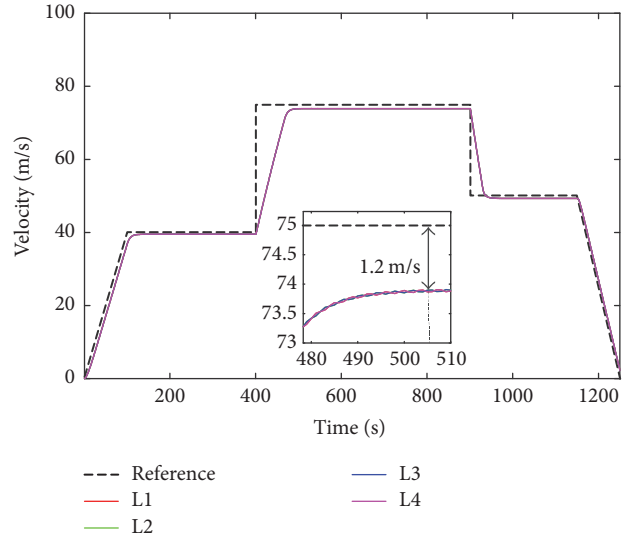


Figure 8: The velocity curve of each vehicle in Scenario B.

almost the same and also satisfy the constraints, which verify the effectiveness of the proposed strategy.

The coupler force among each locomotive is shown in Figure 6. When the coupler force is larger than zero, it means that coupler is pulled. When the coupler force is smaller than zero, the coupler is pushed. From the figure, we can see that the coupler forces among locomotives are always within the safety range. Moreover, the coupler among L1 and L2 is mainly pulled, while the coupler among L3 and L4 is mainly pushed. This is because when we choose the masses of locomotives, the masses of L3 and L4 are relatively smaller than L1 and L2. In order to guarantee the velocity consensus of locomotives, the total forces should be the same and the coupler force can balance the total force of each locomotive. It can be found that, during the constant-velocity stage, constant-accelerating stage, and constant-decelerating stage, the coupler force among locomotives is small. At $t = 200$ s and $t = 400$ s, the coupler force changes dramatically, with the largest coupler pulling force 317 kN and the largest coupler pushing force 241 kN. From the figure, we can see that the ATO controller can guarantee that the coupler force is within the safety range, which verifies the effectiveness.

The operation safety of high-speed trains can be evaluated by the position difference (coupler deformation) between neighboring locomotives. If the deformation is larger than the maximum 0.2 m or less than the minimum −0.2 m, the coupler may have been broken so the safety cannot be guaranteed. The operation safety of high-speed trains can be further illustrated from Figure 7, which shows that the coupler deformation is always with the threshold [−0.2 m, 0.2 m]. This is because the in-train force guarantees that the position difference between neighboring locomotives cannot be too large or too small.

5.2.2. Scenario B. The simulation in Scenario A verifies the effectiveness of the hybrid model predictive controller for all-locomotive case. The metrics include velocity tracking, energy consumption, and coupler force.

In Scenario B, we verified the effectiveness of the ATO controller with 2 locomotives and 2 wagons. The velocity curves are shown in Figure 8. With the same controller

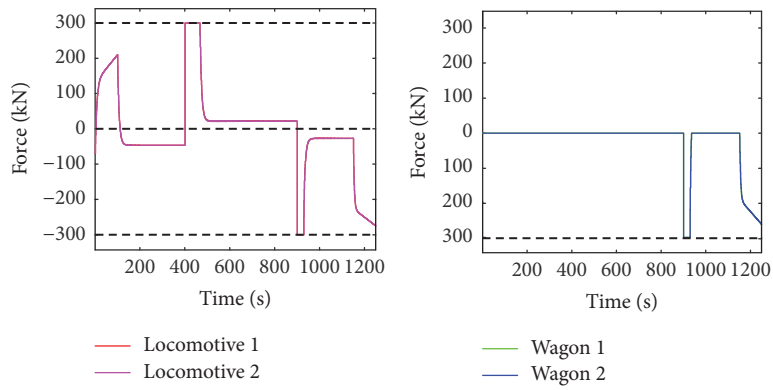


Figure 9: The forces of locomotives and wagons in Scenario B.

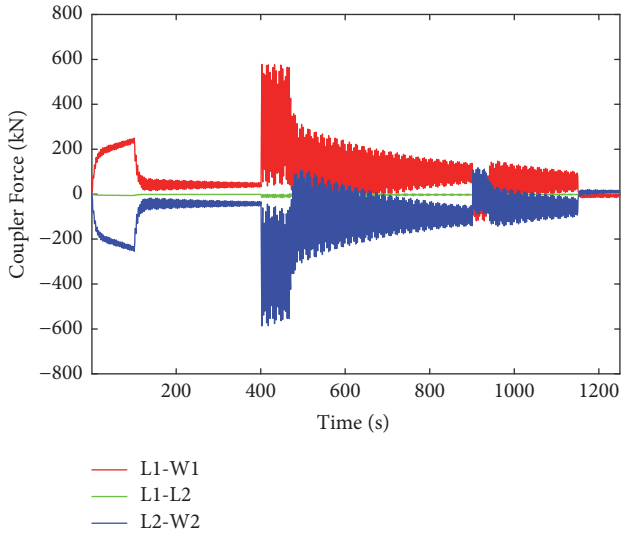


Figure 10: The coupler force among vehicles in Scenario B.

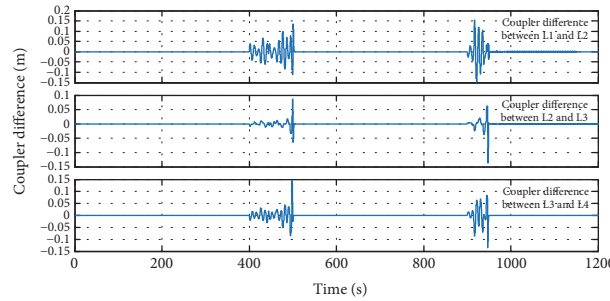


Figure 11: The coupler deformation between neighboring locomotives/wagons in Scenario B.

parameters, at $t = 400$ s, where the reference velocity has a step change, the convergence time of the controller from the change at $t = 400$ s to the steady state at $t = 508$ s implies that the convergence time increases by 18 s compared with the all-locomotive case. The steady error is 1.2 m/s, which increases by 0.7 m/s compared with the all-locomotive case. Moreover, the velocities at the steady stage have some oscillations around the steady value.

The force output of each vehicle is shown in Figure 9, where the left subfigure shows the force output of the locomotive, and the right subfigure shows the force output of the wagon. From the figure, we can see that, during the acceleration and constant-velocity stages, only the locomotive has the tracking force output, and the tracking force of wagons is zero. In the deceleration stage, both the locomotives and wagons supply the braking forces.

Figure 10 shows the coupler force among vehicles. Compared with the all-locomotive trains, the coupler forces between L1 and W1 and between L2 and W2 are kept at large values. This is because locomotive L1 supplies the pulling force and locomotive L2 supplies the pushing force, while wagons W1 and W2 cannot supply the tracking force. The maximum coupler force is 578 kN, which is within the safety threshold. From the above simulations, we can find that the proposed hybrid MPC ATO controller has a good control effect.

The operation safety of high-speed trains can be illustrated from Figure 11, which depicts the spring deformation between neighboring locomotives/wagons. We can see that the largest deformation is 0.14 m, and the smallest deformation is −0.15 m, which is with the safety threshold [−0.2 m, 0.2 m]. As illustrated in Figure 7, this is because the in-train force guarantees that the position difference between neighboring locomotives cannot be too large or too small. Simulation results show that the proposed ATO controller can guarantee the safe operation of high-speed trains.

5.3. Further Discussions. In this subsection, we further discuss the performance of the proposed ATO controller in terms of practicality, computation efficiency, superiority, and robustness. For simplicity, we only consider Scenario A, where the train is comprised of four locomotives. Scenario B can also be analyzed similarly. The train parameters are set the same as in Table 5, unless otherwise specified.

5.3.1. Practicality. In the previous simulation setup, we assume that the maximum traction force $F^T_{\max}$ and the maximum braking force $F^B_{\max}$ are constants. Despite simplifying the implementation, the constant maximum traction/braking

Table 6: Computational time (s) of each MPC step with and without linearization.

	MPC with linearization			MPC without linearization		
	Min.	Max.	Average	Min.	Max.	Average
$N_c = 3$	$1.6e-3$	$3.8e-3$	$2.5e-3$	$2.1e-1$	$3.6e-1$	$2.4e-1$
$N_c = 5$	$1.8e-3$	$6.4e-3$	$3.5e-3$	$2.5e-1$	$8.3e-1$	$4.7e-1$
$N_c = 7$	$1.9e-3$	$7.3e-3$	$4.6e-3$	$3.9e-1$	$9.8e-1$	$6.5e-1$

force are not realistic as the operation is subject to the power limit of the train. This implies that the maximum traction and braking forces evolve with the operation speed; that is, the maximum traction and braking forces for the CRH-3 train are represented as in [33].

$$F_{\max}^{T} = \begin{cases} 300 - 0.284v, & 0 \le v \le 119.7\,\text{km/h} \\ 266 \times \dfrac{119.7}{v}, & 119.7 < v \le 300\,\text{km/h} \end{cases}$$
$$F_{\max}^{B} = \begin{cases} 300 - 0.281v, & 0 \le v \le 106.7\,\text{km/h} \\ 266 \times \dfrac{119.7}{v}, & 106.7 < v \le 300\,\text{km/h}. \end{cases} \quad (27)$$

The force output of each locomotive with power limit is shown in Figure 12, where we find that the output force of each locomotive evolves with time. It is interesting to note that at $t = 400$ s the maximum traction force decreases at the transition stage, where the speed of the train increases from 40 m/s to 75 m/s. Due to the constant-power operation, the maximum traction force will decrease with the increase of velocity. At $t = 900$ s, it is shown that the maximum braking force increases when the velocity decreases from 75 m/s to 50 m/s.

5.3.2. Computational Time. In this subsection, we consider the computation time of the proposed method with different prediction horizons.

Within different control horizons, the nonlinear MPC and the proposed piecewise MPC are both implemented on the original nonlinear system. The simulations are performed under Windows 7 operating system with Intel Core i5-2400 CPU, 4 GB RAM, on a desktop computer. Both the nonlinear MPC and the proposed piecewise MPC provide comparable dynamic performance. Their computational times are summarized in Table 6. From Table 6, it can be seen that the computational time of each proposed MPC step is less than the nonlinear MPC on all control horizon lengths. It is more obvious when the horizon length is extended. The comparative results present the computation efficiency of the proposed method. To obtain the average computational time of each MPC step, the repeated numbers of MPC steps are 1200.

5.3.3. Superiority. In this subsection, we verify the superiority of the proposed method through comparison with the classical PID method.

The velocity difference between each locomotive and the reference velocity value is plotted in the following figures.

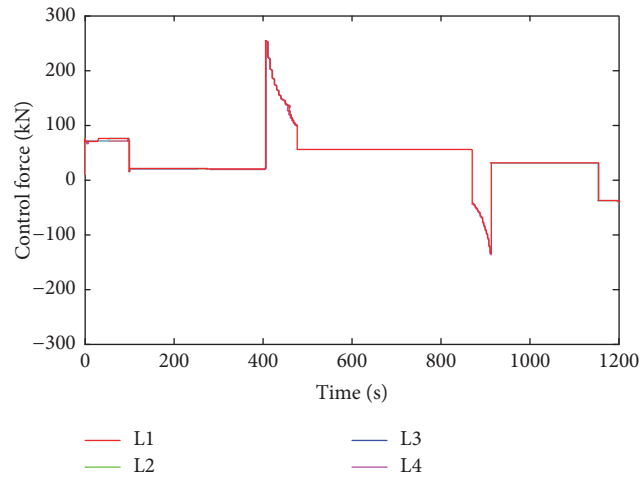


Figure 12: The force output of each locomotive with power limit.

Figures 13(a) and 13(b) show the simulation results under PID and the proposed MPC methods, respectively. From Figure 13(a), the velocities of four locomotives have severe chattering at time 400 s and 900 s when the reference velocity varies, which can further cause drastic changes of the coupler and endanger safe operation of high-speed trains. The proposed method in this paper presents good control performance as shown in Figure 13(b). From this figure, we can see that the velocities of four locomotives do not change frequently and can converge to the desired velocity quickly. This control performance is conducive to the safe operation of the train.

5.3.4. Stochastic Perturbations. In this subsection, we consider the performance of the proposed method with external stochastic perturbations. The stochastic perturbations are added in the system dynamics (6). The maximum amplitude of the perturbation is assumed to be 1. Considering the constant term $\beta_j = 0.0044$, the perturbation is large enough to evaluate the robustness of the proposed method. The external stochastic perturbations are produced by using the Matlab function randn(n), which can create an n-by-n matrix of normally distributed random numbers with mean equal to zero and standard deviation equal to 1. The simulation result of the system with the stochastic perturbations is shown in Figure 14. From this figure, we can see that the practical velocities of locomotives can track the reference values well, though the stochastic perturbations exist. The simulation results show the good robustness of the proposed method.

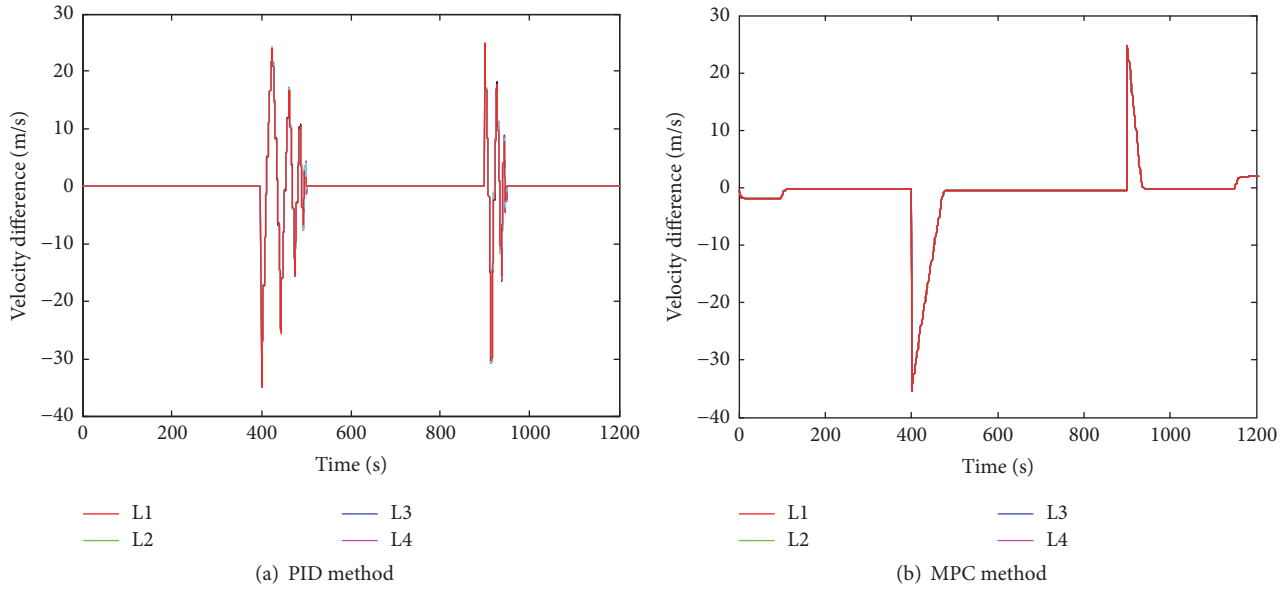


Figure 13: Comparison of MPC method with PID method in velocity error.

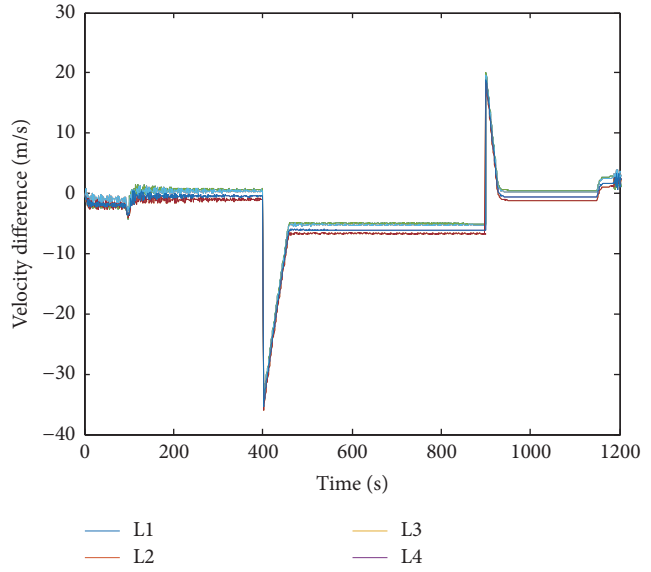


Figure 14: The velocity of each vehicle under MPC method with external stochastic perturbations.

6. Conclusion

In this paper, we investigated the automatic cruising control problem of high-speed trains with the aim of tracking the reference speed and minimizing the energy consumption and guaranteeing the safe operation of high-speed trains. First, we built the piecewise linearization of train resistance and developed the piecewise multiple-point model of trains. Due to the multiple constraints and performance metrics, we proposed a model predictive control method to design the ATO controller. In order to improve the accuracy of the method, we introduced the binary variable and transformed the piecewise model into a mixed logical dynamical model and guarantee that the submodel in the prediction process is still switchable. Then we analyze the operation objective and constraints and transformed the tracking control problem into a constrained finite-time optimal control problem. The optimization problem is solved by the integer linear planning method, and the optimal control solution is obtained. Simulations in the Matlab verified the effectiveness of the proposed algorithm.

In this work, we just consider the nonlinear characteristics of resistance when constructing the train model and neglect the control delay. In the future work, more realistic train models will be investigated to incorporate more nonlinear characteristics and the controller robustness will be improved under time delay.

Disclosure

Part of this paper [2] was presented at the 29th Chinese Control and Decision Conference, Chongqing, China, May 2017.

Conflicts of Interest

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

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (Grants nos. 61379111, 61672537, and 61672539).

References

[1] W. Song, J. Ma, L. Zhou, and X. Feng, "Deadbeat Predictive Power Control of Single-Phase Three-Level Neutral-Point-Clamped Converters Using Space-Vector Modulation for Electric Railway Traction," *IEEE Transactions on Power Electronics*, vol. 31, no. 1, pp. 721–732, 2016.

[2] Z. Xu, Z. Huang, K. Gao, S. Li, R. Zhang, and W. Liu, "Optimal operation of high-speed train using hybrid model predictive control," in *Proceedings of the 2017 29th Chinese Control And Decision Conference (CCDC)*, pp. 3642–3647, Chongqing, China, May 2017.

[3] Y. Luo and H. Xu, "Direct robust adaptive control of high speed train based on nonlinear and time-varying models," *International Journal of Control and Automation*, vol. 6, no. 4, pp. 397–410, 2013.

[4] W. Zhao and C. Gao, "A New Control Method of Automatic Train Operation in Urban Rail Transit Based on Improved Generalized Predictive Control Theory," in *Proceedings of the 2013 International Conference on Electrical and Information Technologies for Rail Transportation (EITRT2013)-Volume I*, vol. 287 of *Lecture Notes in Electrical Engineering*, pp. 567–573, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[5] Y. Wang, R. Luo, Z. Yu, and B. Ning, "Study on ato control algorithm with consideration of atp speed limits," *Journal of the China Railway Society*, vol. 5, p. 017, 2012.

[6] A. Albrecht, P. Howlett, P. Pudney, X. Vu, and P. Zhou, "The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points," *Transportation Research Part B: Methodological*, vol. 94, pp. 482–508, 2016.

[7] C.-D. Yang and Y.-P. Sun, "Robust cruise control of high speed train with hardening/softening nonlinear couple," in *Proceedings of the American Control Conference*, vol. 1999, pp. 2200–2204.

[8] C. Yang and Y. Sun, "Mixed H_2/H_∞ cruise controller design for high speed train," *International Journal of Control*, vol. 74, no. 9, pp. 905–920, 2001.

[9] C-J. Lin, S-H. Tsai, C-L. Chen, and T.-H. S. Li, "Extended sliding-mode controller for high speed train," in *Proceedings of the 2010 International Conference on System Science and Engineering (ICSSE)*, pp. 475–480, Taipei, Taiwan, July 2010.

[10] Q. Song, Y.-D. Song, T. Tang, and B. Ning, "Computationally inexpensive tracking control of high-speed trains with traction/braking saturation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1116–1125, 2011.

[11] X. Zhuan and X. Xia, "Optimal scheduling and control of heavy haul trains equipped with electronically controlled pneumatic braking systems," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 6, pp. 1159–1166, 2007.

[12] X. Zhuan and X. Xia, "Cruise control scheduling of heavy haul trains," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 4, pp. 757–766, 2006.

[13] M. Chou and X. Xia, "Optimal cruise control of heavy-haul trains equipped with electronically controlled pneumatic brake systems," *Control Engineering Practice*, vol. 15, no. 5, pp. 511–519, 2007.

[14] X. Zhuan and X. Xia, "Speed regulation with measured output feedback in the control of heavy haul trains," *Automatica*, vol. 44, no. 1, pp. 242–247, 2008.

[15] G. M. Scheepmaker, R. M. Goverde, and L. . Kroon, "Review of energy-efficient train control and timetabling," *European Journal of Operational Research*, vol. 257, no. 2, pp. 355–376, 2017.

[16] M. Faieghi, A. Jalali, S. K. Mousavi Mashhadi, and D. Baleanu, "Passivity-based cruise control of high speed trains," *Journal of Vibration and Control*, vol. 24, no. 3, pp. 492–504, 2018.

[17] M. Faieghi, A. Jalali, and S. K. E. D. M. Mashhadi, "Robust adaptive cruise control of high speed trains," *ISA Transactions*, vol. 53, no. 2, pp. 533–541, 2014.

[18] T. Tao and H. Xu, "Adaptive fault-tolerant cruise control for a class of high-speed trains with unknown actuator failure and control input saturation," *Mathematical Problems in Engineering*, vol. 2014, Article ID 481315, 2014.

[19] L. Fan, "Iterative learning and adaptive fault-tolerant control with application to high-speed trains under unknown speed delays and control input saturations," *IET Control Theory & Applications*, vol. 8, no. 9, pp. 675–687, 2014.

[20] S. Li, L. Yang, Z. Gao, and K. Li, "Optimal Guaranteed Cost Cruise Control for High-Speed Train Movement," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2879–2887, 2016.

[21] H. Ye and R. Liu, "Nonlinear programming methods based on closed-form expressions for optimal train control," *Transportation Research Part C: Emerging Technologies*, vol. 82, pp. 102–123, 2017.

[22] Y. Zhao, T. Wang, and H. R. Karimi, "Distributed cruise control of high-speed trains," *Journal of The Franklin Institute*, vol. 354, no. 14, pp. 6044–6061, 2017.

[23] L. J. Zhang and X. T. Zhuan, "Optimal operation of heavy-haul trains equipped with electronically controlled pneumatic brake systems using model predictive control methodology," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 13–22, 2014.

[24] S. Li, B. De Schutter, L. Yang, and Z. Gao, "Robust Model Predictive Control for Train Regulation in Underground Railway Transportation," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 1075–1083, 2016.

[25] A. Bemporad, F. Borrelli, and M. Morari, "Model predictive control based on linear programming the explicit solution," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 47, no. 12, pp. 1974–1985, 2002.

[26] E. D. Sontag, "Nonlinear regulation: the piecewise linear approach," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 26, no. 2, pp. 346–358, 1981.

[27] C. E. García, D. M. Prett, and M. Morari, "Model predictive control: theory and practice—a survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.

[28] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, no. 3, pp. 407–427, 1999.

[29] X.-g. Li, D.-j. Gao, and Y.-h. Wang, "Predictive control for hybrid system based on mixed logic dynamic model," *Control and Decision*, vol. 17, no. 3, pp. 315–319, 2002.

[30] Y. Wang, B. De Schutter, T. J. J. van den Boom, and B. Ning, "Optimal trajectory planning for trains - A pseudospectral method and a mixed integer linear programming approach," *Transportation Research Part C: Emerging Technologies*, vol. 29, pp. 97–114, 2013.

[31] F. D. Torrisi and A. Bemporad, "HYSDEL—a tool for generating computational hybrid models for analysis and synthesis problems," *IEEE Transactions on Control Systems Technology*, vol. 12, no. 2, pp. 235–249, 2004.

[32] A. Bemporad, *Hybrid Toolbox-Usern's Guide*, 2003, Hybrid toolbox-userns guide.

[33] L. Li, W. Dong, Y. Ji, Z. Zhang, and L. Tong, "Minimal-energy driving strategy for high-speed electric train with hybrid system model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1642–1653, 2013.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 6164534, 18 pages
https://doi.org/10.1155/2018/6164534

Research Article

Application of Data Clustering to Railway Delay Pattern Recognition

Fabrizio Cerreto,[1] **Bo Friis Nielsen**,[2] **Otto Anker Nielsen**,[1] **and Steven S. Harrod**[1]

[1]*Department of Management Engineering, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*
[2]*Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark*

Correspondence should be addressed to Fabrizio Cerreto; facer@dtu.dk

Received 21 November 2017; Revised 20 February 2018; Accepted 11 March 2018; Published 29 April 2018

Academic Editor: Andrea D'Ariano

K-means clustering is employed to identify recurrent delay patterns on a high traffic railway line north of Copenhagen, Denmark. The clusters identify behavioral patterns in the very large ("big data") datasets generated automatically and continuously by the railway signal system. The results reveal the conditions where corrective actions are necessary, showing the cases where recurrent delay patterns take place. Delay profiles and delay change profiles are generated from timestamps to compare different train runs and to partition the set of observations into groups of similar elements. *K*-means clustering can identify and discriminate different patterns affecting the same stations, which is otherwise difficult in previous approaches based on visual inspection. Classical methods of univariate analysis do not reveal these patterns. The demonstrated methodology is scalable and can be applied to any system of transport.

1. Introduction

Operations analysis is the collection and review of performance data, such as punctuality and process cycle time. It is a key step in the continuous improvement of transport services, and several methods exist to collect and analyze data from operations. The increasing availability of automated data sources is offering new ways to analyze operations, providing deeper insight and more reliable information. Railway management is very accepting of these new possibilities, and considerable effort is made by operators and institutions to use operations analysis in feedback loops for improving the timetabling process [1–4]. A better understanding of the development of delays in railways, and in transportation in general, provides the opportunity to improve the processes and identify the factors affecting reliability. For example, causes of delays might be identified in misallocation of supplements and buffers in timetables, structural conflicts that require mitigation actions, suboptimal design of station processes, and inefficient procedures for preparing a train for departure. This paper demonstrates a data mining technique based on *k*-means clustering to identify recurrent delay patterns in transportation, identify the main reason for cluster membership, and provide managerial insight to improve timetables and processes.

Prior studies propose several methods that are currently in use for operation analysis, deploying sources of automatic data collection. These approaches can be divided into traditional statistical methods and big data techniques, which differ in both the use of data and in the output provided. Traditional methods tend to aggregate and summarize information, so these can provide a general picture or detailed information on specific stations or trains. These are typically proposed in the form of multiple univariate distribution analysis, where the occurrence of different delay patterns at the same station is not visible. Big data techniques can be used to investigate recurring patterns or internal structures in operations. These approaches are expanding, thanks to the growing availability of large amounts of data, and several techniques have been deployed to identify recurrences of delays and describe or predict delays. Advanced techniques such as neural networks, succession rules, Bayesian networks, and various methods of regression have been developed mainly to predict real time delays in railways, as described

in Section 2. However, train delays are necessarily correlated over the progression of a complete journey, and these data relations both along the journey of a train and among adjacent train paths have not received as much attention in the literature.

This paper presents a big data technique to identify recurring delay patterns in railway operations. Big data refer to information assets characterized by high volume, velocity, and variety, whose value is extrapolated by analytical methods [5]. In this application, the absolute delay and delay change are tracked for individual train paths along a railway line, resulting in absolute delay and delay change profiles. In the papers based on univariate statistics, systematic delays in these profiles are identified through visual inspection. The manual search for similarities suffers from subjective interpretation from the operator and is easily biased by common artefacts of the representation. The technique presented in this paper applies k-means clustering to find recurrent patterns in train delay progression, so that management may identify processes for improvement or correction. In this way it is possible to support continuous quality improvement.

In Section 2, a literature survey of contemporary data analysis methods is offered. Section 3 presents the k-means cluster method and the structure of the data to be studied. Section 4 presents results from the study of a high density Danish railway line. The effectiveness of k-means clustering for this application is discussed in Section 5, particularly with regard to its novelty compared to existing literature, while conclusions of this paper are presented in Section 6.

2. Literature Survey

Operations analysis is fundamental in the continuous improvement process to manage and modify railway operations. Data collected from real operations, or from simulation models, has been used in the feedback loop to design and improve railway timetables for decades. Typically, even if timetables may change over time, some of the fundamental infrastructure and service behaviors will not be modified. Timetables are often the result of only minor modifications to the previous editions and need to consider problems discovered in earlier timetables. For example, after a structural change in the Danish railway timetable in 1998, after the opening of the Great Belt fixed link, the service structure remained largely unchanged until 2016 [6].

Data collection systems have proliferated in railway networks since 2000, and very large amounts of data are available today. Widespread systems to collect data increased both volume and variety of data, which are often collected by different systems at the same time. The methods to elaborate and interpret information from past operations evolved together with the amount and quality of data, starting from descriptive and inferential statistic and moving towards big data techniques. For example, delay probability density functions can be extrapolated from historical data and integrated in analytical models to estimate service reliability before operation [7]. Goverde et al. [8] perform extensive statistical analysis and distribution fitting of data from the Dutch railway network. Goverde et al. fit different distributions for arrival and departure delays and find that no general distribution fits all groups of recorded arrival delays.

Primary delay distributions derived from operational data are also often employed as input in simulation models to evaluate the propagation of delays. Sipilä [9] explores the effect of modified running time supplements in railway schedules through microsimulation of a Swedish railway line. The author identifies different strategies for running time supplement allocation by verifying the significance of the change in punctuality recorded in 1600 simulations of selected scenarios. Lindfeldt [10] describes a method to aggregate delay data from real records and isolate distributions of primary delays. These distributions are then used to formulate microsimulation models. The data consists of manual records from dispatchers that assign a delay cause code to every record greater than 4 minutes of delay on the Swedish railways. In absence of other sources of data, the reliability of manual record cannot be validated, although the whole simulation model and its results rely on the derived distributions. Studies from other countries show that manual input can be indeed unreliable [11, 12]. The same method to extract primary delay distributions is later used by Lindfeldt and Sipilä [13] in a simulation model to assess the effect of allowing freight trains to travel outside their assigned path. The authors demonstrate that the realized travel times of freight trains could be shortened considerably without affecting the performance of other trains. The reduction of unnecessary waits for traffic management and the permission to depart before schedule reduced the average travel time on one side but increased its variability on the other.

Historical data also provides insight into the factors that influence service reliability. Olsson and Haugland [14] apply regression analysis on the Norwegian railway network and identify the most relevant factors for punctuality, such as absolute passenger flow and passenger occupation ratio. Gorman [15] uses regression analysis on data from American single-tracked freight railways to identify the factors that contribute the most to prolongation of railway running times. Gorman predicts congestion delay based on meets and passes scheduled as a consequence of speed heterogeneity. Again, in simulation, Shih et al. [16] apply an approach similar to Gorman's to determine the best capacity expansion strategy in terms of reduction of average prolongation of running time for freight trains. Shih et al. identify functional relationships, through regression of simulation results, between average delay per train-mile and several factors, such as the relative length of the doubled-tracked section of a railway line. Lindfeldt [17] applies multilinear regression with special focus on F-statistics to investigate factors generating delays on the Swedish railway network. Lindfeldt measures delay changes over selected routes and analyzes their distributions. In particular, the response variables are the share of trains with a delay increase, the median change in delay, and its standard deviation on the route. Statistically significant explanatory variables are found in the traffic volume for both passenger and freight trains. Among passenger trains the most significant variables are average speed and traffic

heterogeneity, and for freight trains it is the number of stations on the route with at least three tracks.

Timestamps and recorded deviations from schedule can be integrated with information from other sources. For example, incident reports may be compiled in case of larger disruptions. Such reports include information about the typology of the incident, the train affected by the primary delay, other trains involved, the secondary delays generated, and the recovery plans taken by the dispatchers. Schittenhelm and Richter [1] describe the reporting system in the Danish railways (the same system in service at the time of this study) and introduce a quantile-based approach to depict the development of train delays en route. The plots confirm the general understanding of delays from experienced operators and can be used to quantify the magnitude of expected disruption. The quantile-based approach, though, describes operations as a whole, and it is not able to distinguish systematic delays occurring at individual stations, but with different origins, so analysis of individual train services is necessary to identify peculiar delay patterns. Richter [18] introduces new metrics to identify improvement actions, based on data from automatic detection systems. Richter sorts the trains according to recorded delay and identifies the worst one in a percentile approach, associated with recorded delay causes. A similar approach is adopted with regard to change in deviation, or delay jump, recorded on line sections, so that most critical geographical areas are identified. Lastly, Richter proposes a tabular representation of the median delay of individual trains recorded at station, sorted by scheduled time and geographical location. In this way, the analyst can identify which specific trains typically suffer from primary delays, also characterized by geographical location, and which trains are typically affected.

Similarly, Peterson [4] studies the on-time performance along the path of specific train services, using the rolling average delay of the last three timing points. Such on-time performance is plotted for all the repetitions of a specific train service over a time period and compared to the average, standard deviation, and 75th percentile. Peterson identifies empty areas in the pool of plotted delay profiles and interprets these as recurrent delay patterns given by discrete dispatching choices along the train path. Peterson also interprets recurrent increases or decreases of vehicle delay as segments of insufficient or excess running time supplement, respectively. Reliability of service is described by the standard deviation of recorded delays. Peterson used the mentioned measures in a feedback loop to redistribute the running time supplement in train paths according to the recorded performance.

Andersson et al. [19] assess the effectiveness of running time supplement in railway schedules from empirical data collected on a Swedish railway line. The study plots the recorded delays over the train itinerary overlapped with scheduled running time supplement and compares pairwise stacked plots from different railway services, stopping patterns, or directions. The identification of misallocation of running time supplement is based on visual search for recurrent delay patterns, and a few different dispatching tactics are identified in clusters of similar delay profiles. Andersson et al. highlight the existence of a threshold value of delay that triggers prioritization of other trains that are traveling on schedule. The observations are clustered in groups and show recurrent delay patterns, and the analysis is supported by detailed analysis of possible conflicts among individual train itineraries. Noticeably, the authors demonstrate that the measures of punctuality currently in use on the Swedish network hide the effects of running time supplement misallocation and delays developed en route. Even though the punctuality at the final destination is a measure of railway performance very common among railway operators, it does not express how trains increase or recover from delays along their journey. Schittenhelm [20] provides a sample of similar measuring approaches in the European railway industry. In a later study, Andersson et al. [21] underline the relevance of critical points for network robustness by plotting delay profiles and showing that the profiles cluster around critical points according to different dispatching strategies. Advanced clustering techniques may support the identification of different strategies to compute the effects on robustness.

Lastly, van Oort et al. [22] evaluate data collected automatically on public transport services with a combination of statistical methods and visual representation. The study represents delay data similarly to Peterson [4], Andersson et al. [19, 21], and Schittenhelm and Richter [1], plotting the recorded delay over individual repetitions of the same service path, and adds the plot of relevant delay percentiles over the stations. The shape of the percentile-based delay profiles highlights recurrent patterns in the deviation from schedule. The representative delay profiles appear different depending on the percentile they represent. Patterns found included the presence of typical early arrival at stations in bus services, followed by waiting time until the scheduled departure time, or recurrent delay drops or increases at specific stations. The delay plots are combined with the measured headway from the previous vehicle. While the delay plots would suggest allocating more running time supplements at systematic increases of delay, structural delays that cannot be compensated by timetable slack are highlighted in the plots of headways, where service unreliability corresponds to scattered recorded headways. A percentile approach was also presented by van Oort et al. to characterize and sort the stations according to performance, similarly to previous literature.

The statistical analyses presented above are suitable for general description of the system performance but lack specific insight into recurrent delay patterns that occur in operation and into the relationships between delays at different locations. The literature presented in this survey focuses on the univariate analysis of selected measures, such as delays at single stations. Traditional metrics common in the railway industry, such as punctuality, have also been found unrepresentative of the actual service reliability. The methods that include the multidimensional aspect of the problem mostly deal with delay profiles, the sequences of delays recorded on individual train itineraries. The quality of these analyses often relies on visual inspection of plotted data and the observer-operated search for matching delay profiles. This search lacks a standardized methodology and is influenced by

the plotting layout and the subjective interpretation, which is based on personal experience.

Big data techniques have arisen recently and seek to make use of the very large amount of information that is provided by automatic data collection systems, overcoming the mentioned issues of traditional methods. The term big data is rather broad and includes different techniques that serve specific purpose. The common characteristics of these techniques are volume, velocity, and variety, meaning large amounts of data, generated at high speed, possibly by different sources with different or no structure [5]. As opposed to standard statistical analyses, where hypotheses are formulated and tested, big data techniques search for internal structures directly in the data. Data generated by automatic sources typically fit in the big data criteria. In railways, several data mining techniques were developed in the last years, following different approaches and searching for different types of information. The interest in these techniques is rising, together with the increasing availability of structured data. Industrial applications of these techniques are spreading, and new approaches are being studied also among public institutions [2].

Event mining is a technique based on time sorted logs, where relations between different events are found based on their coincidences. Hansen et al. [23] combine an event mining tool and standard statistics to predict the actual running times of trains to the next station, given all the recorded current delays. Dependencies between pairs of events are found or "mined" in timed event graphs created from the timestamps of individual trains, which correspond to events of occupation and release of blocking sections. The process times between events are inspected by standard statistics, resulting in conditional probabilities of process times, given the recorded delays of all relevant trains in the system. Such a model, though, relies considerably on very detailed knowledge about the infrastructure and requires data which is not commonly available from railway infrastructure managers.

Goverde and Meng [11] use the same information source and similar technique to identify and analyze route conflicts and identify delay chains. Infrastructure data and operation data are integrated so that it is possible to identify a train that is occupying a blocking section linked to a signal at danger for another train. Delay trees are built and traced backwards to identify the primary causes, so individual delays can be classified automatically into primary and secondary, and the correct attribution of delay causes can be verified. Interestingly, the authors verify that more than half of the delay cause records were assigned wrongly by the dispatchers, stating that, in the Netherlands, this type of manual input is not reliable and objective enough to be deployed in data analysis.

Kecman and Goverde [24] extend the model to include nonlogged line sections, where it is not possible to distinguish delays due to signaling impositions and delays due to primary causes. Delay chains are also traced in less detailed data by Sørensen et al. [12]. Based on the time sequences at stations experiencing disturbed operations, the authors identify the trains generating the conflicts and the trains suffering from the conflicts. The analysis is used to identify primary delays, describe single days of operation, identify frequent trains originating, or subject to, delay chains, and identify point stations where most of the primary or secondary delays are generated. In a comparison with manually recorded delay causes, the study finds relevant inconsistencies with the primary delays traced in the delay chains, in accordance with Goverde and Meng [11]. The method described, though, is only valid for single track lines and does not identify multiple primary delays.

Cule et al. [25] introduce association rules to identify delays recurring often together and set up an episode mining framework to highlight frequent delay patterns from train timestamps at stations. However, association rules can highlight common recurrences but cannot explain relations of causality between two events, so primary and secondary delays cannot be distinguished. Similarly, Wallander and Mäkitalo [26] identify delay chains according to the manual delay cause records from the dispatchers and based on timestamps at stations with granularity of 1 minute. The succession rules used are very similar to association rules but consider the time dependencies, so that events taking place earlier can be assumed to be the cause of events happening later under the same circumstances. Trains are characterized by the number and magnitude of conflicts they generate so that improvement actions can be concentrated. Association rules have also been adopted to evaluate the effectiveness of delay prevention actions on Japanese suburban networks by Yabuki et al. [27]. Yabuki et al. compare the association among occurrence of delays of different trains, change in delays, extension of running and dwelling times, and realized headway in before/after scenario comparison. The downside of such models is that association rules can be set between binary variables, so the development of delays depicted does not include its magnitude. Further, the number of associations to be analyzed grows exponentially with the number of potential pairs of events, so the analyses must be limited to short time frames of operation.

Neural networks are a big data method that learns from historical records and uses the relations identified among variables to predict an output, given unseen values of the input variables. This technique is particularly suited to delay prediction and has been deployed in multiple studies. Neural networks look for dependencies in the data, as opposed to simulation models, which are based on interaction rules between objects defined initially. Malavasi and Ricci [28] use neural networks to predict the total experienced delay on a railway line, given its geometrical and technological characteristics, and its scheduled utilization over time. In comparison to simulation, Malavasi and Ricci find neural networks more robust against extreme-valued input, which implicates more likely case-overfitting with simulation. Kecman et al. [29] propose a Bayesian network delay prediction model. In this case, input includes the timetable and recorded delays at all stations. Each delay is assumed to depend only on direct connections in a timed event graph, meaning the recorded delay for the same train at a previous station and for the previous train at the same station. Conditional delay distributions are assumed to be Gaussian, and the parameters

are derived through recursive Generalized Linear Models. Chapuis [30] deploys the same assumed delay dependency in a neural network model, where input includes delay of the previous train and at the previous station, and distance to the next station. Such a model can predict the delay of a train at the next station. Independent of the actual infrastructure, this model is generic and can be applied at any station of the railway network. The downside of neural networks, though, is the risk of data overfitting, reducing the prediction capability, although this risk is lower in neural networks than in simulation models.

In response, Marković et al. [31] introduce Support Vector Regression (SVR) to establish a functional relationship between characteristics of the railway system and train delays. Train category, scheduled time, infrastructure, and share of journey completed are identified as most influencing factors to predict the train delay at one station. The authors show that SVR generalizes better than an artificial neural network, which seeks to minimize error of prediction in the historical dataset. Interestingly, the authors assume that the performance of delay prediction can be improved by grouping delays by magnitude, as factors generating smaller delays differ from factors that generate larger disturbances.

Kecman and Goverde [32] apply big data techniques to predict running and dwelling times from actual operation data, based on records from block sections occupations. The study uses random forests of tree-based models, to predict nonlinear relations between input variables and process times, with sufficient robustness to outliers in the data, lowered risk of overfitting, and focus on real time application. Running time predictors are calculated for every block section, and dwelling time predictors are calculated for every station platform. Among the interesting findings, the running times are longer if the headway to the preceding train is short, meaning that the succeeding trains tend to slow down to smoothen the trip and reduce the risk of encountering a yellow signal. Moreover, the authors find no evidence to support the hypothesis that trains run faster when delayed. All the trains were found to run at approximately the maximum performance in any condition. The authors suggest that, in case of insufficient prediction accuracy, new variables might be included in the model, such as platform shape for dwelling times.

Big data techniques focus mainly on the prediction of delays and running times or on the identification of delay chains and realized delay propagation among trains. New applications of these techniques would support the analysis of the realized development of delays along the path of individual train delays. As shown by statistical analysis and visual search for patterns presented by Schittenhelm and Richter [1, 18], Peterson [4], Andersson et al. [19, 21], and van Oort et al. [22], this type of data contains a great deal of information yet to be explored, which would provide insight into the effectiveness of running time supplements and into the presence of structural issues that generate delay in transport operation. In this paper we present a clustering technique to identify recurrent delay patterns among train services, based on readily available data, which leaves room for inference on the factors that generate specific delay patterns. The result shows that, within comparable train trajectories and stopping patterns, different train services accumulate delay at different stations and that recovery shapes differently according to the route direction. Inferences on the cluster composition show the most frequent service characteristics in each cluster. Such information could inform the allocation of correction measures to improve timetables. Tables 1 and 2 summarize the literature just reviewed.

3. Identification of Recurrent Delay Patterns Using Big Data Techniques

In this paper, a delay profile of a train run is defined as the set of recorded deviations throughout its path or a part of it, on a specific date. Note that deviation is reported as the time difference between a scheduled and a realized event, such as arrival, departure, or nonstop timing point. Even though delay is often used to refer to positive deviations, a delay profile can include null and negative values. A delay profile is a powerful representation of operation and the comparison of several delay profiles along the same service path allows the identification of recurrent delay patterns and such a representation method has already been presented in the literature [1, 4, 18, 19, 21, 22]. Delay change, also called delay jump, is the difference in deviation between two consecutive stations and represents the delay recovery or increase. Schittenhelm and Richter [1] use this measure to assess delay increases or time gains between stations, and Goverde and Meng [11] use it to identify delay chains in railway operation. We define a delay change profile of a train as the set of recorded delay changes along its path or a part of it.

A dataset of delay profiles consists of all the delay profiles recorded in a defined period, stacked together. Fields, or variables, of the dataset are the events at every station, whereas observations are individual train runs from a selected service. Such a dataset can refer to a specific train service or to several services following the same stopping pattern, so that the fields can be aggregated. The first case is intended for infrequent services, typically long-distance trains, where every single service may have its own characteristics in terms of planned demand, scheduled rolling stock, or the time of crossing congested nodes. Suburban and regional railway services are often scheduled in constant stopping patterns at high frequency and could, thus, be analyzed together, expecting characteristics of operation to be more homogenous across services. A dataset of delay change profiles is defined analogously to delay profile datasets, where the fields contain the change in deviation in place of the absolute deviation.

Previous researches presented on delay and delay change profiles interpret recurrent patterns by visual search for similarities [1, 4, 19, 21]. The systematic analysis of these two types of datasets through clustering algorithms allows the identification of patterns that are not necessarily visible or that could be wrongly associated by subjective interpretation.

Clustering techniques partition a dataset into a collection of groups of similar observations. In this study, clustering is used to partition the datasets of delay profiles and

Table 1: Review of previous uses of univariate statistics in railway operation analysis.

	Environment		Technique					Purpose
	Real operation	Simulation	Distribution fitting	Test significance	Regression analysis	Percentile sorting	Visual inspection	
Goverde et al. (2001)	X		X					Distributions of primary and secondary delays
Sipilä (2010)		X		X				Comparison running time supplement strategies
Lindfeldt (2010)	X	X	X					Distributions of primary delays from real operation for simulation
Olsson and Haugland (2004)	X				X			Factors affecting punctuality
Gorman (2009)	X				X			Factors that generate delays on single track lines
Lindfeldt (2010)	X				X			Factors that increase delays on line segments
Lindfeldt and Sipilä (2014)		X		X				Travel times with different operation models, with/without free freight operation
Shih et al. (2014)		X			X			Factors affecting average delay per train-mile
Schittenhelm and Richter (2009)	X					X	X	Visual inspection of quantile-based representation of deviations and change in deviation
Richter (2010)	X					X		Delay tabular representation and sorting train service performance
Peterson (2012)	X					X		Rolling average delay for specific train services
Andresson et al. (2011)	X						X	Assessment of effectiveness of running time supplements
Andresson et al. (2013)	X						X	Identification of critical points for robustness
van Oort et al. (2015)	X						X	Delay profiles, headway profiles

Table 2: Review of previous uses of big data techniques in railway operation analysis.

	Technique								Level of detail			
	Event mining	Association rules	Succession rules	Neural networks	Bayesian networks	Random forests	Support Vector Regression	Clustering	Track sections	Station	Input	Purpose
Hansen et al. (2010)	X								X		Current delays of al trains	Prediction of running time to next station
Goverde and Meng (2011) Kecman and Goverde (2012)	X								X		Timestamps	Delay chains, actual primary delay causes
Sørensen et al. (2017)	X									X	Timestamps	Delay chains on single track lines, actual primary delay causes
Cule et al. (2011)		X								X	Timestamps	Delay patterns
Wallander and Mäkitalo (2012)			X							X	Timestamps, delay causes from dispatchers	Delay chains
Yabuki et al. (2015)		X								X	Timestamps	Comparison of real scenarios
Malavasi and Ricci (2001)				X					X		Physical infrastructure and utilization ratio	Prediction of total realized delay on a network
Kecman et al. (2015)a					X					X	Current train delay, last delay at station	Delay prediction at next stations
Chapuis, (2017)				X						X	Current train delay, last delay at station, distance	Delay prediction at next stations
Marković et al. (2015)							X			X	Infrastructure and train journey characteristics	Delay prediction at next stations
Kecman et al. (2015)b						X			X		Current traffic condition, actual train position, delays of the day	Running time and dwelling time prediction
Cerreto et al. *(2018) (this paper)*								*X*		*X*	*Timestamps*	*Recurrent delay patterns across stations*

identify train services that are candidates for identification of common causality. Inference on common factors appearing in observations clustered together facilitates the assessment of delay patterns in association with specific characteristics of a transport service, such as time of the day (peak/off-peak), day of the week, or equipment used. The clustering process is realized through measures of similarity between elements in the same cluster and dissimilarity between elements from different clusters. Several methods and metrics are available to accomplish the task, suitable for different uses. Hierarchical algorithms proceed by splitting or merging observations recursively and are preferred when a nested structure is assumed in the clusters. In contrast, partitional algorithms do not impose a hierarchical structure and find all the clusters at the same time. K-means clustering is a partitional algorithm and was chosen due to its simplicity and frequent appearance in the literature [33].

K-means clustering is an iterative clustering process based on the identification of the mean element in each cluster. Every cluster is represented by its centroid, calculated as the average of the elements of the cluster, and every observation is assigned to the cluster corresponding to the closest centroid. Given a number k of initial centroids, the algorithm executes the following steps:

(1) Assigning every element to the cluster with the closest centroid

(2) Calculating the new centroids of all the clusters as the mean of the elements

(3) Repeating until convergence, which is met when no element changes cluster between consecutive iterations.

This simple method requires three user-specified parameters, which might be hard to determine beforehand: the distance metric, the number of clusters k, and the cluster initialization. Euclidean distance is often used to determine the difference between observations, but other metrics are available, such as the L_1 distance [34]. The number of clusters k is the most difficult parameter to estimate, as there is no perfect mathematical criterion. The parameter k is typically determined according to available knowledge about the data or interpreting and evaluating the meaning of several independent partitions realized for different values of k. The initial centroids might influence the resulting clusters, so the initialization is often chosen among several independent partitions that result from sampling k initial centroids among the observations. The influence of initialization, however, generally diminishes with the dimensionality of the dataset [33].

A substantial contribution to the simplicity of the method is given by the required structure of the data. Contrary to observer-operated search, clustering methods rely on the numerical relations between variable values recorded across single observation. It is, thus, unnecessary for the clustering algorithm to preprocess the data and sort the recorded delays for every train/observation. In the method proposed in this paper, k-means clustering is applied to observations of a multidimensional variable, whose size corresponds to the number of timing points of a fixed stopping pattern, where the fields contain the delays, or delay changes, respectively, recorded at the individual timing points. Every observation of this multidimensional variable is a vector and represents a single train run.

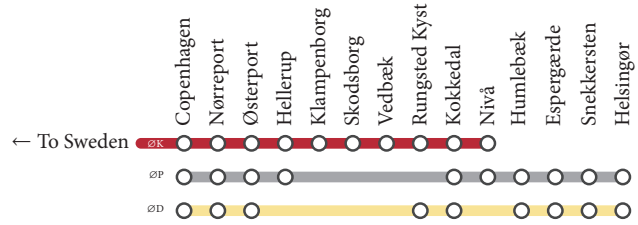


Figure 1: DSB services and stopping patterns on the Kystbanen.

4. Case Study: The Kystbanen, Copenhagen

The *Kystbanen* (coastline) is a double-tracked railway in the Copenhagen region. It is one of the busiest railway lines in the network of Banedanmark, the Danish infrastructure manager, and it is operated with regional standards, with some international services. It is operated nearly entirely by DSB, the largest Danish railway undertaking, which runs three different service types. The timetable is cyclic, and the services operate different stopping patterns during the day, as illustrated in Figure 1.

(i) The Øresund trains ("ØK") run all day every 20 minutes on a limited section of the coastline, between Copenhagen and Nivå. These trains operate between Denmark and Sweden across the Øresund bridge and stop at every station in Danish territory.

(ii) The regional trains ("ØP") run all day every 20 minutes as well, but they only operate in Denmark and run the whole coastline. These trains skip selected stops between Copenhagen and Nivå.

(iii) Additional trains are operated in the morning and afternoon peak hours. The rush hour trains ("ØD") operate every 20 minutes between Copenhagen and Helsingør, skipping other selected stops.

Fewer trains with modified stopping patterns run at night, so only weekday operation between 4:30 and 20:00 is considered in this study. The sections between Copenhagen and Østerport and between Snekkersten and Helsingør are shared with other services and operators.

In the resulting charts, stations are identified by a code specified by the infrastructure manager. Station codes and names are reported in Table 3.

Banedanmark provided a set of timestamps that state the scheduled and realized times of the trains at every timing point from April to December 2014. The records include information about the operation and about the timing points, such as station name, train ID, train category, scheduled time, and recorded deviation. Banedanmark relies on automatic train detection systems, based on the signaling system components. Typically, the track circuit boundaries do not correspond exactly to the platforms, and an offset

Table 3: Station codes and names on the Kystbanen.

Station code	Station name	Distance from KH [km]
KH	København H (Copenhagen central station)	0,0
KN	Nørreport	1,5
KK	Østerport	3,1
HL	Hellerup	7,8
KL	Klampenborg	13,3
SÅ	Skodsborg	18,8
VB	Vedbæk	22,1
RU	Rungsted Kyst	26,1
OK	Kokkedal	29,1
NI	Nivå	32,5
HUM	Humlebæk	36,3
GÆ	Espergærde	40,0
SQ	Snekkersten	42,7
HG	Helsingør	46,2

is generated between the time recorded by the automatic system and the actual time a train arrives at the platform or departs. This is a rather common problem, and it is also reported in the Netherlands [32] and Norway [12]. For the Danish network, a correction factor was calculated by Banedanmark using statistical analyses of GPS positions of train trajectories in collaboration with the main rail operator, DSB. The method and results are described by Richter et al. [35, 36]. Nørreport station is the only underground station on the line, so GPS correction is not available, which is visible as a sawtooth pattern common to all train services in the delay profiles presented below, with a slightly underestimated delay for arrival records at Nørreport and overestimated one for departure records from the same station. Similarly, delay change records are shifted to negative values for arrivals at Nørreport and at Østerport, whereas higher positive values are recorded for delay changes at departures from Nørreport. The bias is systematic and has the same exact effect on all the trains; therefore its influence on clustering can be neglected.

The train timestamps were rearranged by an automatic algorithm to create datasets as described in Section 3, by means of the commercial software SAS 9.4 TS Level 1M4, by SAS Institute Inc., Cary, NC, USA. Observations corresponded to a realized train on a given date, and the fields contained the recorded delay at every station. Data from every station was divided into arrival, departure, and pass-through times, where trains did not stop. Each record is the delay profile or the delay change of a train on a date and represents one observation of the given train. Every variable identifies the station code and the type of timestamp, which can be entrance to the station, I ("Indkørsel"), exit from the station, U ("Udkørsel"), or pass-through station, G ("Gennemkørsel"), which is used where trains do not stop.

The analysis is intended to report delay patterns. Consequently, punctual trains are discarded from the dataset. In Denmark, punctuality measurements are based on a delay threshold of 5 minutes for regional and long-distance trains, such as the Kystbanen. However, for internal management purposes, the infrastructure manager Banedanmark creates a delay report every time a train reaches at least 3 minutes of delay, containing information on the delay cause and on possible other trains hindered. Consequently, only trains with at least one recorded delay greater than or equal to 3 minutes are considered relevant in the present case study. Delay distributions are known to include large shares of trains with short delays, with decreasing frequency for larger delays [7, 8]. Largely unbalanced clusters are a known issue in clustering algorithms and are an object of study to reduce the interference of large clusters [37]. In this case, punctual trains can therefore be considered as a compact cluster derived by prior knowledge, and they can be filtered out from the cluster analysis. The operation of filtering can be considered noise reduction and improves the quality of clustering, as the k-means procedure tends to generate spherical clusters of same radius [38]. According to Marković et al. [31], large delays are influenced by different factors other than smaller delays, which further supports the filtering choice. However, in different contexts, the filtering threshold might be set equal to a different value or not be applied at all.

Given the characteristic high frequency of train services on this line, clustering was operated by stopping patterns rather than by train numbers, so trains were grouped together by direction and service category. Grouping trains with similar characteristics and same stopping patterns increases data availability in the comparison and does not disqualify the result. In fact, such grouping was already proposed by Schittenhelm and Richter [1].

As explained in Section 3, k-means clustering requires choosing the number of clusters k in advance. To set the number of clusters, the k-means algorithm was repeated with different values of k, and the best result was selected using criteria from Jain [33]. The number of clusters k should be large enough to represent different patterns. At the same time, as k increases, the same delay patterns tend to split into more clusters, and k should remain small enough to prevent the generation of duplicate clusters. In detail, for every combination of train category, direction, and clustering variable (delay or delay change), k was set as the highest integer that did not generate duplicate clusters. That is, the univariate distributions of delays, or delay changes, in every cluster should be different from all the other clusters for at least one station. Since k is selected independently for all the mentioned cases, the same set of trains might best be represented by a different number of clusters when the algorithm operates on the delay variables or on the delay change variables. The $L1$ distance was used as a clustering metric between observations, as suggested by Kashima et al. [34].

K-means clustering was performed on the described dataset by the commercial software MATLAB R2017a, by The MathWorks, Inc. Selected results of the method application are reported in Section 4.1, clustering on either delay profiles or delay change profiles.

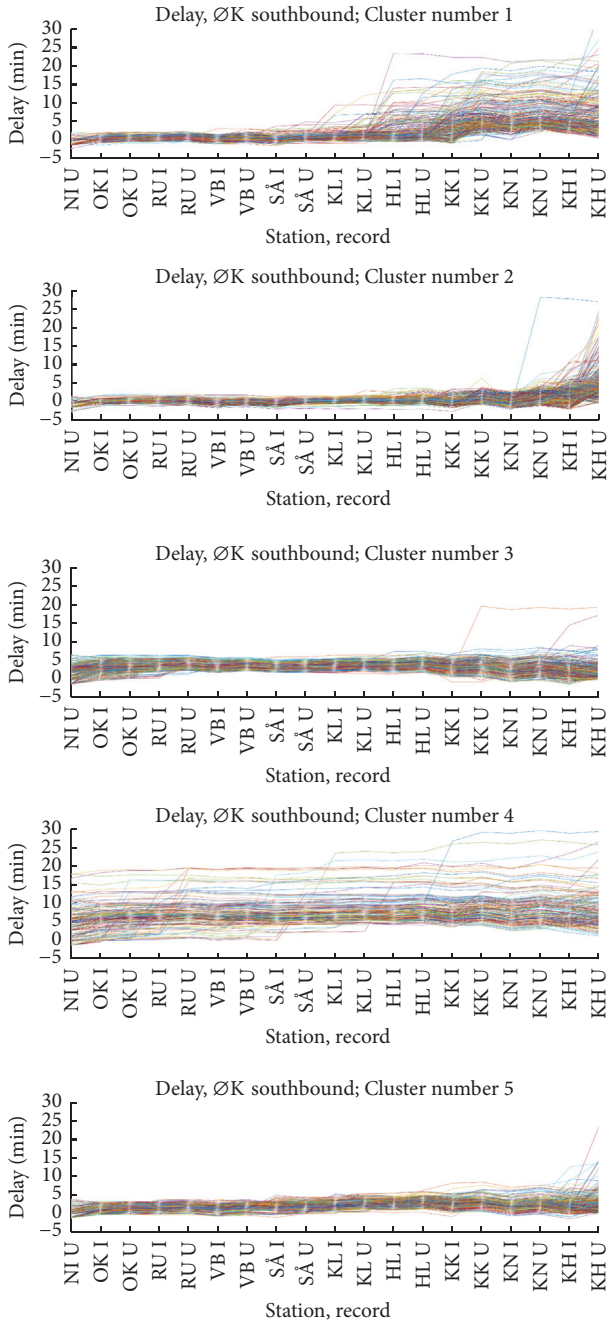


Figure 2: Resulting clusters in southbound ØK trains, Nivå–Copenhagen.

4.1. Clustering Results. Figure 2 illustrates the effectiveness of delay profiles clustering on ØK southbound trains, on the delay variables. Note, after a stop at Copenhagen central station, these trains proceed to Sweden. The charts show that similar delay profiles are grouped together with low variance around the average centroid of each cluster, highlighting recurrent patterns. The resulting clusters can be interpreted as follows:

(1) Cluster 1: it includes trains that are punctual on the first section of the line but suffer delays approaching the most congested area of Copenhagen, mainly from Klampenborg and from Østerport.

(2) Cluster 2: it includes trains that are punctual throughout the complete journey, which receive delays leaving from Copenhagen.

(3) Cluster 3: it includes trains that are nearly punctual or anyway within 5 minutes of delay through the complete journey and across Copenhagen central station; from Hellerup, a marginal delay recovery is visible for these trains.

(4) Cluster 4: it includes the most delayed trains, being delayed throughout the whole itinerary or it largest part.

(5) Cluster 5: it includes punctual trains with slightly, but steadily, increasing delay across stations.

Some clusters present outliers, such as Clusters 2 and 3. Even though some delay profiles may appear considerably different from other profiles in the same clusters, these observations were assigned to the cluster with the closest centroid. This means that, in selected cases, the delay profiles are the representation of rather unique events, which may be neglected after more detailed analysis in the composition of the individual clusters.

Individual clusters are characterized through the mean values of the aforementioned measures. The following measures were computed for each train run to characterize the individual clusters:

(i) Average, minimum, and maximum delay across stations

(ii) Range of delays across stations

(iii) Standard deviation of delays recorded across stations

(iv) Initial delay, the delay at first station

(v) Final delay, the delay at the last station

(vi) Overall delay change, difference between final and initial delay—positive values mean the delay has increased from first to last station

(vii) Maximum delay change across stations.

Cluster characteristics are summarized in Table 4.

4.2. Comparison with Percentile-Based Approaches on Delay Profiles. In this section, a comparison is provided between the pooled data and the clusters on the dataset of delay profiles. The same percentile representation of delay profiles is shown, as proposed by Schittenhelm and Richter [1], Peterson [4], and van Oort et al. [22]. These authors represented different percentiles. For the sake of clarity, only the 15th, 50th, and 85th percentiles and the average are displayed in the following diagrams.

Figure 3 shows the distribution of delays of the entire dataset of ØK southbound trains. The only pattern visible is a slight increase in delay towards Copenhagen, more evident for the more delayed trains, represented by the 85th percentile. Even though a large portion of punctual trains

Table 4: Characterization of delay profile clusters and southbound ØK trains, Nivå–Copenhagen.

Cluster	Number of observations	Mean average delay [min]	Mean STD of delays [min]	Mean initial delay [min]	Mean final delay [min]	Mean min delay [min]	Mean max delay [min]	Mean delay range [min]	Mean max delay change [min]	Mean overall delay change [min]
(1)	270	2,26	2,78	−0,95	6,14	−1,06	7,72	8,78	4,88	7,09
(2)	418	0,55	1,47	−1,05	4,71	−1,24	5,27	6,52	4,69	5,76
(3)	381	3,09	1,12	1,70	1,80	0,53	4,64	4,11	2,69	0,11
(4)	159	7,65	1,92	4,59	8,03	3,73	10,21	6,47	6,79	3,44
(5)	395	1,92	1,14	−0,28	2,23	−0,47	4,10	4,57	2,25	2,51
Total	1623	2,46	1,57	0,35	3,99	−0,12	5,73	5,85	3,87	3,64

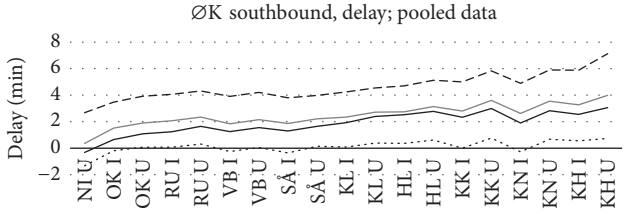


Figure 3: Delays recorded for ØK southbound trains. 15th percentile: dotted; median: solid black; 85th percentile: dashed; average: solid gray.

was discarded from the dataset, the residual distribution of delays remains positively skewed, as shown by the average constantly higher than the median value.

The new information revealed by the clustering algorithm is provided in Figure 4. In this figure, the individual internal distributions of delays are compared to the pooled delay distribution from Figure 3. Figure 4 shows, for each cluster, the difference between the cluster statistic at each station and the equivalent pooled statistic from Figure 3.

In Figure 4, the 15th and 85th percentiles and the median line of the internal cluster delay profiles distributions are compared to the distribution of pooled delay profiles. The clusters where the difference of 85th percentile from the pooled dataset is lower than the difference of the 15th percentile have tighter distributions of delay profiles compared to the pooled dataset, increasing the significance of the identified pattern. The local deviation present in the clusters represents the information hidden in the pooled dataset, which is instead brought to light by the clustering.

4.3. Comparison with Percentile-Based Approaches on Delay Change Profiles. In this section, a comparison is provided between the pooled data and the clusters on the dataset of delay change profiles. The same representation of delay change profiles based on the median is shown, as proposed by Schittenhelm and Richter [1, 18], supplemented with the average, that is, the cluster centroid.

Figure 5 shows the delay change profiles of the entire dataset of ØD northbound trains. A generalized positive delay change is visible at the last station. The large changes in delay from location KN I to KK I are linked to the known deviation in the timestamps at Nørreport.

The differences between the pooled median and average delay change profile and the same profiles from individual clusters are represented in Figure 6. In this case, the information gained by clustering is more evident. All the clusters remain similar to the pooled data at most stations, except few stations, where a large difference is recorded in the delay change.

Every cluster is characterized by at least one larger delay change at one station, which would be hidden in the pooled distribution of delay change profiles. Noticeably, the negative effect of different delay patterns overlapping is evident for *KN I* records. All the clusters deviate negatively from the pooled data by around 0,5 minutes, except Cluster 2, which deviates positively by around 1,5 minutes from the pooled profile. This means that the pooled profile was shifted by one single cluster to a central value, hiding both the frequent delay recovery and the specific delay increase from Cluster 2.

4.4. Inference on the Clusters. In this section, results from clustering of delay profiles and delay change profiles are investigated to identify relations with cluster characteristics, using heuristic classification. For the sake of conciseness, only cluster centroids are reported in Figures 7 and 10, and only a sample of the results is reported, which is ØD northbound trains and ØK southbound trains. Figure 7 shows results from clustering delay change profiles for ØD trains to Helsingør.

Clusters can be interpreted as follows:

(1) Cluster 1: regular delay increases at the last three stations, where trains become unpunctual.

(2) Cluster 2: delay increase arriving at the first stop, Nørreport.

(3) Cluster 3: trains are considerably delayed arriving at the final destination, Helsingør.

(4) Cluster 4: trains are without remarkable delay changes: these trains tend to keep the same delay throughout the whole journey.

(5) Cluster 5: specific delay increases at Humlebæk arrival; trains in this cluster show also smaller recovery at Skodsborg arrival, compared to other clusters.

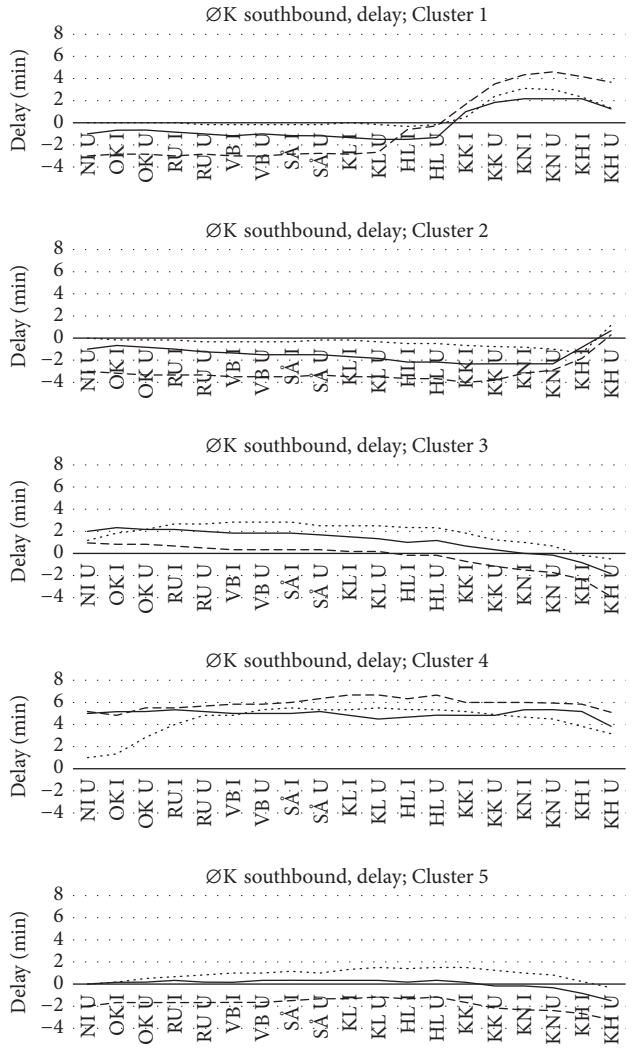


Figure 4: Differences in the distributions of delays recorded for ØK southbound trains. Each cluster's internal distribution is compared to the pooled distribution. 15th percentile: dotted; median: solid; and 85th percentile: dashed.

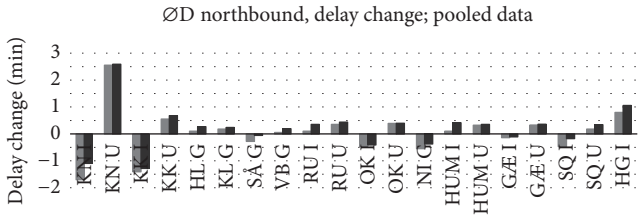


Figure 5: Delay changes recorded for ØD northbound trains. Median in bright shade and average in dark shade.

(6) Cluster 6: these trains accumulate delays passing the stations of Hellerup and Klampenborg; on the other side, compared to other clusters, the average delay increase at final destination Helsingør is smaller.

Inference on the cluster population shows that some patterns are specific for selected train services, identified by their

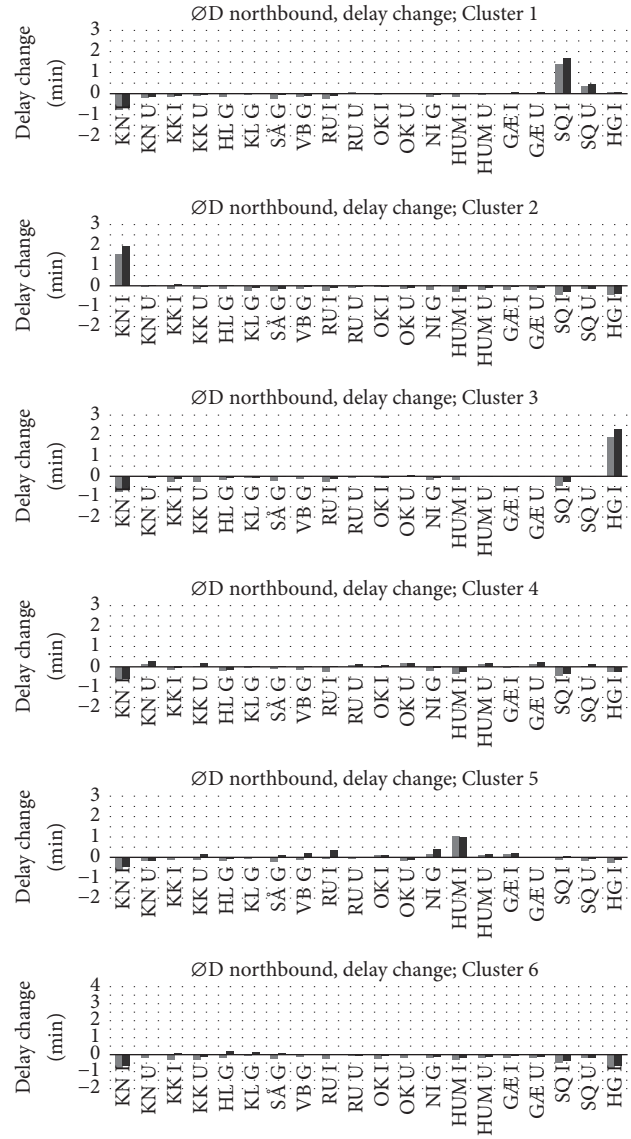


Figure 6: Delay changes recorded for ØD northbound train, by clusters. Median in bright shade and average in dark shade.

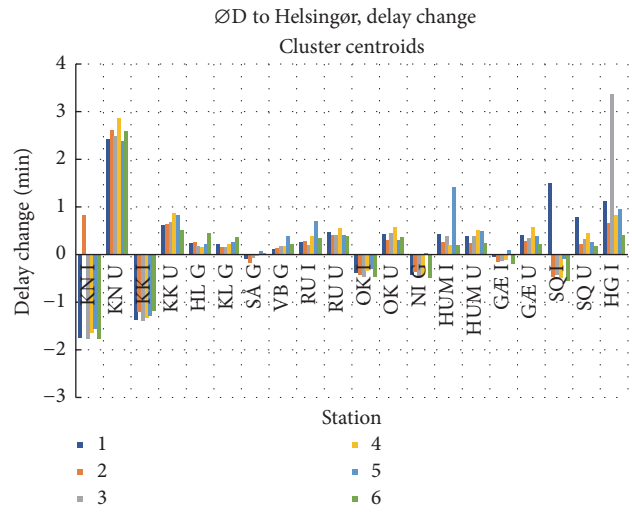


Figure 7: Cluster centroids for northbound ØD trains: delay change.

Time band	Departure time from KH	Train number	Cluster 2	6	4	1	3	5
2 AM Peak	06:18	4413	18%	25%	14%	14%	18%	11%
	06:38	4415	4%	36%	4%	16%	32%	8%
	06:58	4417	26%	33%	7%	19%	4%	11%
	07:18	4419	6%	22%	8%	31%	8%	25%
4 PM Peak	15:18	4467	21%	25%	17%	4%	21%	13%
	15:38	4469	19%	30%	7%	12%	23%	9%
	15:58	4471	44%	16%	16%	4%	8%	12%
	16:18	4473	5%	15%	28%	18%	20%	15%
	16:38	4475	43%	13%	21%	13%	10%	2%
	16:58	4477	20%	15%	39%	9%	2%	15%
	17:18	4479	16%	32%	15%	12%	9%	16%
	17:38	4481	31%	14%	19%	19%	5%	12%
	17:58	4483	46%	14%	14%	6%	10%	10%
		Cluster share	*24%*	*22%*	*17%*	*14%*	*12%*	*12%*

Figure 8: Northbound ØD trains. Cluster share by train service ID. The color code compares the individual row's distributions among clusters to the overall distribution among clusters reported in the last row. Clusters sorted by size.

Time	Type	Time band	Cluster 2	6	4	1	3	5
6:20–8:20	Peak AM	2	13%	28%	9%	21%	15%	15%
15:20–18:00	Peak PM	4	27%	20%	19%	12%	11%	11%
		Cluster share	*24%*	*22%*	*17%*	*14%*	*12%*	*12%*

Figure 9: Northbound ØD trains. Cluster share by time band. The color code compares the individual row's distributions among clusters to the overall distribution among clusters reported in the last row. Clusters sorted by size.

train number. Figure 8 shows how every train service ID is spread across clusters. In each column, the shade represents the difference between individual percentages and the cluster share, where the brightest colors are associated with the values furthest from the cluster share. Green is positive difference, that is, larger percentages than the cluster share; red is negative difference, that is, smaller percentages than the cluster share.

Delay change profiles in Clusters 1 and 5 represent typical behavior of service 4419, whereas Cluster 2 is considerably more frequent in services 4471, 4473, and 4483. Cluster 3 is more common among services 4415, three times more frequent than the whole population distribution across clusters, 4467, 4469, and 4473, which double the average frequencies. Cluster 4 is typical for service 4477, and, lastly, Cluster 6 represents a large share of services 4417 and, again, 4415. Further investigation on other factors may reveal the causes that rule the train services' cluster membership.

The analysis of Figure 8 shows the existence of a relation between train IDs in a specific time band and cluster membership. This is shown in detail in Figure 9, where cluster membership is aggregated in time bands. The same color coding as Figure 8 is applied.

The timetable is divided into time bands according to the overall service frequency on the line, so that time bands 2 and 4 are the AM and PM peak periods, respectively, when 9 trains/h per direction are operated. Time bands 1, 3, and 5 are the remaining off-peak periods, when ØD trains are not operated, so only 6 trains/h occupy the line in each direction, allowing for larger headway buffers between trains. At the same time, smaller congestion is expected, in off-peak periods, both on the train traffic and on the number of passengers to board or alight at the stations.

In this case, morning peak shows recurrent delay patterns presented by Clusters 1 and 6, whereas patterns represented by Clusters 2 and 4 are rare in this time band. Oppositely, the distribution of trains in the PM peak hour is similar to the overall distribution.

Further inference on the clusters of ØD northbound trains might highlight interferences from other trains. Lokaltog trains run mostly on a network independent of Banedanmark's and share with ØD and ØP trains the line section between Snekkersten and Helsingør. ØD northbound trains are scheduled at a short headway after Lokaltog trains from Snekkersten to Helsingør. The analysis of timestamps from Lokaltog trains on this section and of the realized headways between Lokaltog and ØD northbound trains might suggest that Clusters 1 and 3, which increase the delay near Helsingør, are actually the result of delay propagation from Lokaltog trains to ØD trains.

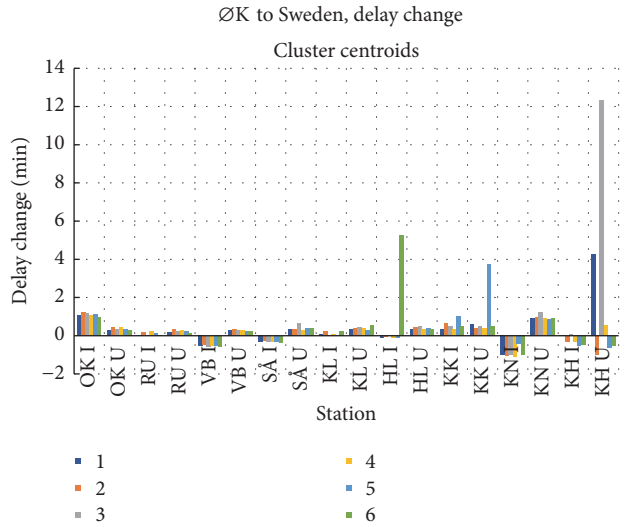


Figure 10: Cluster centroids for southbound ØK trains: delay change.

The clustering results from other service categories, with different stopping patterns, can be related to the time periods of the day. For example, ØK southbound trains are reported in Figures 10 and 11.

Figure 10 represents the centroids of resulting clusters in train category ØK southbound, according to delay change. Besides, the distribution of trains across clusters is summarized in Figure 11, disaggregated by time bands, highlighted in the same color code as Figures 8 and 9. Note that the number of clusters in the delay change profiles of ØK southbound trains is different from the number of clusters in delay profiles. This is not necessarily inconsistent, as the two variables express different aspects of the development of delays. In this case, the cluster share by time band explains the nature of two clusters. In particular, Cluster 1 is considerably more frequent in the PM peak hour, whereas Cluster 2 is more typical of trains in the AM peak hour. This result can be reasonably interpreted as delays generated by passenger congestion. In fact, delay increases in the PM peak hour appear at departures from Copenhagen, where a large number of passengers leave towards Sweden. On the contrary, Cluster 2 represents delays increases collected across stations towards Copenhagen and a delay recovery departing from Copenhagen, where fewer passengers are expected to board. The cluster share for Clusters 3 to 6 is comparable with the overall distribution across different time bands, so these delay patterns cannot directly be associated with time of the day. Further research may reveal factors that rule the cluster membership for these clusters.

More disaggregated analysis of cluster composition according to train number, or service ID, is in accordance with aggregated time bands. This is valuable especially for time band 3, which is the most populated time band according to the timetable. Figure 12 shows that even if the overall distribution of trains in time band 3 across clusters is very similar to the overall distribution, specific train services present different typical delay patterns. In this case, further analysis of train service characteristics should indicate a better disaggregation of train services in a specific time band. The same color code as in tables in Figures 8, 9, and 11 is applied in Figure 12.

Even though recurrent patterns are also clear in the delay profiles dataset, the results could not be explained by the available variables. Further research might identify relations that guide the clustering of delay profiles on this line, such as realized headways, weather conditions, passenger counts, and recorded delay causes.

5. Discussion

The clustering method proposed in this paper finds it strengths in being automatic, unbiased, flexible, and simple. A comparison to methods presented in the literature is provided in this section. Previous approaches [1, 4, 18, 19, 21, 22] extracted information from delay profiles mainly through observation, occasionally combined with multiple univariate statistical analyses and observation ranking. In most studies, the complete dataset was plotted in the form of delay profiles, and the identification of frequent patterns among the observations relied on the observer's ability. Visual inspection is typically affected by subjective interpretation, which can differ across analysts, and suffers from low effectiveness of naked eye to average data represented graphically. In some studies, supporting measures were plotted with the full dataset, such as average profile, median, and selected percentiles to represent the distributions.

The application of these measures as multiple univariate distributions, though, does not catch the interdependencies of delays at different stations and does not provide information about the development of delays along the train journey. The method proposed in this paper allows automatic identification of delay patterns, removing, thus, the influence of subjective interpretation of delay profiles. Furthermore, profile clustering allows the identification of similar delay profiles in the entire pool of records. Note that even though the clustered delay profiles were plotted in this paper, the observation of the profiles did not play a role in the identification of similarities. This exact process is indeed performed by the clustering algorithm, and the results are then plotted for an easier comprehension of the development of delays in the individual clusters. The metrics provided as 15th, 50th, and 85th percentile would be sufficient to describe the distributions within individual clusters and might be used in replacement of the cluster plots.

Compared to big data techniques proposed in the literature for other purposes in analysis of transport operation [11, 23, 24, 32], this method relies on readily available data and does not need detailed knowledge on the infrastructure and occupation of individual blocking sections. It can therefore be scaled to different levels of detail or transferred to other modes of transportation where delay can be measured at fixed points on a given path, such as bus networks or air traffic. It is a very common practice of transport operators to provide live data on delays recorded on their own network, which can be recorded accessing public websites.

Start time	Type	Time band	Cluster 2	4	1	5	3	6
04:30	Off-peak	1	35%	30%	15%	12%	3%	5%
06:20	Peak AM	2	49%	26%	10%	11%	1%	3%
08:20	Off-peak	3	39%	29%	13%	13%	3%	3%
15:20	Peak PM	4	30%	26%	24%	9%	9%	2%
18:00	Off-peak	5	40%	29%	18%	9%	3%	1%
		Cluster share	*40%*	*28%*	*15%*	*12%*	*3%*	*2%*

FIGURE 11: Southbound ØK trains. Cluster share by time band. The color code compares the individual row's distributions among clusters to the overall distribution among clusters reported in the last row. Clusters sorted by size.

Time band	Departure time from NI	Train number	Cluster 2	4	1	5	3	6
1	06:01	1314	38%	20%	23%	13%	3%	5%
	06:21	1316	32%	44%	8%	8%	4%	4%
	06:41	1318	33%	33%	10%	14%	5%	5%
2	07:01	1320	59%	12%	12%	15%	0%	2%
	07:21	1322	50%	29%	13%	4%	0%	4%
	07:41	1324	51%	31%	10%	5%	0%	3%
	08:01	1326	41%	32%	14%	11%	2%	2%
	08:21	1328	53%	18%	0%	22%	2%	5%
	08:41	1330	42%	33%	14%	3%	3%	6%
3	09:01	1332	43%	30%	4%	20%	0%	4%
	09:21	1334	39%	32%	16%	13%	0%	0%
	09:41	1336	41%	24%	15%	17%	0%	2%
	10:01	1338	24%	24%	27%	16%	5%	3%
	10:21	1340	31%	39%	19%	3%	0%	8%
	10:41	1342	57%	14%	11%	14%	0%	5%
	11:01	1344	45%	24%	12%	15%	0%	3%
	11:21	1346	40%	47%	10%	3%	0%	0%
	11:41	1348	53%	35%	7%	5%	0%	0%
	12:01	1350	26%	29%	21%	15%	3%	6%
	12:21	1352	47%	32%	8%	8%	0%	5%
	12:41	1354	38%	16%	28%	16%	0%	3%
	13:01	1356	46%	19%	15%	8%	8%	4%
	13:21	1358	50%	31%	15%	0%	4%	0%
	13:41	1360	59%	24%	7%	3%	7%	0%
	14:01	1362	29%	34%	17%	17%	3%	0%
	14:21	1364	44%	32%	12%	4%	8%	0%
	14:41	1366	37%	37%	10%	15%	2%	0%
	15:01	1368	24%	24%	16%	30%	4%	1%
	15:21	1370	43%	25%	6%	14%	10%	2%
	15:41	1372	35%	41%	5%	11%	5%	3%
	16:01	1374	36%	21%	17%	17%	8%	2%
	16:21	1376	36%	33%	14%	6%	6%	6%
4	16:41	1378	37%	27%	22%	7%	5%	2%
	17:01	1380	19%	29%	29%	13%	6%	3%
	17:21	1382	39%	18%	18%	5%	18%	0%
5	17:41	1384	45%	29%	23%	3%	0%	0%
	18:01	1386	18%	43%	30%	8%	3%	0%
	18:21	1388	47%	28%	16%	9%	0%	0%
	18:41	1390	40%	24%	16%	12%	4%	4%
	19:01	1392	58%	15%	13%	15%	0%	0%
	19:21	1394	34%	31%	17%	10%	7%	0%
	19:41	1396	48%	26%	7%	7%	11%	0%
	20:01	1398	31%	38%	22%	6%	0%	3%
		Cluster share	*40%*	*28%*	*15%*	*12%*	*3%*	*2%*

FIGURE 12: Southbound ØK trains. Cluster share by service ID. The color code compares the individual row's distributions among clusters to the overall distribution among clusters reported in the last row. Clusters sorted by size.

Furthermore, the partition of operation into recurrent delay patterns allows inference on individual clusters, which is not possible with association or succession rules [25–27]. These methods do not provide causality connection and can only be used to compare scenarios, for example, before and after delay mitigation countermeasures have been implemented. Results from clustering can be inferred with other mining techniques to identify further connections between specific system factors and delay membership, so that the causes of specific delays can be identified, and the effects of corrective actions can be estimated beforehand.

Alongside flexibility, the strength of this method resides in its simplicity. Unsupervised learning methods, such as clustering, aim at the identification of internal structures of the system. Supervised learning methods, in contrast, attempt to predict results, based on assumed connections in the input. For these reasons, neural networks [28, 30], Bayesian networks [29], and supporting vector regression methods [31] require initial assumptions on the factors that have direct effect on the desired output, which can be cumbersome to identify and could be hidden. The clustering method proposed here does not require initial assumptions, so any recurrent delay pattern can be identified. In particular, the k-means algorithm was selected, being the most common algorithm for partitional clustering. Even though several clustering methods and algorithms exist in the literature, none of them is clearly preferred to the others [33]. It is important to stress the fact that the output of clustering algorithms only suggests hypotheses and that the interpretation of results plays a more relevant role than seeking the best clustering principle. However, further research might improve the method. For example, a different choice of the clustering statistic between observations might be explored. In addition, the choice of the parameter k might be supported by advanced techniques and metrics. In this paper, k was set through statistical analysis of the associated clusters, but further studies might reveal more efficient methods integrated in the clustering algorithm itself. Lastly, the clustering results might depend on the punctuality threshold selected to filter out punctual trains, if applied.

The relations found in inference from resulting clusters can, eventually, be considered and implemented in the mentioned supervised data mining methods. The use of other sources of information can be further investigated, for example, the rolling stock equipment deployed or information on delay causes collected by train dispatchers. The clustering algorithm itself cannot provide information on the causes of delays, but relevant relationships with external variables might be found through the inference on the clusters. The implementation of information recorded by the dispatchers on primary and secondary delays could support the identification of delay propagation. However, previous studies in Europe highlighted the unreliability of such manually recorded data [11, 12]. These procedures are different for each infrastructure manager and should comply with different national regulations. This input should be analyzed in detail before being implemented in the inference on clusters. The timestamps might be integrated with data from other railway undertakings, so that the realized headways could be investigated and included in the cluster inferences. The effects of delay propagation might be thus investigated and the dispatching strategies possibly improved. Passenger counts or boarding/alighting timings could also reveal that specific localized delay increases are linked to passenger exchange and might suggest modifications in the scheduled dwelling times. Useful information from the railway undertakings might include differences between planned and realized train compositions or the use of energy saving strategies. Driving support systems are spreading among train operators to reduce the energy consumption and thus the operating cost, especially for diesel-powered railways. The effects of such systematic patterns in the driving style are, in any event, expected to emerge in the clustering algorithm, especially with more detailed data in the positioning. Further development of this method might expand its application to other industrial processes or other transportation modes. The service timekeeping could be measured at designated checkpoints, to build standard delay profiles and delay change profiles.

6. Conclusions

In this paper, a new method is presented to analyze railway operations, based on big data techniques. Previous studies highlighted the need for tools to analyze railway operation based on data from automatic data collection sources and to automatically detect delay patterns [1]. K-means clustering is here applied to train delay records from automatic train detection systems to identify systematic delays, rearranged in delay profiles and delay change profiles. This method is automatic, unbiased, flexible, and simple.

Both institutions and industry are showing great interest in big data applications [2]. The method described in this paper provides a managerial tool to identify recurrent delay patterns that affect the service reliability. A localized analysis with additional information supports the identification of the causes of individual patterns, so that specific countermeasures can be designed. For example, dispatching strategies might be modified when a structural conflict is detected; the boarding and alighting process might be improved at stations where delay increases recurrently. If the causes of recurrent delays are identified in frequent conflicts, small modifications to the timetable slack might be a solution to reduce delay propagation.

The effectiveness of this approach is demonstrated in an application on a Danish regional railway line. The application shows that it is possible to identify systematic delays at specific stations in a congested area and to identify different delay patterns. Furthermore, delay patterns can be conveniently associated with specific time periods of the day, showing time dependency, reasonably explained by the prevailing passenger flow direction. Specific delay patterns are demonstrated to be characteristic of individual train service IDs, which could depend on other service characteristics, such as structural conflicts with other trains in specific sections of the line, use of specific rolling stock equipment, or connections to other transport services. The implementation of other sources of information might improve the inference on the clusters, such

as weather conditions, passenger counts, information from the dispatchers, or rolling stock characteristics.

Further development of this method might improve the selection of the number of clusters, identify new clustering metrics between observations, or integrate additional sources of information to improve the inference on clusters.

Conflicts of Interest

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

Acknowledgments

This work was funded by a Dean Grant from the Technical University of Denmark (DTU) and by the Danish Innovation Fund through the IPTOP (Integrated Public Transport Optimisation and Planning) project.

References

[1] B. Schittenhelm and T. Richter, "Railway Timetabling Based on Systematic Follow-up on Realized Railway Operations," in *Proceedings of the Proceedings from the Annual Transport Conference at Aalborg University*, H. Lahrmann, Ed., Aalborg: Traffic Research Group, Aalborg University, 2009.

[2] A. D'agostino, *Big Data in Railways - Common Occurrence Reporting Programme*, 2016, Big Data in Railways - Common Occurrence Reporting Programme.

[3] T. Richter, "En bedre jernbane gennem højere datakvalitet," in *Trafikdage på Aalborgs Universitet. Aalborg: Traffic Research Group*, H. Lahrmann, Ed., Aalborg University, 2008.

[4] A. Peterson, "Towards a robust traffic timetable for the swedish southern mainline," in *Proceedings of the Computers in Railways XIII*, pp. 473–484, WIT Transactions on The Built Environment, New Forest, UK.

[5] A. De Mauro, M. Greco, and M. Grimaldi, "A formal definition of Big Data based on its essential features," *Library Review*, vol. 65, no. 3, pp. 122–135, 2016.

[6] J. Hansen, "Future Railway Development and Performance," in *Proceedings of the Danish Rail Conference*, The Danish Rail Sector Association, Copenhagen, 2015.

[7] M. Carey, "Ex ante heuristic measures of schedule reliability," *Transportation Research Part B: Methodological*, vol. 33, no. 7, pp. 473–494, 1999.

[8] R. M. P. Goverde, G. Hooghiemstra, and H. P. Lopuhaä, *Statistical Analysis of Train Traffic: The Eindhoven Case*, IOS Press, Incorporated, 2001.

[9] H. Sipilä, "Simulation of modified timetables for high speed trains Stockholm – Göteborg," in *Proceedings of the First International Conference on Road and Rail Infrastructure - CETRA2010*, S. Lakušić, Ed., p. 1, University of Zagreb, Opatija, Croatia, 2010.

[10] O. Lindfeldt, "Evaluation of punctuality on a heavily utilised railway line with mixed traffic," in *Proceedings of the 11th International Conference on Computer System Design and Operation in the Railway and Other Transit Systems, COMPRAIL 2008*, pp. 545–553, esp, September 2008.

[11] R. M. P. Goverde and L. Meng, "Advanced monitoring and management information of railway operations," *Journal of Rail Transport Planning and Management*, vol. 1, no. 2, pp. 69–79, 2011.

[12] A. Ø. Sørensen, A. D. Landmark, N. O. E. Olsson, and A. A. Seim, "Method of analysis for delay propagation in a single-track network," *Journal of Rail Transport Planning and Management*, vol. 7, no. 1-2, pp. 77–97, 2017.

[13] A. Lindfeldt and H. Sipilä, "Simulation of freight train operations with departures ahead of schedule," in *Proceedings of the 14th International Conference on Railway Engineering Design and Optimization, COMPRAIL 2014*, pp. 239–249, Italy, June 2014.

[14] N. O. E. Olsson and H. Haugland, "Influencing factors on train punctuality - Results from some Norwegian studies," *Transport Policy*, vol. 11, no. 4, pp. 387–397, 2004.

[15] M. F. Gorman, "Statistical estimation of railroad congestion delay," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 3, pp. 446–456, 2009.

[16] M.-C. Shih, C. T. Dick, S. L. Sogin, and C. R. L. Barkan, "Comparison of capacity expansion strategies for single-track railway lines with sparse sidings," *Transportation Research Record*, vol. 2448, pp. 53–61, 2014.

[17] A. Lindfeldt, "A study of the performance and utilization of the Swedish railway network," in *Proceedings of the First International Conference on Road and Rail Infrastructure - CETRA2010*, S. Lakušić, Ed., University of Zagreb, Opatija, Croatia, 2010.

[18] T. Richter, "Systematic analyses of train run deviations from the timetable," in *Proceedings of the 12th International Conference on Computer System Design and Operation in the Railways and other Transit Systems, COMPRAIL 2010*, pp. 651–662, China, September 2010.

[19] E. Andersson, A. Peterson, and J. Törnquist Krasemann, "Robustness in Swedish Railway Traffic Timetables," in *Proceedings of the 4th International Seminar on Railway Operations Modelling and Analysis*, S. Ricci, I. A. Hansen, G. L. Longo, D. Pacciarelli, J. Rodriguez, and E. Wendler, Eds., pp. 1–18, Rome, 2011.

[20] B. H. Schittenhelm, "lanning With Timetable Supplements in Railway Timetables," in *Proceedings of the Annual Transport Conference at Aalborg University*, Aalborg, DK: trafikdage, 2011.

[21] E. Andersson, A. Peterson, and J. Törnquist Krasemann, "Introducing a New Quantitative Measure of Railway Timetable Robustness Based on Critical Points," in *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013*, vol. 3, pp. 1–19, Copenhagen, 2013.

[22] N. van Oort, D. Sparing, T. Brands, and R. M. P. Goverde, "Data driven improvements in public transport: the Dutch example," *Public Transport*, vol. 7, no. 3, pp. 369–389, 2015.

[23] I. A. Hansen, R. M. P. Goverde, and D. J. Van Der Meer, "Online train delay recognition and running time prediction," in *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, ITSC 2010*, pp. 1783–1788, Portugal, September 2010.

[24] P. Kecman and R. M. P. Goverde, "Process mining of train describer event data and automatic conflict identification," in *Computers in Railways XIII, WIT Transactions on the Built Environment*, C. A. Brebbia, N. Tomii, P. Tzieropoulos, and J. M. Mera, Eds., vol. 127, pp. 227–238, 2012.

[25] B. Cule, B. Goethals, S. Tassenoy, and S. Verboven, "Mining train delays," *Lecture Notes in Computer Science (including*

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface, vol. 7014, pp. 113–124, 2011.

[26] J. Wallander and M. Mäkitalo, "Data mining in rail transport delay chain analysis," *International Journal of Shipping and Transport Logistics*, vol. 4, no. 3, pp. 269–285, 2012.

[27] H. Yabuki, T. Ageishi, and N. Tomii, "Mining the Cause of Delays in Urban Railways based on Association Rules," in *CASPT2015*, pp. 1–16, 2015.

[28] G. Malavasi and S. Ricci, "Simulation of stochastic elements in railway systems using self-learning processes," *European Journal of Operational Research*, vol. 131, no. 2, pp. 262–272, 2001.

[29] P. Kecman, F. Corman, A. Peterson, and M. Joborn, "Stochastic prediction of train delays in real-time using Bayesian networks," in *Proceedings of the in. Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT*, p. 18, Rotterdam (Netherlands), 2015.

[30] X. Chapuis, "Arrival Time Prediction Using Neural Networks," in *7th International Conference on Railway Operations Modelling and Analysis*, N. Tomii, I. A. Hansen, J. Rodriguez, P. Pellegrini, S. Dauzère-Pérès, and D. De Almeida, Eds., pp. 1500–1510, International Association of Railway Operations Research, Lille, Fra, 2017.

[31] N. Marković, S. Milinković, K. S. Tikhonov, and P. Schonfeld, "Analyzing passenger train arrival delays with support vector regression," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 251–262, 2015.

[32] P. Kecman and R. M. P. Goverde, "Predictive modelling of running and dwell times in railway traffic," *Public Transport*, vol. 7, no. 3, pp. 295–319, 2015.

[33] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[34] H. Kashima, J. Hu, B. Ray, and M. Singh, "K-means clustering of proportional data using L1 distance," in *Proceedings of the 2008 19th International Conference on Pattern Recognition (ICPR)*, pp. 1–4, Tampa, FL, USA, December 2008.

[35] T. Richter, A. Landex, and J. L. E. Andersen, "Precise and accurate train run data: Approximation of actual arrival and departure times," in *Proceedings of the Precise and accurate train run data: Approximation of actual arrival and departure times*, International Association of Railways, Sydney, Australia, 2013.

[36] T. Richter, *Data aggregation for detailed analysis of train delays*, C. A. Brebbia, N. Tomii, J. M. Mera, B. Ning, and P. Tzieropoulos, Eds., WIT Transactions on the Built Environment. WIT-Press, 2012.

[37] J. Wu, *Advances in K-means Clustering*, Springer, Berlin, Heidelberg, 2012.

[38] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, New York, NY, USA, 2008.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 7905820, 19 pages
https://doi.org/10.1155/2018/7905820

Research Article

Integrated Optimization on Train Control and Timetable to Minimize Net Energy Consumption of Metro Lines

Yuhe Zhou[1], **Yun Bai**[1], **Jiajie Li**[1], **Baohua Mao**[1] and **Tang Li**[2]

[1]*MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China*
[2]*Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK*

Correspondence should be addressed to Yun Bai; yunbai@bjtu.edu.cn

Received 15 December 2017; Revised 21 February 2018; Accepted 19 March 2018; Published 26 April 2018

Academic Editor: Andrea D'Ariano

Energy-efficient metro operation has received increasing attention because of the energy cost and environmental concerns. This paper developed an integrated optimization model on train control and timetable to minimize the net energy consumption. The extents of train motoring and braking as well as timetable configurations such as train headway and interstation runtime are optimized to minimize the net energy consumption with consideration of utilizing regenerative energy. An improved model on train control is proposed to reduce traction energy by allowing coasting on downhill slopes as much as possible. Variations of train mass due to the change of onboard passengers are taken into account. The brute force algorithm is applied to attain energy-efficient speed profiles and an NS-GSA algorithm is designed to attain the optimal extents of motoring/braking and timetable configurations. Case studies on Beijing Metro Line 5 illustrate that the improved train control approach can save traction energy consumption by 20% in the sections with steep downhill slopes, in comparison with the commonly adopted train control sequence in timetable optimization. Moreover, the integrated model is able to significantly prolong the overlapping time between motoring and braking trains, and the net energy consumption is accordingly reduced by 4.97%.

1. Introduction

Metro systems develop rapidly across the world because of their high reliability and large capacity. For example, the total length of metro lines in mainland China was only about 1500 km in 2010 and more than 3700 km in 2016 and is expected to reach 6000 km by 2020. Although the metro is one of the most energy-efficient transport means, the long operating mileage and high traffic volume make the total energy consumption of metro systems become significant. Metro operators are facing more pressure to save energy due to increasing environmental concerns and operational costs. Energy-efficient train control and service timetable optimization are preferred to reduce metro train energy consumption in existing urban rail systems, as significant energy saving could be achieved with relatively low capital investment and minor modifications [1].

Energy-efficient train control (EETC) has been widely applied to real-world train operation since the 1980s. It aims to find the optimal sequence of train control modes and the switching points among the modes. Ichikawa [2] pointed out that energy-efficient train control on level tracks includes maximum acceleration (MA), cruising (CR), coasting (CO), and maximum braking (MB). Strobel et al. [3] found that it is also applicable to trains running in downhill and uphill sections with small slopes, where partial motoring and braking might be implemented to maintain cruising. Milroy [4] concluded that cruising is not necessarily required in short interstation runs. Subsequently, Benjamin et al. [5] and Howlett et al. [6] demonstrated that coasting presents opportunities for energy saving especially in long interstation runs. Khmelnitsky [7] researched the optimal operation of a train on a variable grade profile subject to arbitrary speed restrictions and presented a numerical algorithm to find the optimal velocity profile. Liu and Golovitcher [8] gave the analytical solution of optimal control and equations to find the control change points; speed limit and steep gradients are taken into account. Dongen and Schuit [9] and Scheepmaker

and Goverde [10] proposed two models to find the optimal cruising speed and the proper coasting location to minimize energy consumption for train interstation runs.

In recent years, Howlett et al. [11] proposed energy-efficient control approaches for freight trains on steep tracks. Their survey indicated that coasting before a steep downhill section is very important for energy saving. Bai et al. [12] found that up to 9% of the energy consumption for train interstation runs could be saved by applying coasting as much as possible before braking. Domínguez et al. [13] and Carvajal-Carreño et al. [14] presented approaches to attain the optimal schemes of the ATO speed profile to obtain the Pareto front between train runtime and energy consumption. Sicre et al. [15] considered a specific model for energy-efficient manual driving in high speed lines by means of fuzzy parameters; uncertainty in manual control and punctuality requirement are taken into account. Keskin and Karamancioglu [16] attached more importance to train mass and compared the effectiveness of different evolutionary algorithms in attaining solutions for energy-efficient train control. Shangguan et al. [17] developed a multiobjective optimization approach for single-train speed trajectory.

Energy-efficient train timetabling (EETT) aims to reduce the net energy consumption of metro lines via optimizing the timetable configurations, such as the distribution of runtime supplements among different interstations, dwell time at stations, and train headways. Albrecht and Oettich [18] applied dynamic programming to find the optimal timetable considering both train energy consumption and passenger waiting time. Sicre et al. [19] attained the Pareto curves of journey time and energy consumption for high speed train interstation itinerary and then optimized the distribution of runtime supplements among different interstations. Chevrier et al. [20] proposed a biobjective evolutionary algorithm to attain speed profiles for train interstation runs in which runtime and energy consumption were optimized concurrently.

EETT has become more popular in recent years with the application of regenerative braking [21], which is able to convert kinetic energy into electricity during train braking. Regenerative electricity could provide up to 40% of the energy supplied to trains [22]. The regenerative energy is firstly used to support onboard auxiliary devices of the braking train and then fed back into the overhead contact line to assist motoring and auxiliary equipment of other trains in the same power supply interval (PSI). If the feedback energy cannot be used in time, it will be consumed by protective resistor, as most metro systems have not installed energy storage devices. The motoring and braking trains could be synchronized for better utilization of regenerative energy by optimizing timetable configurations, which determine train motoring and braking timing when the train departs from and approaches stations. Albrecht [23] explored the optimal allocation of runtime supplement for train interstation runs, to minimize the net energy consumption and reduce the maximal load of power systems. Yang et al. [24] built a model to maximize the overlapping time between the motoring and braking of successive trains by regulating train headway and dwell time. Further, Yang et al. [25] and Le et al. [26] proposed models to maximize the utilization of regenerative braking energy (RBE) as well. Peña-Alcaraz et al. [27] designed a mathematical programming model to synchronize the braking of trains arriving at station with the acceleration of trains exiting from the same or another station.

More energy saving could be achieved by integrated optimization on train timetable and control scheme. Ding et al. [28] proposed a two-level optimization model to find the energy-efficient train trajectory as well as the distribution of runtime supplements to minimize the traction energy without changing the predefined journey time. Su et al. [29, 30] developed an integrated optimization model, consisting of both train control and timetable formulation, to minimize the total energy consumption of multiple trains. Yang et al. [31] developed a stochastic programming model for the integrated optimization on train timetable and speed profile, where train mass in each interstation was set as a stochastic variable with a given probability distribution. Li and Lo [32] developed an integrated model on train control and timetable formulation to minimize the net energy consumption, where regenerative braking was taken into account. Zhao et al. [33] presented an integrated optimization model on train trajectory and timetable to reduce the net energy consumption and peak power of substations; the effectiveness of the proposed model was verified by a multiple-train movement simulator.

Many previous researches have explored the integrated optimization of train operation for energy saving, including both train control and timetabling, especially for the metro lines where RBE is available. In these studies, the maximum traction and braking force are commonly applied in train control. However, partial traction and braking when trains depart from and approach stations are able to improve the utilization of regenerative energy via prolonging train motoring and braking time, at the expense of a slight increment of traction energy consumption. Moreover, the EETC applied in EETT is usually based on the control sequence of MA-CR-CO-MB. Howlett et al. [11] proved that CR-CO-CR saves more energy than the control scheme of only CR when the train runs on steep downhill slopes. Liu and Golovitcher [8] also presented an equation to derive the conditions where different control modes should be applied with consideration of steep downhill slopes. Nevertheless, these improved train control strategies have not been employed in energy-efficient timetable optimization problems.

In this paper, an integrated optimization model on train control and timetable is developed to minimize the system net energy consumption, considering both traction energy reduction and utilization of regenerative braking energy. Synchronization of motoring and braking trains to better use the regenerative braking energy is realized by optimizing the extents of train motoring and braking as well as timetable configurations including headway and scheduled interstation runtime. An improved train control model allowing coasting on downhill slopes as much as possible is proposed to further reduce traction energy consumption compared with the simple control sequence of MA-CR-CO-MB, which is commonly adopted in previous energy-efficient timetabling studies. The proposed model takes into account the practical operation conditions such as varied train mass in different interstations, power peak of traction power supply system,

Table 1: Literature on energy-efficient timetabling with RBE.

Publication	Objective(s)	Variables	Speed profile	Mechanical braking	Power peak	Auxiliary devices	Varied mass
Albrecht (2004)	Power peak + energy consumption	Trip time	MA-CO-MB		√		
Peña-Alcaraz et al. (2011)	Overlapping time	Arrival & departure time			√	√	
Li and Lo Hong (2014)	Net energy consumption	Speed profile, headway, arrival & departure time	MA-CO-MB				
Yang et al. (2015)	Net energy consumption	Dwell time	MA-CO-MB			√	
Yang et al. (2016)	Traction energy	Speed profile, arrival & departure time	MA-CO-MB				√
Su et al. (2016)	Traction energy	Varied train characteristic	MA-CO-MB	√	√	√	√
Zhao et al. (2017)	Substation energy consumption	Speed profile, interstation journey time, service intervals	MA-CO-MB		√		
This paper	Net energy consumption	Speed profile, headway, interstation runtime	MA-$[CO\text{-}CR]_n$-MB	√	√	√	√

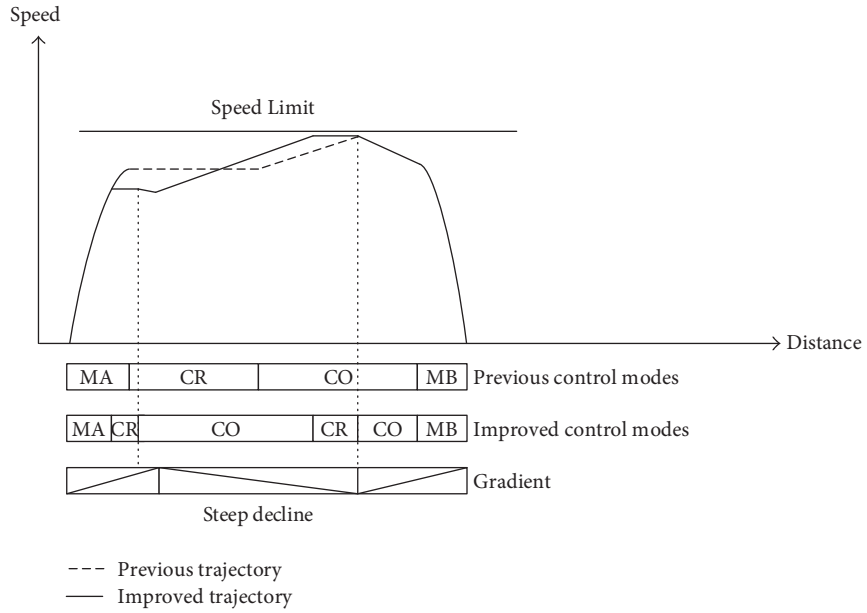


Figure 1: The improvements of the energy-efficient train control strategy.

and application of mechanical braking when the train adopts brakes at a very high speed. Table 1 gives a comparison between this study and the existing literatures.

The rest of this paper is organized as follows. In Section 2, an improved model on energy-efficient train control is presented. In Section 3, the integrated optimization model taking into account energy-efficient train control is developed to minimize the net energy consumption of a whole metro line. In Section 4, the brute force algorithm and an NS-GSA algorithm are presented to attain the optimal train trajectory and timetable, respectively. In Section 5, case studies on Beijing Metro Line 5 are conducted to verify the effectiveness of the proposed approach. Conclusions are provided in Section 6.

2. Energy-Efficient Train Control

There is usually more than one control scheme to drive a train from one station to the next even with the same runtime. The energy-efficient train control is to find the optimal control scheme leading to the minimal traction energy consumption while the scheduled runtime and speed limits are respected. Most previous studies proposed that the energy-efficient train control consists of the following four successive modes (i.e., MA, CR, CO, and MB) in the interstation with constant speed limit and no steep slopes, as illustrated by the dashed line in Figure 1.

With such a control, the train adopts maximum motoring followed by cruising at a constant speed and then coasting

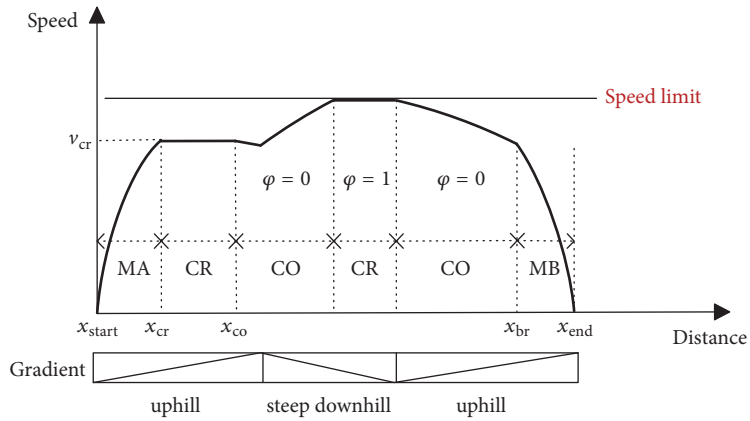


Figure 2: Improved energy-efficient train control scheme.

till braking for station stops. However, trains must brake to keep the speed constant in steep downhills, which is not preferred to minimize traction energy. The train could start cruising earlier and then adopt coasting on downhill slopes, followed by employing cruising and coasting again before stop braking, as illustrated by the solid line in Figure 1. As such, the energy consumption of the train interstation run could be saved, as the time for MA in the new trajectory is shortened while the interstation runtime remains the same. To this end, an improved model on train control is proposed to reduce the traction energy consumption by taking the full advantage of potential energy in downhill tracks, with the assumption that there is no change of speed limit in the whole interstation run.

2.1. Objective Function. The train adopts MA and MB in accelerating from and approaching stations in the proposed train control model, while the pairs of CR and CO are implemented in the interstation runs, as shown in Figure 2. It should be noted that Figure 2 is a sketch and switching points among different control modes might not be precise. With the proposed model, train control scheme is based on MA-$[\text{CR-CO}]_n$-MB and the positive integer n represents the number of CR and CO in one interstation journey. n is equal to 1 in most cases and the control scheme is MA-CR-CO-MB, which is the same as that attained by the previous model. n might be greater than 1 when the train is running on steep downhill tracks. Therefore, the proposed model is a supplement to previous work and it could contribute toward energy saving in the interstations with steep downhill profiles since more coasting is allowed.

The decision variables in the improved model are two switching points during one interstation run. The first point is the location where the first CR is applied, that is, x_{cr}. The second point is the location where the first CO is applied, that is, x_{co}. The other switching locations could be obtained once the decision variables are known. For example, the location to start the second CR is where the train reaches the speed limit. The train continues CR till it leaves the steep downhill slopes and CO is implemented again when the train does not accelerate with no traction power. The location to start applying MB, which is denoted by x_{br}, depends on the intersection point of coasting and braking profiles before the station stop. The objective function of the proposed model is to minimize the traction energy consumption for one interstation run, which is described as

$$\min \quad E_{\text{trac}}(x_{cr}, x_{co}) = \frac{\sum_{x=0}^{X} \max(\mathbf{F}(v,x), 0)}{3600 \cdot \eta_1}, \tag{1}$$

where E_{trac} is traction energy consumption with the given x_{cr} and x_{co} which are decision variables in EETT; η_1 is the energy conversion efficiency factor from electrical to mechanical energy; X is the length of one interstation; the units for distance, force, power, energy, and time are defined as m, kN, kW, kWh, and s in this paper; $\mathbf{F}(v,x)$ stands for the output force acting upon the train given by ATO. A positive value of $\mathbf{F}(v,x)$ denotes that the train is motoring. Zero indicates that the train is coasting and a negative value represents braking.

Traction force and braking force are expressed as a vector force that can be described as

$$\mathbf{F}(v,x) = \begin{cases} F_{\text{tr}}(v,x) & 0 \le x \le x_{cr} \\ F_{\text{cr}}(v,x) & x_{cr} < x \le x_{co} \\ \varphi \cdot F_{\text{cr}}(v,x) & x_{co} < x \le x_{br} \\ -F_{\text{br}}(v,x) & x_{br} < x \le X, \end{cases} \tag{2}$$

where $\mathbf{F}(v,x)$ is a vector force which is greater than 0 in traction mode or less than 0 in braking mode; $F_{\text{tr}}(v,x)$ is the maximum available traction force according to motor characteristic; $F_{\text{cr}}(v,x)$ is the force to keep cruising at position x; $F_{\text{br}}(v,x)$ is the maximum available braking force; φ is a binary variable with no unit.

$\varphi = 0$ indicates that the train control mode is CO, while $\varphi = 1$ stands for the case in which the control mode is CR. For example, coasting is implemented when the train arrives at x_{co} and φ is thus set as 0. When train speed increased to the limit, CR will be implemented again and φ becomes 1. Then resistance force F_R is used to decide the following

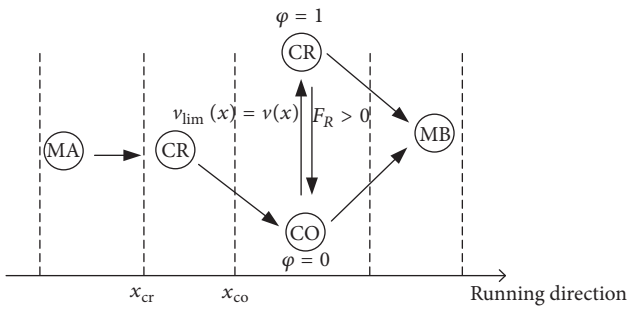


Figure 3: Transitions of train control modes with the improved model.

control regimes. φ can be therefore attained by the following equation:

$$\varphi(x) = \begin{cases} 0 & (v_{\lim}(x) - v(x)) + \varphi(x - \Delta x) \cdot F_R(v, x) > 0 \\ 1 & (v_{\lim}(x) - v(x)) + \varphi(x - \Delta x) \cdot F_R(v, x) \le 0, \end{cases} \quad (3)$$

where $v_{\lim}(x)$ and $v(x)$ are speed limit and train speed at position x, respectively. $\varphi(x-\Delta x)$ represents the train control mode in the previous distance step, where Δx represents the length of distance step. $F_R(v, x)$ is resultant resistance, consisting of friction resistance, air drag, and additional resistance caused by grades and curves.

Figure 3 demonstrates the possible transitions among different control modes. At the beginning stage, MA should be applied till the location where the train starts cruising, which is the first decision variable x_{cr}. In the mode of CR, the effort acting upon the train by ATO is equal to the resistance to keep train speed constant. Train continues its travel at a constant speed till it reaches the location to start CO, which is the second decision variable x_{co}. The following driving mode depends on train speed and running resistance according to (3). If train speed $v(x)$ does not exceed the speed limit $v_{\lim}(x)$, coasting phase will be continued. If train speed exceeds the limit, cruising phase will be adopted again. The coasting will be applied again only when the resultant resistance is positive, which indicates train speed decreases when coasting is applied. Finally, MB is applied before the station stop and the braking curve is attained by backward calculation from the target stop point using the maximum available braking effort.

According to the characteristics of train motor, the traction force varies with train speed. When MA is applied, there are three traction force curves to be chosen automatically in response to three train weights, as shown in Figure 4(a). When the train is empty, the traction force curve of AW0, which is the lowest one among all the three curves, is chosen because the train does not need too much traction force for acceleration. In case the number of passengers on the train is equal to the nominal capacity, the traction force curve of AW1 is adopted to allow a bigger acceleration rate than that of AW0. The traction force curve of AW2, which is the highest one, is selected when the train carries the most passengers. For the circumstances in which the actual train weight is not equal to one of the three weights above, the traction force is calculated as follows.

Firstly, we need to calculate train mass to obtain the available traction force.

$$m(x) = M_{\text{train}} + \mu \cdot \tau(n), \quad (4)$$

where $m(x)$ is the actual train mass; M_{train} is the rolling stock mass; $\tau(n)$ is the number of passengers on the train in the interstation of n; μ is the average mass of a person.

Secondly, linear interpolation is used to attain the practical traction force, as shown in

$$F_{\text{tr}}(v, x) = \begin{cases} F_0(v) + (F_1(v) - F_0(v)) \cdot \dfrac{m(x) - M_0}{M_1 - M_0} & M_0 \le m(x) < M \\ F_1(v) + (F_2(v) - F_1(v)) \cdot \dfrac{m(x) - M_1}{M_2 - M_1} & M_1 \le m(x) \le M_2, \end{cases} \quad (5)$$

where M_0, M_1, and M_2 represent train masses when the train is empty, nominally loaded, and maximum loaded, respectively; $F_0(v)$, $F_1(v)$, and $F_2(v)$ indicate the traction force in the above three circumstances.

Train mass also has impacts on train braking force. There are usually three different braking force curves corresponding to three different load factors, that is, AW0-AW1-AW2. The braking force under different load factors can be calculated similarly. However, the braking force usually keeps constant when train speed varies [34–36]. It should be noted that the practical braking force (F_{br}) is the combination of both mechanical braking and regenerative braking. Figure 4(b) shows the components of braking force, taking the case of AW1 as an example. Regenerative braking force (F_{eb}) declines gradually when train speed becomes very large or very small, and mechanical braking (F_{mech}) is applied to compensate the shortage of braking force.

As the traction force $F_{\text{tr}}(v, x)$ and braking force $F_{\text{br}}(v, x)$ are obtained, the traction energy consumption can be calculated by (1) and (2) once we obtained the value of $F_{\text{cr}}(v, x)$. The resultant force $F(v, x)$ acting upon the train can be calculated as

$$\begin{aligned} F(v, x) &= \mathbf{F}(v, x) - F_R(v, x) \\ F_R(v, x) &= R_{\text{basic}}(v) + R_{\text{grad}}(x) + R_c(x), \end{aligned} \quad (6)$$

where $R_{\text{basic}}(v) = a + bv + cv^2$ and the coefficients of a, b, and c are usually given by rolling stock manufacturers; $R_{\text{grad}}(x)$ is the resistance caused by the gradient; $R_c(x)$ is the resistance caused by the curve.

When cruising is applied, the resultant force $F(v, x)$ should be equal to zero; then, $F_{\text{cr}}(v, x)$ can be obtained based on Newton's law of motion:

$$\begin{aligned} F_{\text{cr}}(v, x) &= F_R(v, x) \\ &= m(x)\, g \cdot \left[\left(a + bv + cv^2 \right) + \sin\theta_x + \frac{600}{r_x} \right], \end{aligned} \quad (7)$$

where g is gravitational acceleration; θ_x is the angle of slope (negative means downhill); r_x is the radius of the curve at location x.

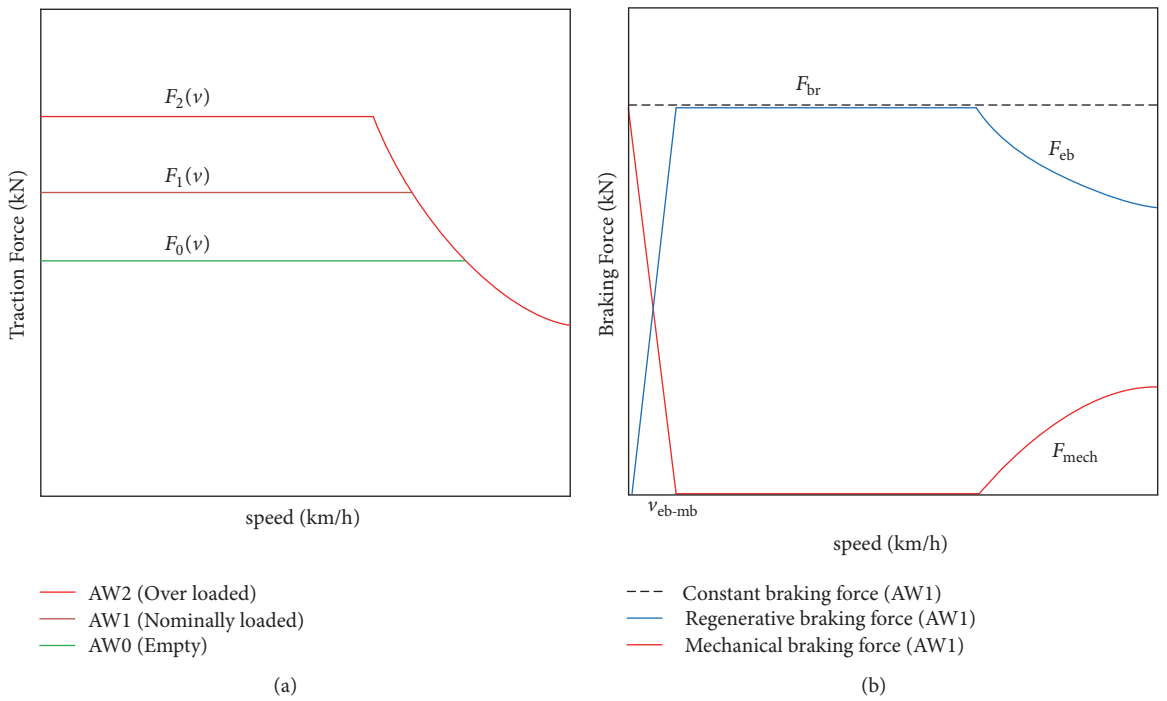


FIGURE 4: Maximum force against train speed. (a) Maximum traction force with different load factors. (b) Available braking forces in AW1.

2.2. Constraints. Train control schemes should be subject to the following constraints.

(1) For operational safety and passenger comfort, the acceleration should be limited within a proper range:

$$a_{\text{dec}} \leq \frac{dv(x)}{dt} \leq a_{\text{acc}}. \tag{8}$$

(2) The train must stop when it arrives at a station:

$$v(x_n) = 0, \quad \forall n \in [1, 2N]. \tag{9}$$

(3) Train velocity must not exceed the speed limit or be negative:

$$0 \leq v(x) \leq v_{\text{lim}}(x). \tag{10}$$

(4) The error between practical runtime and scheduled runtime should be less than a certain threshold. In this paper, we set this threshold as $\delta = 0.5$:

$$\left| \text{RT}_n - \sum_{x=x_n}^{x_{n+1}} \frac{\Delta x}{v(x)} \right| \leq \delta, \tag{11}$$

where RT_n is the scheduled runtime in interstation n; x_n refers to the location of station n.

(5) This constraint aims to calculate regenerated braking energy for timetable optimization in Section 3. As a matter of fact, regenerative braking force varies with train speed as well as load factor, and it will be entirely replaced by mechanical braking when train speed is lower than $v_{\text{eb-mb}}$.

$$F_{\text{eb}}(v, x) = \begin{cases} \min\left(F_{\text{eb}}(v, x), F_{\text{br}}\right), & v \geq v_{\text{eb-mb}} \\ 0, & v < v_{\text{eb-mb}}. \end{cases} \tag{12}$$

3. Integrated Optimization Model on Train Control and Timetable

With the application of regenerative braking, synchronization of motoring and braking trains in the same PSI for better utilization of regenerative energy becomes very important in minimizing the net energy consumption of metro lines, since the regenerative energy can provide a significant proportion of the total energy fed into rolling stocks. Train control also has a great influence on timetable optimization, as different control schemes lead to different energy consumption even when the interstation runtime remains the same.

An integrated model on train control and timetable optimization is proposed in this section to minimize the net energy consumption, taking into account the improved train control model described in Section 2. Train headway and interstation runtimes are optimized to improve the utilization of RBE as well as reduce the traction energy consumption by allocating the runtime supplements properly. In addition, power control system could be modified to constrain the output of traction and braking force when trains depart from and approach stations, which is able to enlarge the space for synchronization of motoring and braking trains via prolonging train motoring and braking time, although traction energy consumption increases slightly. Different from the previous studies, the new integrated model allows the train to adopt partial motoring and braking in accelerating from and approaching stations.

Figure 5 illustrates the influences on energy consumption by train headway, interstation runtime, and percentages of the full traction and braking force applied in train control. The motoring of train B and the braking of train A could be

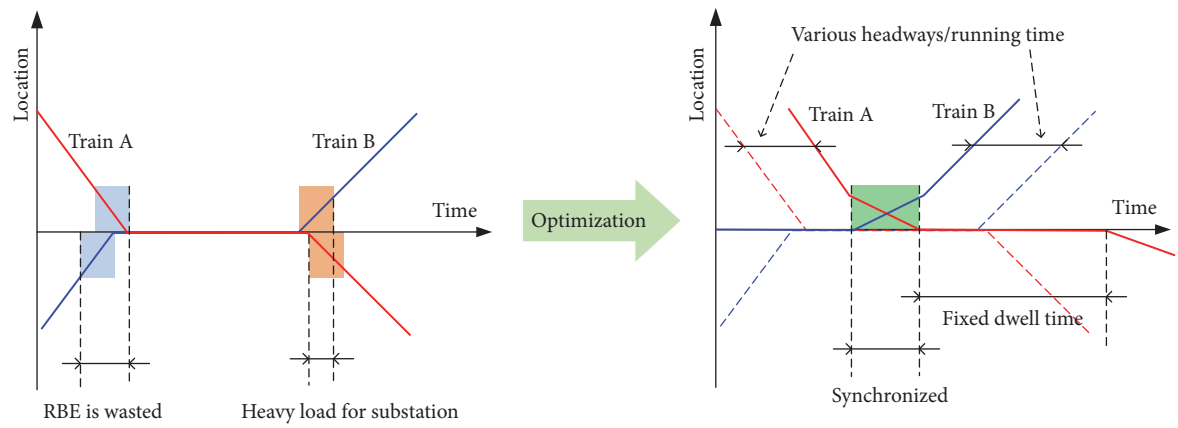


Figure 5: Energy-saving methods in integrated optimization.

synchronized by regulating train headway and interstation runtime; therefore, the regenerated energy produced by braking train A can be transferred to and utilized by traction train B. As a result, net energy consumption is reduced because the energy used in the accelerating phase of train B is provided by the regenerative braking of train A. The overlapping time, which means the time slot in which a train adopts regenerative braking while there is at least one train motoring in the same PSI, is further extended using partial traction and braking force.

Net energy consumption could be saved due to the much higher utilization of RBE, although the traction energy consumption may increase slightly. It should be pointed out that dwell time is not allowed to be changed in the integrated optimization model because it usually depends on the number of passengers getting on and off the train.

3.1. Objective Function. In this study, the whole problem of energy-efficient train operation is decomposed into train trajectory formulation and timetable optimization, which are successively processed. Firstly, train control scheme in each interstation run is optimized with the given runtime (RT), k_t, and k_b. Secondly, a database, including the optimal train control scheme and the corresponding energy consumption with all possible groups of RT, k_t, and k_b, is then built. Finally, the optimal HD, RT, k_t, and k_b are obtained by an evolutionary algorithm on the basis of the database. It should be noted that k_t and k_b are assumed as two constants for the sake of calculating efficiency and convenience of train control implementation.

The objective of the integrated model is to find the optimal utilization coefficients of traction and braking force (k_t, k_b), interstation runtime (RT), and headway (HD) to minimize the net energy consumption for all trains $j \in \{1, 2, 3, \ldots, J\}$ operating on the whole line, which is the difference between the required traction energy and the utilized regenerative energy. Additionally, energy consumed for auxiliary devices is also taken into account. The objective function is expressed as

$$\begin{aligned} \min \quad & E\left(k_t, k_b, \mathrm{RT}, \mathrm{HD}\right) \\ & = \sum_{j=1}^{J} \sum_{n=1}^{2N} E_{\mathrm{trac}}^{n,j}\left(k_t, k_b, \mathrm{RT}_n\right) + E_{\mathrm{aux}} \\ & \quad - \sum_{t=1}^{T_{\mathrm{total}}} \sum_{n_p=1}^{N_p} \sum_{j \in \kappa} E_{\mathrm{reg\_u}}\left(t, n_p, j, \mathrm{RT}, \mathrm{HD}\right), \end{aligned} \tag{13}$$

where k_t and k_b represent traction and braking force utilization coefficients; N_p is the number of PSIs; n_p represents a PSI on the metro line and $n_p \in [1, N_p]$; $E_{\mathrm{trac}}^{n,j}(k_t, k_b, \mathrm{RT}_n)$ is total traction energy consumption in interstation n for train j; E_{aux} is energy consumed by auxiliary devices; $E_{\mathrm{reg\_u}}(t, n_p, \mathrm{RT}, \mathrm{HD})$ is the amount of utilized RBE in n_p at time step t; T_{total} is the length of operation period; κ denotes the set of trains located in the same power supply interval n_p.

Partial traction and partial braking are applied to extend the motoring and braking phases, which result in a modification on the EETC model. It is known that the efficiency of traction systems usually increases with the extent of motoring [37] and a constant efficiency factor could lead to inaccurate results [38]. Therefore, energy conversion factor η_1 in (1) is defined as a variable calculated as follows [39]:

$$\eta_1 = \begin{cases} 1 & 0.95 \le k_t < 1 \\ 0.6620 + 0.3558 \cdot k_t & 0.5 \le k_t < 0.95 \\ 0.5714 + 0.537 \cdot k_t & k_t < 0.5. \end{cases} \tag{14}$$

Due to the force utilization coefficients, (2) in Section 2.1 is rewritten as

$$\mathbf{F}\left(k_t, k_b, v, x\right) = \begin{cases} k_t \cdot F_{\mathrm{tr}}\left(v, x\right) & 0 \le x \le x_{\mathrm{cr}} \\ F_{\mathrm{cr}}\left(v, x\right) & x_{\mathrm{cr}} < x \le x_{\mathrm{co}} \\ \varphi \cdot F_{\mathrm{cr}}\left(v, x\right) & x_{\mathrm{co}} < x \le x_{\mathrm{br}} \\ -k_b \cdot F_{\mathrm{br}}\left(v, x\right) & x_{\mathrm{br}} < x \le X. \end{cases} \tag{15}$$

Then, traction energy consumption could be calculated by attaining the optimal (x_{cr}, x_{co}) in energy-efficient train control model.

The traction energy comprises not only propulsion of the train, but also consumption of auxiliary systems aboard the train such as lighting, air conditioning, and the signal system. Usually, the power of onboard auxiliary devices P_{aux} remains constant during train movements. Hence, the total energy consumption for onboard auxiliary devices is only related to the cycle time, that is, the period required for one train to complete one cycle, which is expressed as C.

$$E_{aux} = \sum_{j=1}^{J} \frac{P_{aux}^{j} C}{3600}. \tag{16}$$

The calculation of utilized RBE in each PSI during the operation period is more complex and the framework on utilization of RBE is given in Figure 6. The RBE is sequentially used to support auxiliary devices on the braking trains and traction trains in the same PSI and auxiliary devices on the other trains except braking trains. It should be noted that the regenerative energy from braking trains can only be transmitted among trains which are operating in the same PSI.

The utilization of RBE consists of two parts. The first part is used to support onboard auxiliary devices in braking trains, and the second part is used by traction trains and onboard auxiliary devices on other trains except braking trains, which is distinguished by different lines in Figure 6. Therefore, utilized RBE for all trains j in the same PSI at time step t can be calculated as

$$E_{reg\_u}(\xi) = E_{reg\_u}^{self}(\xi) + \eta_2 \cdot E_{reg\_u}^{other}(\xi), \tag{17}$$

where $\xi = \{t, n_P, j, \mathrm{RT}, \mathrm{HD} \mid 1 \le t \le T_{total},\ 1 \le n_P \le N_P,\ j \in \kappa\}$; η_2 is the transmission loss factor of the regenerative energy.

The first part of utilized RBE which is used by braking trains is calculated by (18). If train j is braking in power supply interval n_p at time t, the vector force $\mathbf{F}$ will be negative. This means the train is providing regenerative energy. A portion of RBE is used for onboard auxiliary devices on the braking train j firstly.

$$E_{reg\_u}^{self}(\xi) = \sum_{j\in\kappa} \min\left(\eta_3 \frac{\left|\min\left(\mathbf{F}_j(v,x)\, v\Delta t, 0\right)\right|}{3600}, \frac{P_{aux}^{j}\Delta t}{3600}\right), \tag{18}$$

where η_3 denotes the conversion factor from mechanical energy to regenerative energy.

The rest of the RBE, which can be attained by (19), is available to accelerate the motoring train followed by providing power for the auxiliary devices of other trains except braking trains in the same PSI. The energy required by motoring trains is calculated by (20).

$$E_{reg}^{avai}(\xi) = \sum_{j\in\kappa} \max\left(\eta_3 \frac{\left|\min\left(\mathbf{F}_j(v,x), 0\right)\right| v\Delta t}{3600} - \frac{P_{aux}^{j}\Delta t}{3600}, 0\right) \tag{19}$$

$$Ee_{trac}(\xi) = \sum_{j\in\kappa} \frac{\max\left(\mathbf{F}_j(v,x), 0\right) v\Delta t}{3600 \cdot \eta_1}. \tag{20}$$

If the available RBE is more than the total traction energy required by motoring train in the same PSI, the surplus RBE is then used to support auxiliary devices for other trains except the braking trains in the same PSI. The total energy required by all the trains in the same PSI except the braking trains can be calculated by (21). Finally, the total utilized regenerated energy by traction trains and other trains except the braking trains is calculated in (22).

$$Ee_{aux}(\xi) = \sum_{j\in\kappa} \frac{P_{aux}^{j}\Delta t}{3600}, \quad \forall \mathbf{F}_j(v,x) \ge 0 \tag{21}$$

$$E_{reg\_u}^{others}(\xi) = \begin{cases} Ee_{trac}(\xi) + \min\left(E_{reg}^{avai}(\xi) - Ee_{trac}(\xi), Ee_{aux}(\xi)\right) & Ee_{trac}(\xi) \le E_{reg}^{avai}(\xi) \\ E_{reg}^{avai}(\xi) & Ee_{trac}(\xi) > E_{reg}^{avai}(\xi), \end{cases} \tag{22}$$

where $Ee_{aux}(\xi)$ is the total energy consumption of onboard devices of trains in the same PSI, except the braking trains.

3.2. Constraints. In this paper, we consider the operation period from the time when the first train is put into operation in the up direction to the time when the last train returns to the depot from the down direction. During the operation period, all trains have the same cycle time, traction and braking force utilization coefficients, interstation runtime, and dwell time. The constraints for energy-efficient timetable are as follows:

(1) Runtime constraints: the interstation runtime should be limited within a certain range for the consideration of service quality and the maximum traction force.

$$\underline{T_{cn}} \le T_n \le \overline{T_{cn}}, \quad \forall n \in [1, N-1] \cup [N+1, 2N-1]. \tag{23}$$

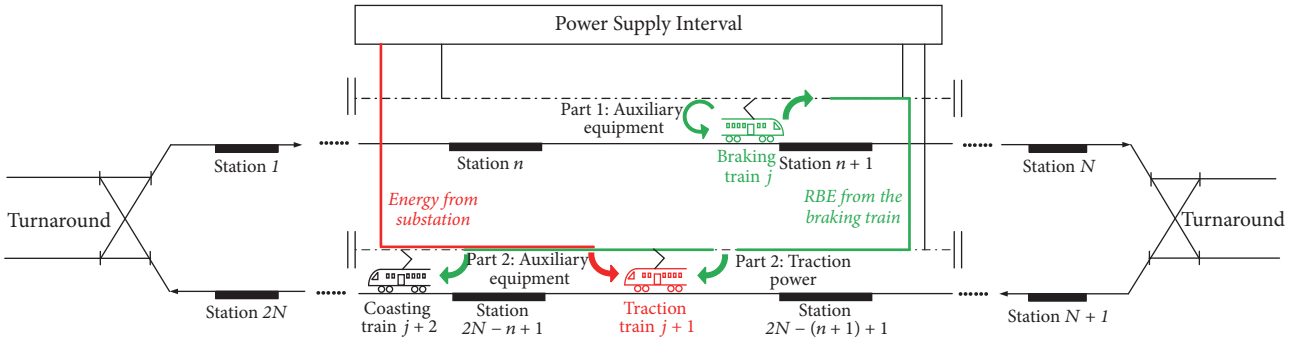


Figure 6: The utilization of regenerative braking energy.

(2) Turnaround time constraints: the turnaround time should be larger than the minimum time due to the track length of turnaround and limited travel speed, and less than a certain upper bound for the consideration of service quality and rolling stock operation.

$$\underline{T_{\text{turn}}} \le T_n \le \overline{T_{\text{turn}}}, \quad n \in \{N, 2N\}. \tag{24}$$

(3) Cycle time constraints: the total travel time for one cycle in the optimal timetable should be equal to the one in the original timetable, to keep the service frequency unchanged. In this study, we assume turnaround time as a special interstation runtime.

$$C = \sum_{n=1}^{2N} T_n + \sum_{n=1}^{2N} T_n^{\text{dwell}}. \tag{25}$$

(4) Headway constraints: the headway between any two successive trains must be less than the maximal headway required by service quality and larger than the minimal one for safety concerns.

$$\underline{H} \le H_j \le \overline{H}. \tag{26}$$

(5) Power peak: the total traction power of all trains located in the same PSI at each time moment should be no more than the maximal load of the power supply systems $P_{\text{sub}}^{\max}$.

$$\sum_j P_j(t) \le P_{\text{sub}}^{\max}. \tag{27}$$

(6) The percentages of full traction and braking force applied in motoring and braking should be limited in a reasonable range, considering the service quality requirement.

$$\begin{aligned} k_b &\in [0.5, 1] \\ k_f &\in [0.5, 1]. \end{aligned} \tag{28}$$

4. Solution Algorithm

Evolutionary algorithms, particularly the genetic algorithm (GA), have been widely adopted to solve such kind of problems especially the scheduling optimization [40–42]. However, it is easy to fall into a local solution rather than the global optimum if a standard GA is employed to solve such a complex integrated optimization problem. To improve the computing efficiency, the whole problem of energy-efficient train operation is decomposed into train trajectory formulation and timetable optimization, which are successively processed in this study. Figure 7 shows the whole structure of the optimization process.

In the single-train trajectory optimization, the brute force algorithm is applied to find the exact solution for train control with the given traction and braking force utilization coefficients as well as runtime in each interstation, and the results are saved in a database. Then, a genetic annealing algorithm with neighborhood search strategy is implemented in the integrated optimization to calculate the proper traction and braking force utilization coefficients, runtime, and headway. Train trajectories and traction energy consumption are directly taken from the database.

4.1. Brute Force Algorithms to Attain Train Trajectories. Brute force is an exact algorithm by searching all the possible solutions and evaluating their fitness, which has been successfully used in railway operation optimization [43, 44]. The main steps of brute force algorithm are as follows.

(1) Enumerating: the energy consumption and interstation runtime of all the possible solutions (x_{co}, x_{cr}) in the interstation run will be calculated, with the given force utilization coefficients (k_t, k_b) and running direction (d). The results will be stored in the database.

(2) Selection: the database may include a number of solutions with the same interstation runtime (RT), but the traction energy consumption is different. For each runtime, only the solution with the least traction energy consumption will be retained and other solutions with the same runtime will be removed from the database.

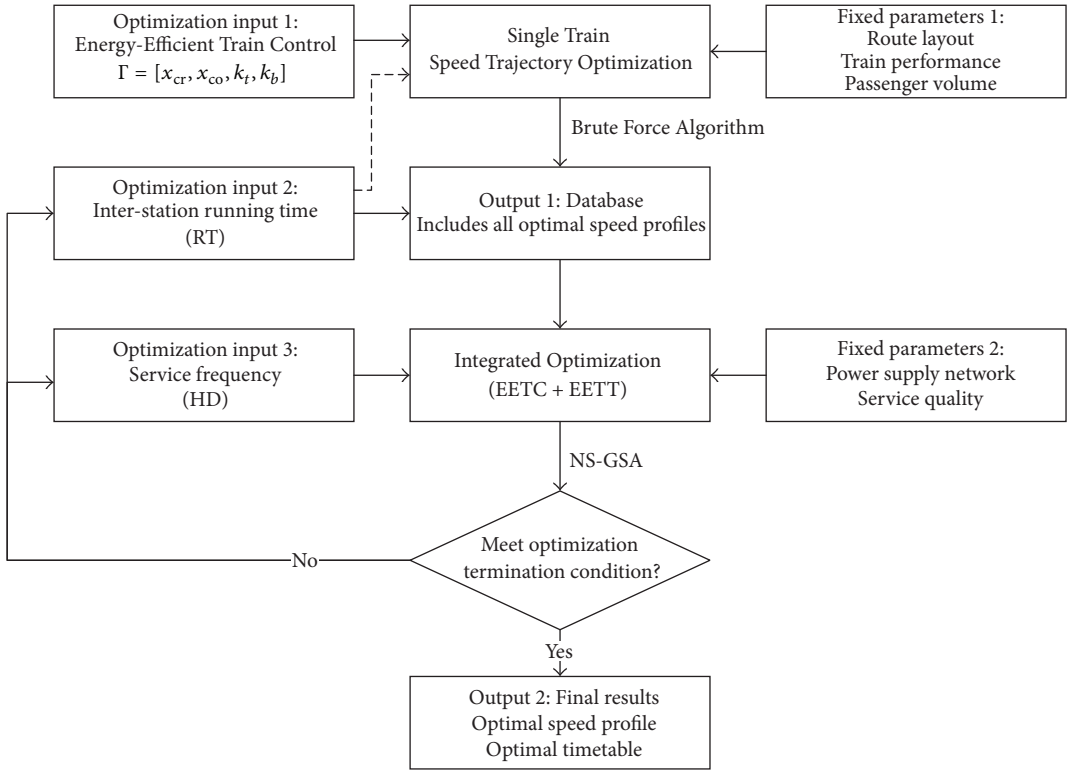


Figure 7: Flowchart of the integrated optimization.

(3) Updating: the force utilization coefficients (k_t, k_b) and running direction (d) will be updated, while the value should satisfy the constraint in (28). Then, the algorithm returns back to Step (1). If all the possible combinations of k_t, k_b, and d in this interstation are calculated, go to Step (4).

(4) Termination: the optimal train control schemes in the interstation with different runtimes, running directions, traction force utilization coefficients k_t, and braking force utilization coefficients k_b are obtained and stored in the database.

The algorithm takes about 1.1 hours to generate a database for one interstation, which gives the optimal solutions and the corresponding energy performance under 2160 different operation conditions, involving 6 traction coefficients, 6 braking coefficients, 30 runtimes, and 2 run directions.

4.2. NS-GSA Algorithm in Timetable Optimization. An improved genetic annealing algorithm with neighborhood search strategy (NS-GSA) is used to optimize the timetable. To overcome the premature convergence in the standard genetic algorithm, a simulated annealing (SA) algorithm is introduced to escape from the local optimum and approach the global optimum. Metropolis criterion of simulated annealing (SA) allows a decreasing probability of accepting worse solutions, which provides a diverse population for GA without compromising solution convergence. Additionally, the solution space is quite large in this integrated optimization problem, and the variables are continuous. Therefore, it is necessary to enhance the local search ability of genetic simulated annealing (GSA) algorithm, and neighborhood search strategy (or local search) is used which has been proved to be an effective search technique and is widely implemented in railway operation and management [45–48]. The procedure of NS-GSA is shown in Figure 8.

The main steps of the developed algorithm are as follows:

(1) Initialization: a random initial population of the solutions is produced to form the first generation. A number of individuals are included in the population and each individual represents a set of traction force utilization coefficient (k_t), braking force utilization coefficient (k_b), interstation runtime (RT), and headway (HD), which are the decision variables in the integrated optimization model.

(2) Neighborhood search: each individual explores its neighborhood to enhance the local search ability of the algorithm. The details on neighborhood search are presented below.

   (a) Initialize the set of neighborhood structures SOL(p, q), $p \in \{1, \ldots, P\}$, $q \in \{1, \ldots, Q\}$. P is the number of individuals and Q is the maximum number of local searches for each individual. Set $p = 1$, $q = 1$.

   (b) Until $q = Q + 1$, repeat the following operations. Through randomly generating a set of RT within a reasonable range, a set of solutions SOL(p, q) in the neighborhood of SOL(p) is

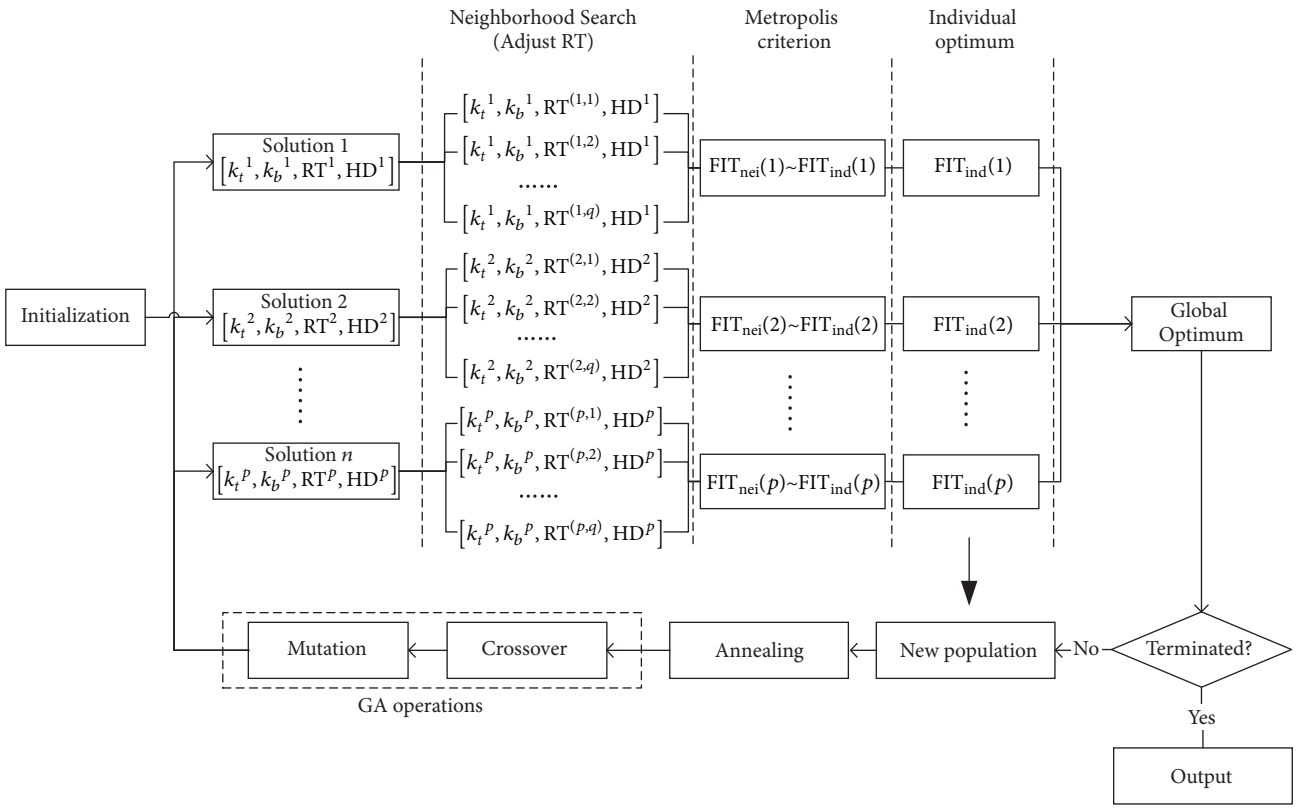


Figure 8: Flowchart of NS-GSA algorithm.

produced by replacing the RT in SOL(p) with the newly generated RT. It should be noted that the journey time should not be changed in each local searching. Set $q \leftarrow q + 1$.

(c) Set $p \leftarrow p+1$. If $p > P$, finish the neighborhood search. Otherwise, set $q = 1$ and go to Step (b).

(3) Evaluation: each solution in the population will be evaluated. Equation (13) is used to calculate the fitness FIT(p, q) of solution SOL(p, q).

(4) Acceptance: if the best neighborhood solution of the individual p with the fitness $FIT_{nei}(p)$ is better than the individual fitness $FIT_{ind}(p)$, the individual will be replaced by the neighborhood solution. Otherwise, the individual will be replaced by the best neighborhood solution according to the Metropolis criterion in simulated annealing algorithm. The acceptance criterion is listed as follows:

$$\begin{aligned} &\forall p \in P, \text{if } FIT_{nei}(p) < FIT_{ind}(p) \\ &FIT_{ind}(p) = FIT_{nei}(p) \\ &\text{elseif } e^{(FIT_{ind}(p) - FIT_{nei}(p))/Temper} < Rand(0,1) \\ &FIT_{ind}(p) = FIT_{nei}(p). \end{aligned} \tag{29}$$

(5) New generation: if the termination conditions are not satisfied, the optimal individuals after neighborhood searching will be used to create a new generation. The parameter Temper in Metropolis criterion decreases at a certain rate at the same time.

(6) Crossover: set the random number k_c ($0 \le k_c \le 1$) and the probability of crossover P_c. If $k_c \le P_c$, then crossover operation will randomly select three positions of genes in three parts of the chromosome, corresponding to force utilization coefficients, interstation runtime, and headway. Then, two chromosomes in the population will exchange these genes. As shown in Figure 9, genes which have been exchanged in crossover are marked in green.

(7) Mutation: set random number k_m ($0 \le k_m \le 1$) and the probability of crossover P_m. If $k_m \le P_m$, the mutation operation will randomly select one gene from one chromosome and replace the gene with a random number within the reasonable range. It should be noted that constraints in Sections 2.2 and 3.2 should be satisfied.

(8) Termination: the algorithm returns to Step (2) and repeats until the maximum generation is reached.

The required CPU time of NS-GSA algorithm is about 20 hours to attain the energy-efficient timetable as well as the optimal control schemes using a computer with 3.2 GHz processor speed and 4 GB memory, when the number of individuals in GA is 20, the number of neighborhood solutions for each individual is 2, and the maximum generation is 40.

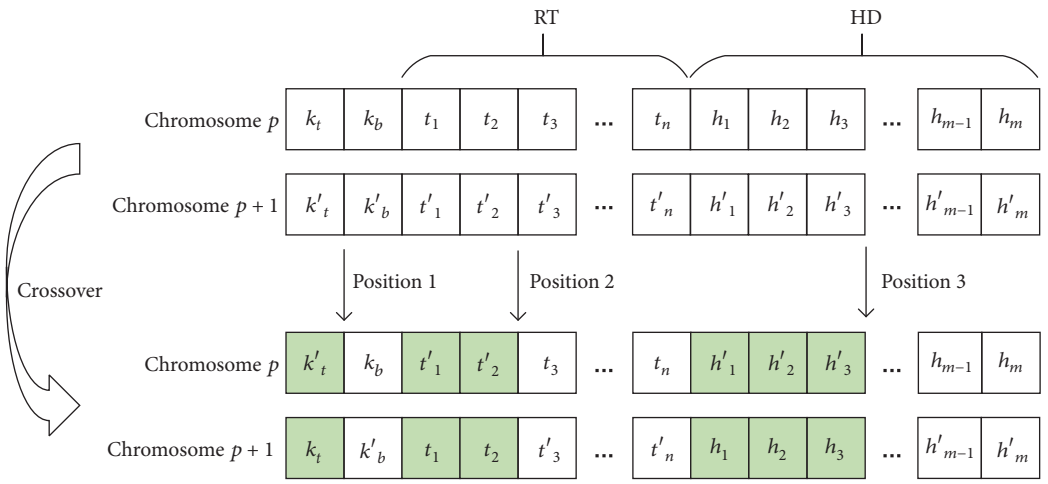


FIGURE 9: The structure of chromosome and crossover operation.

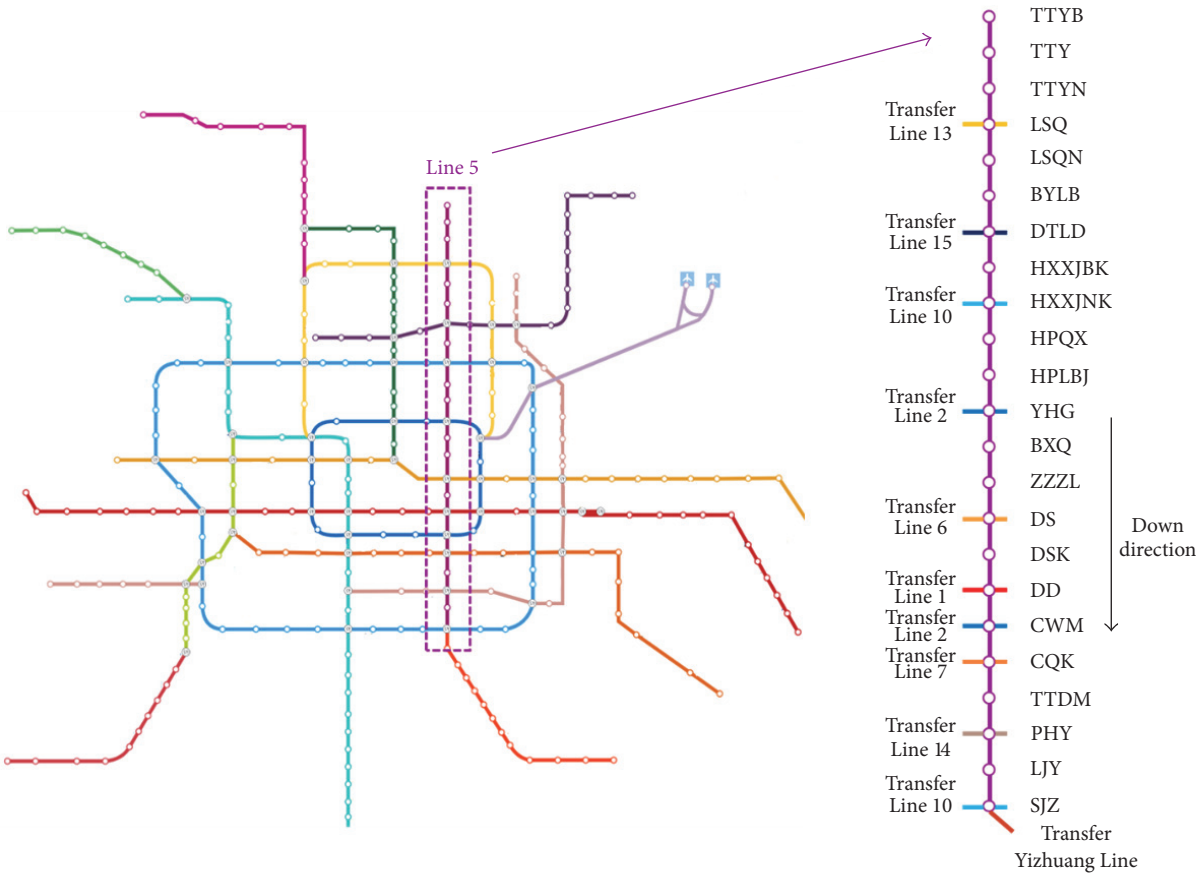


FIGURE 10: Beijing Metro network and Line 5.

5. Case Study

In this section, case studies on Beijing Metro Line 5 in peak hours are employed to illustrate the effectiveness of the improved train control and integrated optimization model.

5.1. Basic Data. The Beijing Metro Line 5 is 27.6 km long with 23 stations. It covers several commercial centers and links the residential districts in north and south of Beijing (see Figure 10). Table 2 gives the basic information of trains servicing on Beijing Metro Line 5 and Figure 11 shows the traction and regenerative braking characteristics of the train.

The number of passengers on the train in different interstation runs is shown in Figure 12, taking peak hour as an example. It is used to attain train mass, assuming the mass of each person is 60 kg. There are 11 PSIs on Beijing Metro

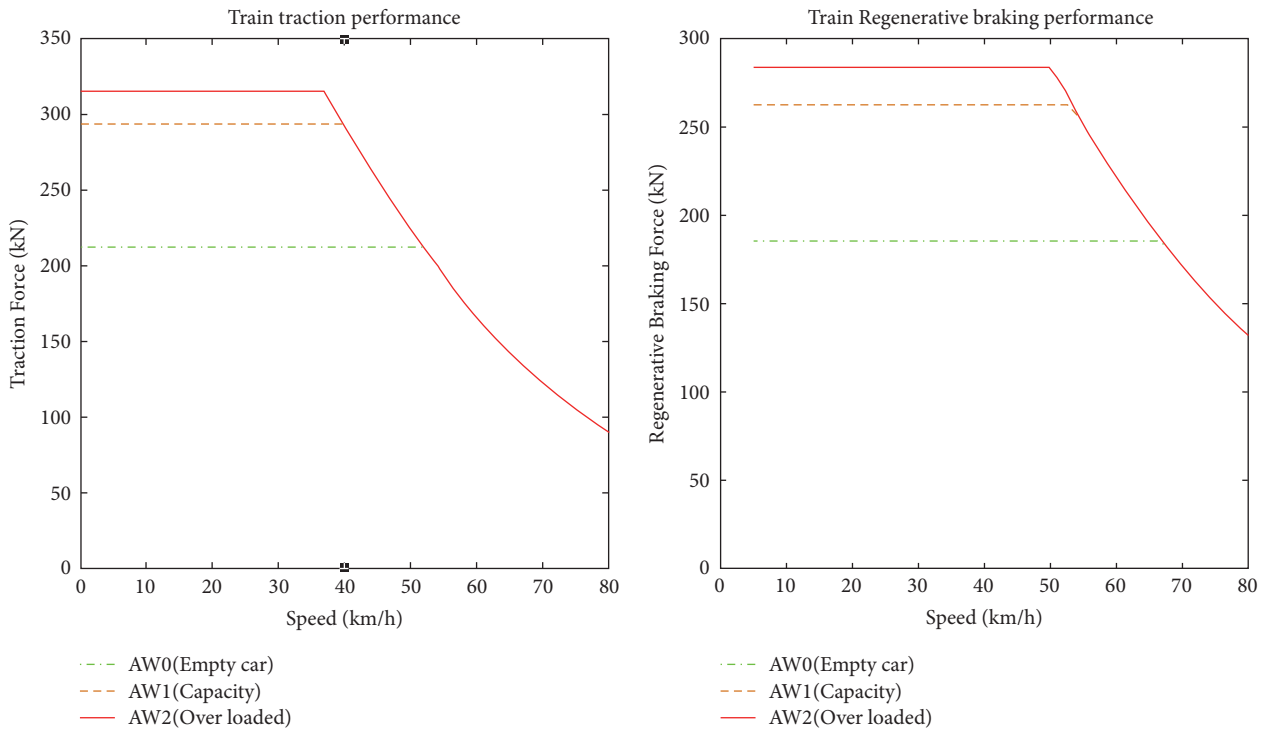


Figure 11: The power performance under different load factors.

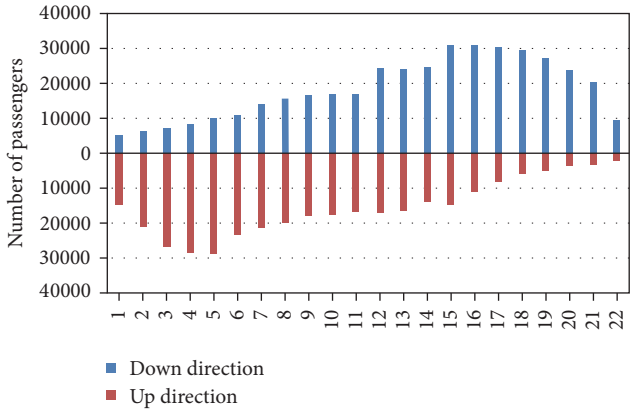


Figure 12: Number of passengers on the train in different interstations.

Table 2: Train characteristics.

Rolling stock mass	203,000 kg
Max speed limit	80 km/h
Number of cars	6 pcs
Maximum acceleration	1.0 m/s^2
Maximum deceleration	–1.0 m/s^2
Minimum speed for RBE	5 km/h
Resistance coefficient a	1.2414
Resistance coefficient b	0.0144
Resistance coefficient c	0.000221

Line 5, and power peak in every PSI is 8500 kW. For safety and quality of service, the maximal headway is 160 s and the minimal headway is 140 s. In the NS-GSA solution algorithm, we set population $P = 100$, $Q = 10$, $P_c = 0.8$, $P_m = 0.1$, the maximum generation as 50, and the annealing rate as 0.95 with an initial temperature of 100°C.

5.2. Energy Performance of Improved Energy-Efficient Train Control. Table 3 gives the energy performance of the improved model (MA-[CR-CO]$_n$-MB) and previous model (MA-CR-CO-MB) with the same runtime, from DTLD to HXXJNK in down direction. The improved train control strategy is able to reduce traction energy consumption by 1.07 kWh in the first interstation with steep downhill slopes and the potential annual traction energy saving in this interstation could be up to 50772 kWh. On the other hand, two driving strategies consumed the same traction energy in the second interstation where there is no steep downhill slope.

Figure 13 demonstrates the train trajectories of two driving strategies to further explain the reason of traction energy reduction by improved control model in a more intuitive way. In the first interstation, the trajectory obtained by the improved control consists of MA-[CR-CO]$_2$-MB; both CR and CO are implemented twice but the second CO is omitted because MB has to be implemented for station stop. In contrast, the trajectory attained by the previous model is sequentially composed of MA, CR, CO, and MB. The improved trajectory extends the coasting distance in steep downhill tracks in order to take full advantage of potential energy to increase the kinetic energy. As such, MA can be

Table 3: Energy performance of the improved train control model.

Interstation	DTLD-HXXJBK (with steep downhills)		HXXJBK-HXXJNK (without a steep downhill)	
Driving strategy	Improved	MA-CR-CO-MB	Improved	MA-CR-CO-MB
Runtime (s)	129	129	94	94
Traction energy consumption (kWh)	6.42	7.49	6.09	6.09

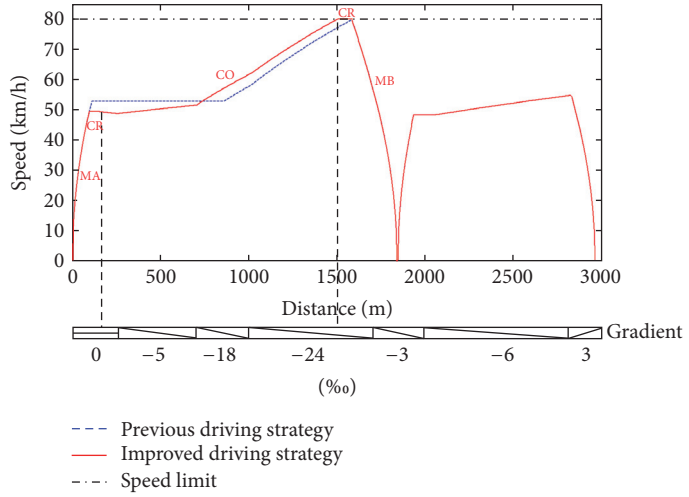


Figure 13: The comparison between two driving strategies.

shortened and the traction energy consumption is saved, while the interstation runtime remains the same. In the second interstation, two control strategies obtained the same trajectory as well as the traction energy consumption, as there is no steep downhill track in this interstation.

It can be concluded that the traction energy performance of the improved control strategy is better than or equal to the MA-CR-CO-MB strategy. Additionally, runtime supplements in each interstation run have significant influences on train trajectory optimization. We studied the energy-saving rates of the improved train control in comparison with the previous model under different runtimes. Figure 14 shows traction energy consumption of two control strategies in interstation DTLD-HXXJNK with different runtimes.

When the scheduled runtime is equal to the minimum interstation runtime, the train has to accelerate to speed limit using maximum traction force and keep the speed constant to satisfy the runtime constraint until the train brakes for station stop. Under this condition, two control strategies attained the same trajectory and consumed the same energy because there is no room to apply CO. Along with the increment of runtime supplements, the improved train control can realize increased energy saving because more CR is allowed in downhill sections, which reduces the traction energy consumption by taking advantage of potential energy. Once the interstation runtime reaches a certain level, the energy-saving ratio decreases because the motoring distances in both control strategies decrease. In case the scheduled runtime is large enough, the trajectories as well as the energy consumption of the two models are the same.

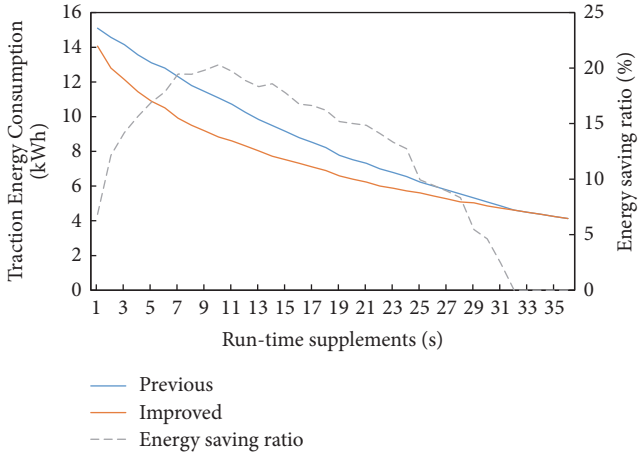


Figure 14: Energy-saving effect of the improved driving strategy at different runtimes.

In addition to interstation runtime and track profile, train mass also has a great impact on traction energy consumption because a large mass results in heavier resistance during the interstation runs and also helps the train accelerate in downhill slopes. We used the proposed control strategy and the one applied in previous studies to obtain the optimal trajectories in the interstation DTLD-HXXJBK, with different train masses and different interstation runtimes. The traction energy reduction by the improved control in comparison with the previous model is shown in Figure 15. It is found that

Table 4: The original timetable and the optimized timetable.

ID	Interstation	Length (m)	Up direction			Down direction			PSI
			Runtime (s)		Dwell time (s)	Runtime (s)		Dwell time (s)	
			Original	Optimized		Original	Optimized		
1	SJZ-LJY	1671	126	122	45	130	123		1
2	LJY-PHY	905	81	92	30	82	88	41	
3	PHY-TTDM	1900	133	140	30	133	127	30	2
4	TTDM-CQK	1183	98	96	45	97	92	30	3
5	CQK-CWM	877	80	80	50	79	84	45	
6	CWM-DD	822	79	92	60	79	75	50	
7	DD-DSK	945	83	93	30	82	88	60	4
8	DSK-DS	848	81	82	45	81	88	30	
9	DS-ZZZL	1017	89	86	30	87	84	45	5
10	ZZZL-BXQ	791	78	82	30	74	75	30	
11	BXQ-YHG	866	81	88	50	85	86	30	
12	YHG-HPLBJ	1151	98	96	30	103	91	50	6
13	HPLBJ-HPXQ	1059	93	90	30	93	91	30	
14	HPXQ-HXXJNK	1025	89	87	55	91	87	30	7
15	HXXKNK-HXXJBK	1122	99	92	30	94	97	55	
16	HXXJBK-DTLD	1838	133	137	40	129	127	30	8
17	DTLD-BYLB	3000	180	180	30	181	181	40	9
18	BYLB-LSQN	1286	110	110	30	108	100	30	10
19	LSQN-LSQ	1306	111	109	50	112	112	30	
20	LSQ-TTYN	1544	120	109	30	119	111	50	11
21	TTYN-TTY	966	86	93	30	86	84	30	
22	TTY-TTYB	939	85	94		86	85	40	
Total		27061	2213	2250	800	2211	2176	806	

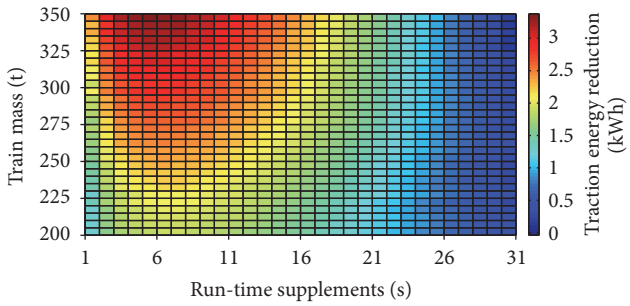


Figure 15: Traction energy reduction under various conditions using improved driving strategies.

the improved model achieves better energy saving when the train mass increases. The reason is that heavier trains acquire greater acceleration when coasting in steep downhills, and they can gain higher kinetic energy in steep downhill tracks and achieve better energy saving.

To sum up, the improved model is a supplement to the MA-CR-CO-MB control strategy in interstations with constant speed limit and it is able to save traction energy in downhill sections by taking advantage of potential energy. Energy saving is influenced by the scheduled runtime and train mass. The energy saving firstly increases with scheduled runtime and then decreases, finally reaching zero when the runtime is sufficient. Heavier trains can achieve a better traction energy reduction because the acceleration is larger when the train is coasting in downhills where the potential energy could help the train more in accelerating.

5.3. Energy Performance of Integrated Optimization. Separate train control to minimize the traction energy consumption of each train does not necessarily lead to the minimum net energy consumption of whole metro lines. The energy consumption of the whole metro lines can be further reduced by optimizing train control and timetable concurrently. Table 4 shows the original timetable and train control configurations (RT, HD, k_t, and k_b) provided by operators and the optimized one obtained by the proposed integrated optimization model. The original journey time is 2213 s in the up direction and 2211 s in the down direction. Compared with the original timetable, the optimized timetable increases by 37 s in the up direction and decreases by 35 s in the down direction. Therefore, the cycle time is extended by 2 s, which is acceptable. Additionally, train headway is regulated to improve the utilization of RBE. The headway is set as a fixed value (i.e., 150 s) in the original timetable. In the optimized one, the headway between every two adjacent trains varies within a reasonable range from 140 s to 159 s, without changing the number of service provisions and dwell time at each station.

Table 5: Comparison of energy consumption between the original timetable and optimized timetable.

	Original timetable	Optimized timetable	
$[k_t, k_b]$	[1, 1]	[0.9, 0.9]	
Net energy consumption (kWh)	15059	14311	(−4.97%)
Auxiliary devices (kWh)	530	530	(+0.00%)
Traction energy (kWh)	16567	17554	(+5.96%)
Total RBE (kWh)	13016	14360	(+10.33%)
Utilized RBE (kWh)			
Auxiliary (self)	52	57	(+9.62%)
Traction	1840	3570	(+94.02%)
Auxiliary (other)	146	147	(+0.68%)
Total	2038	3773	(+85.13%)
Overlapping time (s)	4194	4543	(+8.32%)
Utilization percentage of RBE (%)	15.66	26.27	(+67.81%)

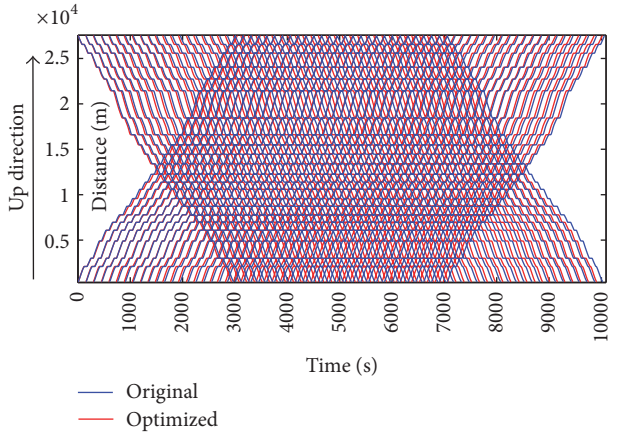


Figure 16: Comparison of the original and optimal timetable.

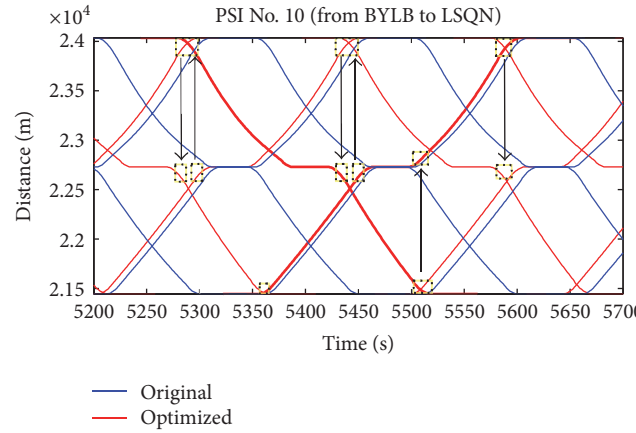


Figure 17: The utilization of the regenerative braking energy in the same power supply interval.

Table 5 shows the comparison of energy consumption between the original timetable with previous train control model and the optimized timetable with the improved train control model. Although traction energy consumption with the optimized timetable increases by 5.96%, the net energy consumption of the whole line decreases by 4.97% because of the significant improvements on utilization of RBE. This verifies the argument that minimum traction energy consumption does not necessarily lead to the minimum net energy consumption. Although partial traction and braking result in a slight increment in traction energy consumption, the utilization of RBE with the optimized timetable increases by 67.81% as the overlapping time increases by 8.32%. It is concluded that the integrated optimization model can notably improve the utilization of RBE, thus reducing the net energy consumption.

Figure 16 shows the difference between the original timetable and the optimized timetable. The total journey time and the operating time for service provision remain the same in the optimized timetable, although the interstation runtime (RT) and train headway (HD) are regulated slightly.

An enlarged view of part of Figure 16, taking a section from the station of BYLB to the station of LSQ in the same PSI as an example, is shown in Figure 17 to explicate how the utilization of RBE is optimized by the integrated model. The arrows in the figure illustrate the flow of regenerative braking energy; that is, the RBE generated by the braking train can be transmitted to the traction train. It is found that motoring and braking of trains in the same PSI could be synchronized by adjusting headway and interstation runtime in timetable formulation. Moreover, overlapping time could be further extended using partial traction/braking to prolong the motoring and braking time. As a result, the net energy consumption decreases, although the traction energy consumption is not the minimal one.

Figure 18 gives the trajectories of two successive trains moving in the same PSI. Under the original timetable with full motoring and braking control, the RBE is not used timely because motoring train and braking train are not synchronized well. By optimizing headway and interstation runtime, the departure of train 2 from DD station just located in the same time window when train 1 arrives at the TTDM station. Furthermore, the synchronized time between traction and braking train is prolonged by applying partial motoring and braking, which enable a larger overlapping

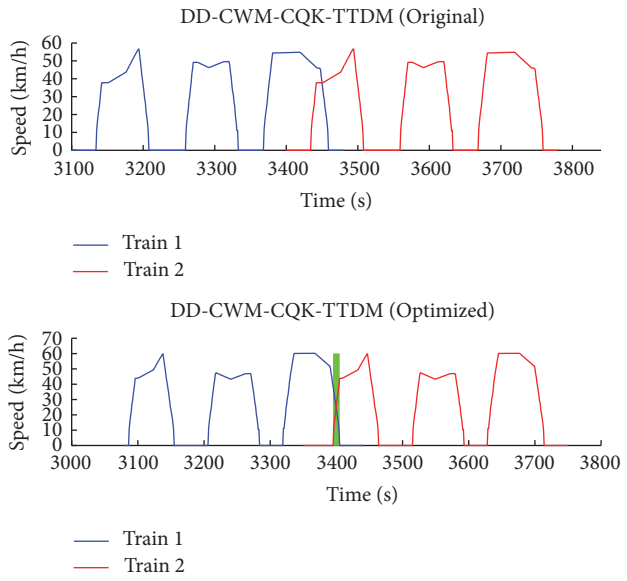


Figure 18: Synchronized traction and braking train in one power supply interval.

time as well as a better utilization of RBE. Thus, most of the RBE produced by train 1 is timely used by train 2, which helps to reduce the energy feeding from substations and in turn save the net energy consumption of the whole metro line.

6. Conclusion

An integrated optimization model on train control and timetable formulation is proposed in this study to minimize the net energy consumption of all trains servicing on the metro line. The proposed train control is still based on finding the optimal switching points among the control modes of MA, CR, CO, and MB to minimize the traction energy consumption, while cruising and coasting regimes might be adopted for more than one time according to track profiles. For better utilization of RBE, timetable configurations such as interstation runtime and train headway as well as the extents of motoring and braking in train control are optimized concurrently, taking into account the synchronization of motoring and braking trains in the same PSI. Practical operation condition and constraints, such as varied train mass in different interstations, the limitation on maximal loading of power system, are taken into consideration in the proposed model. The brute force algorithm is employed to attain the energy-efficient trajectory of train interstation runs and an NS-GSA algorithm is developed to attain the optimal extents of motoring/braking and timetable configurations.

Case studies on Beijing Metro Line 5 demonstrated the energy performance of the proposed integrated model. It is found that the improved train control can reduce the traction energy by 20% in the interstations with steep downhill slopes. The energy-saving rate of the improved train control in comparison with the previous model depends on track profiles and the scheduled runtime. The integrated optimization on train control and timetable formulation is able to save net energy consumption by 4.97% through regulating interstation runtime, train headway, and the extents of motoring and braking in train control. Although the traction energy slightly increases by 5.96% because of the application of partial motoring and braking in train control, the utilization of RBE is significantly improved by prolonging the overlapping time of motoring and braking trains in the same PSI.

This paper explored the energy-efficient train control with the constant speed limit, while the speed limit may vary during interstation runs due to curves or temporary maintenance. In such case, the proposed control strategy might not be the optimal one. One of our future researches is to take the varied speed limits into consideration in train control and apply a more general energy-efficient train control model in timetable optimization. Additionally, this study assumes that all trains have the same extents of motoring and braking in different interstation runs. Different extents of motoring and braking for different trains in different interstation runs might lead to more significant energy savings, which is another direction of future research.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71390332, 71621001, and 71571016) and the Fundamental Research Funds for the Central Universities (2016JBM031).

References

[1] A. González-Gil, R. Palacin, P. Batty, and J. P. Powell, "A systems approach to reduce urban rail energy consumption," *Energy Conversion and Management*, vol. 80, pp. 509–524, 2014.

[2] K. Ichikawa, "Application of optimization theory for bounded state variable problems to the operation of train," *Journal of the Japan Society of Mechanical Engineers*, vol. 11, no. 47, pp. 857–865, 1968.

[3] H. Strobel, P. Horn, and M. Kosemund, "A contribution to optimum computer-aided control of train operation," in *Proceedings of the 2nd FAC/IFIP/IFORS Symposium on Traffic Control and Transportation Systems*, pp. 377–387, 1974.

[4] I. P. Milroy, *Aspect of automatics train control [Ph.D. dissertation]*, Loughborough University of Technology, Sydney, Australia, 1980.

[5] B. R. Benjamin, A. M. Long, I. P. Milroy, R. L. Payne, and P. J. Pudney, "Control of railway vehicles for energy conservation and improved timekeeping," in *Proceedings of the Conference on Railway Engineering*, pp. 41–47, Perth, Australia, 1987.

[6] P. G. Howlett, I. P. Milroy, and P. J. Pudney, "Energy-efficient train control," *Control Engineering Practice*, vol. 2, no. 2, pp. 193–200, 1994.

[7] E. Khmelnitsky, "On an optimal control problem of train operation," *IEEE Transactions on Automatic Control*, vol. 45, no. 7, pp. 1257–1266, 2000.

[8] R. Liu and I. M. Golovitcher, "Energy-efficient operation of rail vehicles," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 917–932, 2003.

[9] L. A. M. V. Dongen and J. H. Schuit, "Energy-efficient driving patterns in electric railway traction," in *Proceedings of the International Conference on Main Line Railway Electrification (IET '02)*, pp. 154–158, 2002.

[10] G. M. Scheepmaker and R. M. P. Goverde, "The interplay between energy-efficient train control and scheduled running time supplements," *Journal of Rail Transport Planning and Management*, vol. 5, no. 4, pp. 225–239, 2015.

[11] P. G. Howlett, P. J. Pudney, and X. Vu, "Local energy minimization in optimal train control," *Automatica*, vol. 45, no. 11, pp. 2692–2698, 2009.

[12] Y. Bai, B. H. Mao, F. M. Zhou, Y. Ding, and C. B. Dong, "Energy-efficient driving strategy for freight trains based on power consumption analysis," *Journal of Transportation Systems Engineering & Information Technology*, vol. 9, no. 3, pp. 43–50, 2009.

[13] M. Domínguez, A. Fernández-Cardador, A. P. Cucala, T. Gonsalves, and A. Fernández, "Multi objective particle swarm optimization algorithm for the design of efficient ATO speed profiles in metro lines," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 43–53, 2014.

[14] W. Carvajal-Carreño, A. P. Cucala, and A. Fernández-Cardador, "Optimal design of energy-efficient ATO CBTC driving for metro lines based on NSGA-II with fuzzy parameters," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 164–177, 2014.

[15] C. Sicre, A. P. Cucala, and A. Fernández-Cardador, "Real time regulation of efficient driving of high speed trains based on a genetic algorithm and a fuzzy model of manual driving," *Engineering Applications of Artificial Intelligence*, vol. 29, no. 29, pp. 79–92, 2014.

[16] K. Keskin and A. Karamancioglu, "Energy-efficient train operation using nature-inspired algorithms," *Journal of Advanced Transportation*, vol. 2017, Article ID 6173795, 12 pages, 2017.

[17] W. Shangguan, X.-H. Yan, B.-G. Cai, and J. Wang, "Multiobjective optimization for train speed trajectory in CTCS high-speed railway with hybrid evolutionary algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2215–2225, 2015.

[18] T. Albrecht and S. Oettich, *A New Integrated Approach to Dynamic Schedule Synchronization and Energy-Saving Train Control*, WIT Press, Southampton, UK, 2002.

[19] C. Sicre, P. Cucala, A. Fernández, J. A. Jiménez, I. Ribera, and A. Serrano, "A method to optimise train energy consumption combining manual energy efficient driving and scheduling," in *Proceedings of the 12th International Conference on Computer System Design and Operation in the Railways and other Transit Systems (COMPRAIL '10)*, pp. 549–560, September 2010.

[20] R. Chevrier, P. Pellegrini, and J. Rodriguez, "Energy saving in railway timetabling: a bi-objective evolutionary approach for computing alternative running times," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 20–41, 2013.

[21] G. M. Scheepmaker, R. M. Goverde, and L. Kroon, "Review of energy-efficient train control and timetabling," *European Journal of Operational Research*, vol. 257, no. 2, pp. 355–376, 2017.

[22] R. Barrero, J. van Mierlo, and X. Tackoen, "Energy savings in public transport," *IEEE Vehicular Technology Magazine*, vol. 3, no. 3, pp. 26–36, 2008.

[23] T. Albrecht, "Reducing power peaks and energy consumption in rail transit systems by simultaneous train running time control," *Computers in Railways*, pp. 885–894, 2004.

[24] X. Yang, X. Li, Z. Gao, H. Wang, and T. Tang, "A cooperative scheduling model for timetable optimization in subway systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 438–447, 2013.

[25] X. Yang, B. Ning, X. Li, and T. Tang, "A two-objective timetable optimization model in subway systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1913–1921, 2014.

[26] Z. Le, K. Li, J. Ye J, and X. Xu, "Optimizing the train timetable for a subway system," *Proceedings of the Institution of Mechanical Engineers Part F: Journal of Rail & Rapid Transit*, vol. 229, no. 8, pp. 2532–2542, 2015.

[27] M. Peña-Alcaraz, A. Fernández, A. P. Cucala, A. Ramos, and R. R. Pecharromán, "Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 226, no. 4, pp. 397–408, 2011.

[28] Y. Ding, H. D. Liu, Y. Bai, and F. M. Zhou, "A two-level optimization model and algorithm for energy-efficient urban train operation," *Journal of Transportation Systems Engineering & Information Technology*, vol. 11, no. 1, pp. 96–101, 2011.

[29] S. Su, X. Li, T. Tang, and Z. Gao, "A subway train timetable optimization approach based on energy-efficient operation strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 883–893, 2013.

[30] S. Su, T. Tang, X. Li, and Z. Gao, "Optimization of multitrain operations in a subway system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 673–684, 2014.

[31] X. Yang, A. Chen, B. Ning, and T. Tang, "A stochastic model for the integrated optimization on metro timetable and speed profile with uncertain train mass," *Transportation Research Part B: Methodological*, vol. 91, pp. 424–445, 2016.

[32] X. Li and H. K. Lo, "An energy-efficient scheduling and speed control approach for metro rail operations," *Transportation Research Part B: Methodological*, vol. 64, pp. 73–89, 2014.

[33] N. Zhao, C. Roberts, S. Hillmansen, Z. Tian, P. Weston, and L. Chen, "An integrated metro operation optimization to minimize energy consumption," *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 168–182, 2017.

[34] S. Su, T. Tang, and Y. H. Wang, "Evaluation of strategies to reducing traction energy consumption of metro systems using an optimal train control simulation model," *Energies*, vol. 9, no. 2, pp. 105–123, 2016.

[35] S. Watanabe and T. Koseki, "Energy-saving train scheduling diagram for automatically operated electric railway," *Journal of Rail Transport Planning and Management*, vol. 5, no. 3, pp. 183–193, 2015.

[36] M. Miyatake and H. Ko, "Optimization of train speed profile for minimum energy consumption," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 5, no. 3, pp. 263–269, 2010.

[37] A. Albrecht, P. Howlett, P. Pudney, X. Vu, P. Zhou, and D. Chen, "Using maximum power to save energy," in *Proceedings of the 17th IEEE International Conference on Intelligent Transportation Systems*, pp. 1205–1208, Qingdao, China, November 2014.

[38] Á. J. López-López, R. R. Pecharromán, A. Fernández-Cardador, and A. P. Cucala, "Assessment of energy-saving techniques in direct-current-electrified mass transit systems," *Transportation Research Part C: Emerging Technologies*, vol. 38, pp. 85–100, 2014.

[39] M. Domínguez, A. Fernández, A. P. Cucala, and P. Lukaszewicz, "Optimal design of metro automatic train operation speed

profiles for reducing energy consumption," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 225, no. 5, pp. 463–474, 2011.

[40] A. P. Cucala, A. Fernández, C. Sicre, and M. Domínguez, "Fuzzy optimal schedule of high speed train operation to minimize energy consumption with uncertain delays and drivers behavioral response," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 8, pp. 1548–1557, 2012.

[41] X. Yang, X. Li, B. Ning, and T. Tang, "An optimisation method for train scheduling with minimum energy consumption and travel time in metro rail systems," *Transportmetrica B: Transport Dynamics*, vol. 3, no. 2, pp. 79–98, 2015.

[42] P. Shrivastava, S. L. Dhingra, and P. J. Gundaliya, "Application of genetic algorithm for scheduling and schedule coordination problems," *Journal of Advanced Transportation*, vol. 36, no. 1, pp. 23–41, 2002.

[43] B. Fan, C. Roberts, and P. Weston, "A comparison of algorithms for minimising delay costs in disturbed railway traffic scenarios," *Journal of Rail Transport Planning and Management*, vol. 2, no. 1-2, pp. 23–33, 2012.

[44] N. Zhao, C. Roberts, S. Hillmansen, and G. Nicholson, "A multiple train trajectory optimization to minimize energy consumption and delay," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2363–2372, 2015.

[45] S. Yang, K. Yang, Z. Gao, L. Yang, and J. Shi, "Last-train timetabling under transfer demand uncertainty: mean-variance model and heuristic solution," *Journal of Advanced Transportation*, vol. 2017, Article ID 5095021, 13 pages, 2017.

[46] M. Samà, A. D'Ariano, F. Corman, and D. Pacciarelli, "A variable neighbourhood search for fast train scheduling and routing during disturbed railway traffic situations," *Computers & Operations Research*, vol. 78, pp. 480–499, 2017.

[47] A. D'Ariano, F. Corman, D. Pacciarelli, and M. Pranzo, "Reordering and local rerouting strategies to manage train traffic in real time," *Transportation Science*, vol. 42, no. 4, pp. 405–419, 2008.

[48] P. Pellegrini and M. Birattari, "Implementation effort and performance," in *Engineering Stochastic Local Search Algorithms. Designing, Implementing and Analyzing Effective Heuristics, International Workshop, SLS 2007, Brussels, Belgium, September 2007*, Springer, 2007.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 6318516, 12 pages
https://doi.org/10.1155/2018/6318516

Research Article

Calculation Method for Load Capacity of Urban Rail Transit Station considering Cascading Failure

Jiajun Huang, Feng Zhou, and Mengru Xi

The Key Laboratory of Road and Traffic Engineering of the State Ministry of Education, Tongji University, 4800 Cao'an Road, Jiading District, Shanghai 201804, China

Correspondence should be addressed to Feng Zhou; zhoufeng24@tongji.edu.cn

Received 24 November 2017; Accepted 1 March 2018; Published 26 April 2018

Academic Editor: Paola Pellegrini

The load capacity of urban rail transit station is of great significance to provide reference in station design and operation management. However, it is difficult to carry out quantitative calculation quickly and accurately due to the complex interaction among passenger behaviors, facility layout, and the limit capacity of single facility. In this paper, the association network of facilities is set up based on the analysis of passenger service chain in station. Then the concept of cascading failure is introduced to the dynamic calculation model of load capacity, which is established on the user-equilibrium allocation model. The solution algorithm is optimized with node attack strategy of complex network to effectively reduce the computational complexity. Finally, a case study of Lujiabang Road Station in Shanghai is carried out and compared with the simulation results of StaPass, verifying the feasibility of this approach. The proposed method can not only search for the bottleneck of capacity, but also help to trace the loading variation of facilities network in different scenarios, providing theoretical supports on passenger flow organization.

1. Introduction

The networked process of urban rail transit (URT) in China's major cities has been expedited, and the demand of passenger flow is further unleashed. It brings great challenge on operational security, transport capacity and efficiency, service level, and other aspects to URT management department. Taking Shanghai as an example, in 2016, the whole rail transit system provided service to over 9 million passengers every weekday averagely, the extreme passenger volume hit a new record by exceeding 10 million people, and over 10 metro stations handled more than 100,000 passengers every day.

In the context of this, pressures are increased on URT stations which served as the basic operation unit of URT and the distribution hub for passenger. Prominent problems emerge from the daily work of passenger organization, which can be summarized as follows:

(1) The design load capacity of some stations does not match with the actual passenger flow, resulting in the increasing risk of emergencies like passenger stranded, severe congestion, and so on.

(2) Varying levels of inflow-limiting strategies have been carried out in order to coordinate mass passenger flow, however lacking the quantitative basis for formulation.

(3) The specific treatments are usually formulated by the subjective work experience of URT operators in emergency strategy, while disregarding the interrelation of different facilities in stations and failing to take full advantage of it.

The problems mentioned above could be ascribed to the inaccuracy of passenger flow forecasting during the design phase. Yet the underlying reason is the lack of mathematical assessment on the load capacity of URT stations.

The load capacity of URT stations is defined as the quantity of passengers when the passenger services cannot be provided because some key facilities are unavailable or in congestion. At present, static calculation methods for capacity are generally adopted in the design stage of the rail transit station. In China, a Cannikin Law based method is applied to analyze the load capacity of the station, that is, taking the minimum

Table 1: Calculation methods for load capacity of URT station.

Method	Precision	Complexity	Quantified	Comprehensiveness
Queueing Theory	☆☆	☆☆	☆☆☆	Not consider the influence of passenger motor behaviors
Macroscopic Simulation (System Dynamics)	☆	☆☆	☆☆	Not consider the influence of passenger motor behaviors and station layout
Microscopic Simulation (simulation of passenger motor process)	☆☆	☆☆☆	☆	Basically comprehensive

value as the overall load capacity from all facilities and equipment whose maximum capacities have already stipulated by national standard [1]. Some European urban rail transit, such as the London Subway [2], divided service quality into several levels with the consideration of passenger characteristics. The design work and capacity assessment of the station is carried out under the guidance of service and safety level. However, URT station is a complex system consisting of various types of facility, providing multiple routes for passengers to reach their destination, and passenger motional characteristics are closely related to the layout of facilities and equipment in station. Thus static or discrete calculation for capacity would be a straightforward solution, but it is not feasible in reflecting the load capacity of station in practice.

In the case of dynamic methods, Queueing Theory and system simulation are methods mostly used to evaluate the load capacity of rail transit station (Table 1). Approaches based on Queueing Theory establish particular congestion state-dependent queueing model [3–5] for each facility in station such as gates, staircases, and corridors according to the analysis of the passenger flow characteristics, then modeling the M/G/C/C state-dependent queueing network [6, 7] in a systematic way. It takes the coordination between capacities of different facilities into account, neglecting the dynamic impact exerted on the load capacity of station when passengers make choice on routing. The system simulation method is to simulate passengers' motion in the urban rail transit station through specific models or tools, which are separated into two large fields of microscopic and macroscopic researches. The latter commonly regard the station as a dynamics system and models with diverse theories, including mixed Petri net [8, 9] and system dynamics [10–12]. But the model fails to consider the influence of the facilities layout on the load capacity. At the microscopic level, cellular automata [13], social force model [14], potential field [15], and other approaches are used to simulate individual behaviors; meanwhile some commercial pedestrian simulation software programs like Legion, Step, StaPass, and so on are also applied to search the bottleneck of station capacity. They can evaluate the load capacity of URT station in different scenarios, but have many defects such as the complexity of modeling and time-consuming simulation.

Since the load capacity of URT station is not only restricted to the capacity of single piece of equipment but also influenced by passenger behaviors and the layout of facilities in the station, it is insufficient to calculate the accurate load capacity if only considering one of these factors. The review of the literature indicates the necessity to develop a novel method for load capacity calculation. In URT station, passengers receive the service from a series of facilities having strong interrelation with each other. Though the capacity limit of a single facility does not necessarily represent the vulnerability of the whole station, the cascading failure properties of the network composed of all facilities can lead to congestion [16]. This is quite similar to the dynamics of network flow in the traffic system.

Therefore, this paper sets up the association network of facilities and its passenger flow assignment mechanism considering the cascading failure effect. On the basis of the user-equilibrium assignment model, we propose a dynamic method to calculate the load capacity of URT station. It is able to trace the loading variation of facilities network, search for the bottleneck of load capacity, and provide staff with the theoretical support on passenger flow organization.

The rest of this paper is organized as follows. We analyze the service chain of passenger flow in the URT station and set up the association network of facilities and its property in Section 2. Section 3 provides the methodology to assign passenger flow with the cascading failure effect and the user-equilibrium assignment model considering passenger choice behavior is presented. Then the solution algorithm that combines node attack strategy with Frank-Wolfe algorithm is given in Section 4. Afterwards, a case study on real-world station is expatiated with the comparison to pedestrian simulation in Section 5. Finally, conclusions and future research are discussed.

2. Association Network of Facilities in URT Station

In this section, we propose the association network of facilities in URT station based on the service chain of passenger flow.

2.1. Service Chain of Passenger Flow. Passenger flow in URT stations can be classified into three distinctive categories, that is, ingress passenger flow, egress passenger flow, and transfer passenger flow. And the gathering and distributing process of passenger is denoted as receiving specific services from a series of facilities successively, which is the definition of "service chain" in the URT station. Indeed, the service chain varies with the type of passenger flow [17]:

(1) Ingress passenger flow: enter station → purchase tickets → check tickets → go through staircases or escalators (if station hall and platform are on different floors) → wait for the train and board

(2) Egress passenger flow: alight → go through staircases or escalators (if station hall and platform are on different floors) → check tickets → exit station

(3) Transfer passenger flow: alight → go through staircases or escalators → check tickets (if necessary) → walk to another platform (through corridors, staircases, or escalators) → wait for the train and board.

According to the description of the service chain, each service is provided by one kind of facility, including corridors. To be noteworthy, train is not strictly the facility that belongs to the station, but it is the only server in the event of boarding and alighting. Thus we regard the train as one piece of equipment in this paper. Consequently, the service chain can be translated into the facility chain in station, shown in Figure 1.

2.2. Association Network of Facilities. The motion of passenger flows brings forth the coupling between facilities, and the facility chain in station makes it feasible to depict that relationship. In each strand of facility chain, every single node represents a specific piece (or group) of equipment, fusing together and then forming an open-loop and directed association network of facilities.

Explanations of association network are given as follows.

Item 1. A set of nodes D_i ($i \in N^*$) denotes a certain link L_u ($u \in N^*$) in the service chain. In Figure 2, the automatic fare gates (AFG) D_5 to D_8 constitute the facility set of checking tickets service for ingress and egress passenger flow.

Item 2. Directed edge $E_{i\text{-}j}$ indicates the accessibility of the path from node D_i to node D_j ($j \in N^*$, $j \neq i$), while there would not be a directed edge if two nodes are disconnected. $l_{i\text{-}j}$ is the length of directed edge $E_{i\text{-}j}$, defined as the linear distance between midpoints of two connected facilities. Meanwhile, the transition from one link to another is completed via directed edges. Figure 2 shows that $E_{3\text{-}5}$, $E_{3\text{-}7}$, and $E_{4\text{-}7}$ are involved in the process that passenger flow moves from purchasing link L_2 to checking link L_3.

Item 3. C_i is the limit capacity of node D_i, except for those representing entrance and exit, under a certain service level. It is defined as maximum passenger flow which the facility can handle in unit time, quantified with

$$C_i = C_{si} + C_{qi}. \tag{1}$$

The maximum number of people that facility node D_i can serve in unit time without queueing is defined as the maximum service capacity C_{si}, and its formula is given in [1]. C_{qi} is the maximum queueing number in unit time under a certain service level, called maximum queueing capacity, and defined as

$$C_{qi} = S_{\text{queue\_}i} * p_i, \tag{2}$$

where $S_{\text{queue\_}i}$ is the size of queueing area for facility node and p_i is the number of passengers per unit area, which suggested quantifying with Fruin level of service (LOS) [18] in this study.

Item 4. Let VI_i denote the inflow volume of node D_i and VO_j denote the outflow volume of node D_j. Then VI_i should satisfy (3) if node D_j is the former point that connected to node D_i at steady state.

$$\text{VI}_i = \sum_{j \to i} \text{VO}_j. \tag{3}$$

For each node D_i, there is an upper limit to how many passengers could be handled. Thus the outflow volume of node D_i is supposed to be updated according to the inflow volume VI_i which is in the same flow direction.

If the inflow volume of node D_i exceeds its maximum service capacity, only part of passengers can move to the next node while the rest of them are counted as queueing volume VQ_i. Moreover, if VQ_i is beyond the maximum queueing capacity, node D_i is considered as overloaded. Whereas the inflow volume of node D_i is less than the maximum service capacity, all passengers receive service in time and leave the node. Then the formula for outflow volume VO_i is given by

$$\text{VO}_i = \begin{cases} C_{si}, \text{VQ}_i = \text{VI}_i - C_{si}, & \text{if } \text{VI}_i > C_{si} \\ \text{VI}_i, \text{VQ}_i = 0, & \text{if } \text{VI}_i \leq C_{si}. \end{cases} \tag{4}$$

3. Methodology

Cascading failure is a failure in a system of interconnected parts where the crash of one part can trigger the failure of successive parts [19]. In a similar way, if the inflow volume is far beyond the maximum capacity of the facility in URT station, failure occurs and there is a call for the reassignment of passenger flow. Yet the crash of one node in the association network of facilities will not alter the network structure, either the volume or distribution of that node remains in the collapsing state. Only passenger flows on other nodes will be reallocated and trigger the crash of vulnerable nodes.

The failure will radiate from the crash node successively until the station can no longer maintain the service chain of any type of passenger flow. In other words, the load capacity of the station is the quantity of passengers when all nodes in one link of the facility chain have collapsed. Hence, the model in this section elaborates on the mechanism of assigning passenger flow to the facility network.

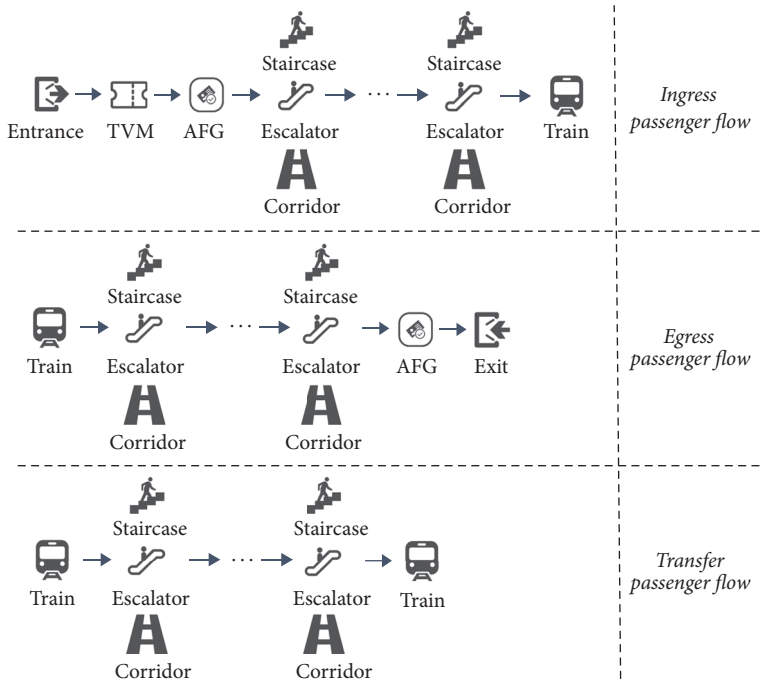


Figure 1: Illustration for facility chains in station.

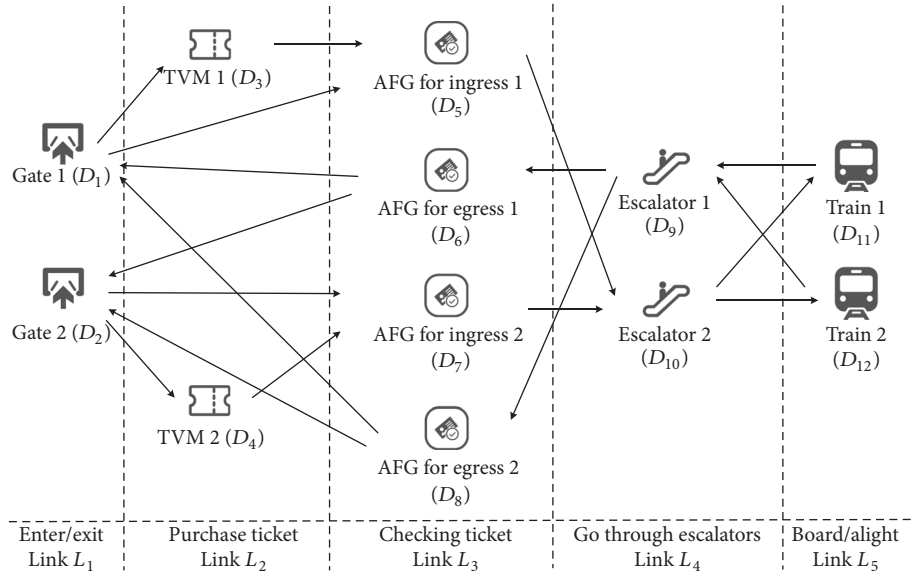


Figure 2: The association network of facilities in nontransfer station.

3.1. Assumptions. Due to the complex interaction among passenger behaviors, facility layout, and single facility capacity, assumptions are proposed as follows to ensure a high computational efficiency and the appropriate accuracy of the model.

(1) The capacity of directed edge $E_{i\text{-}j}$ is unlimited, which means the walking space of accessible paths from node D_i to node D_j is able to accommodate infinite passengers. This term is supposed since the load capacity of opening area is usually much higher than other facilities. Meanwhile, it can be improved by avoiding intercross and queueing in serpentine line in either the designing or operation phase.

(2) The maximum load capacity is an intrinsic property of the facility (node D_i). It does not change with the service intensity and time.

(3) It is assumed that velocity of passenger flow remains constant, taking no account of individual attributes, motor process, and the loss of speed.

(4) The service of boarding and alighting is available when trains arrive periodically. It needs to be

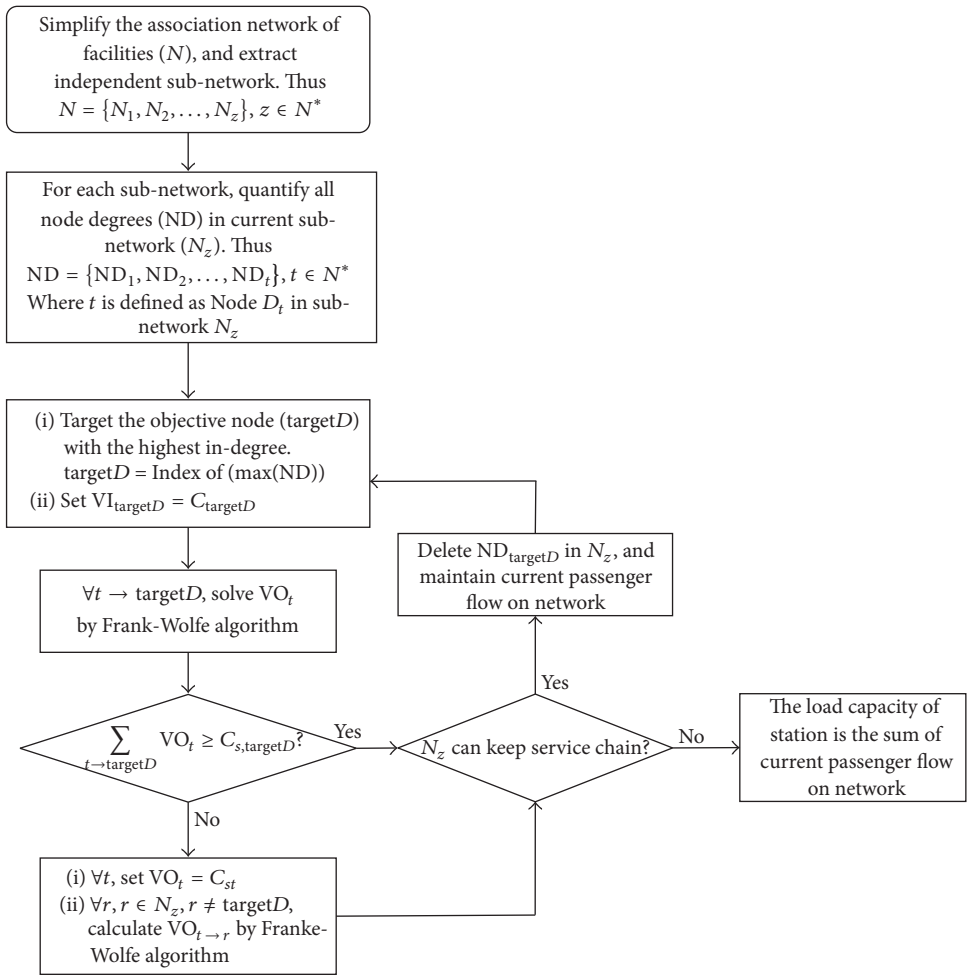


Figure 3: The flow chart of algorithm for calculating load capacity.

converted into the passenger flow volume handled per unit time so that all fundamental capacities of facilities are in uniform dimension.

(5) Passengers need to decide on which node to choose next every time before moving to the successive link of service chain. Once the choice is determined, it is unable to be reselected.

(6) The action of leaving the node is completed in a moment; that is, passengers will not be stranded in the node after receiving the service.

(7) All individual motor processes in the same passenger flow category are viewed as a whole fluid motion. On this basis, we suppose that passengers arrive at the same time and receive service from different nodes in one link.

3.2. Passenger Flow Assignment Mechanism. Commuter and residents are the majority of URT users, having a command of the layout of facilities and equipment in stations. The frequent trip by URT enables these passengers to acquire the guidance information in a short time. Therefore, it is reasonable to assume that passengers make decisions on which route to take with a complete knowledge of information in station.

The user-equilibrium (UE) model [20, 21] is a typical method for traffic assignment. It is based on the fact that people choose a route so as to minimize their travel time and on the assumption that such a behavior on the individual level creates an equilibrium on the network. In this paper, the flow loading on the association network of facilities and equipment is described by the UE model.

Let x_{ij} denote the flow volume on directed edge $E_{i\text{-}j}$. And the impedance function is defined as $w_{ij}(x)$ to quantify the choice behavior of passengers. f_k^{uv} is the flow volume of edge k ($k \in N^*$) between links L_u and L_v ($v \in N^*$, $v \neq u$), while d_{uv} denotes the total flow of links (L_u, L_v). Then the UE model is formulated as follows:

$$\min \quad Z(X) = \sum_{(i,j)} \int_0^{x_{ij}} w_{ij}(x)\, dx \tag{5}$$

$$\text{st.} \quad \sum_k f_k^{uv} = d_{uv} \quad \forall u, v \tag{6}$$

$$f_k^{uv} \geq 0 \quad \forall u, v, k \tag{7}$$

$$x_{ij} = f_k^{uv} \delta_{ij,k}^{uv} \quad \forall i, j. \tag{8}$$

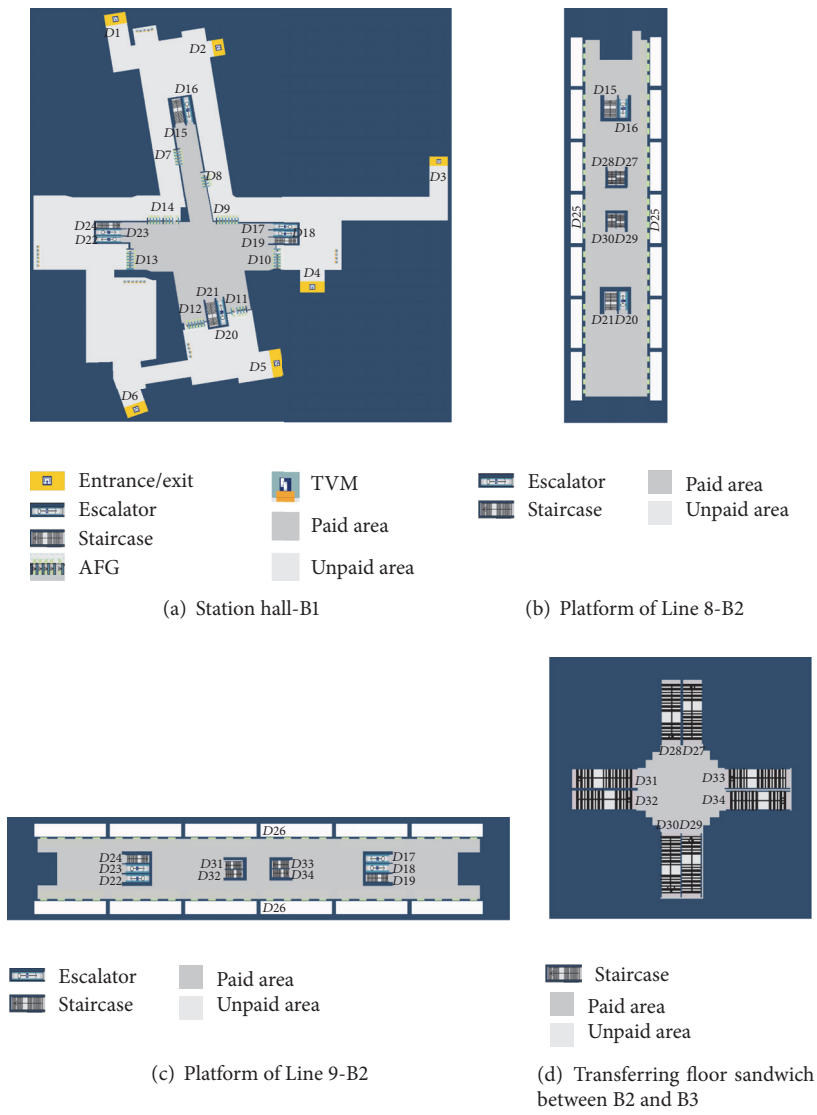


Figure 4: Sketches of LJB station: (a) station hall, (b) platform of Line 8, (c) platform of Line 9, and (d) interlayer for transfer.

If the directed edge is the right edge k between links (L_u, L_v), $\delta_{ij,k}^{uv} = 1$ in (8). Otherwise, $\delta_{ij,k}^{uv} = 0$.

3.3. Choice Behavior on Nodes. Actually, a multitude of passengers prefer to choose the node which is characterized by short distance and convenient service [17]. Thus we take distance, number of people, and congestion into account and formulate the impedance function $w_{ij}(x)$ in Section 3.2, using BPR function as reference.

$$w_{ij}(x) = w_{ij}^0 \left[1 + a_0 \left(\frac{x_{ij}}{C_j}\right)^b + a_1 * \frac{\left(x_{ij} - C_{sj}\right)}{C_{qj}}\right] \quad (9)$$

$$a_0, b, a_1 \geq 0.$$

Equation (9) is composed of the following elements.

Firstly, the distance impedance from node D_i to node D_j indicates the initial impedance of a node before being selected, denoted as w_{ij}^0. It is an innate property of the node, positively associated with $l_{i\text{-}j}$, given in

$$w_{ij}^0 = e^{h * l_{i\text{-}j}}, \quad h > 0, \quad (10)$$

where h has relationship with the scale of URT station (0.1 is suggested).

Secondly, $a_0(x_{ij}/C_j)^b$ is used to denote the impedance of passenger number at node D_j. It refers to circumstances such as low velocity at node caused by the increasing number of people which reduces service level. For parameters a_0 and b, $a_0 = 0.15$ and $b = 4$ are set in general [22].

Thirdly, the congestion impedance $a_1 * (x_{ij} - C_{sj})/C_{qj}$ indicates the crowded degree of queueing area in the case that inflow volume at node D_j exceeds the maximum service capacity C_{sj}. The parameter a_1 is initialized as zero and updated only when $x_{ij} > C_{sj}$ (0.2 is suggested).

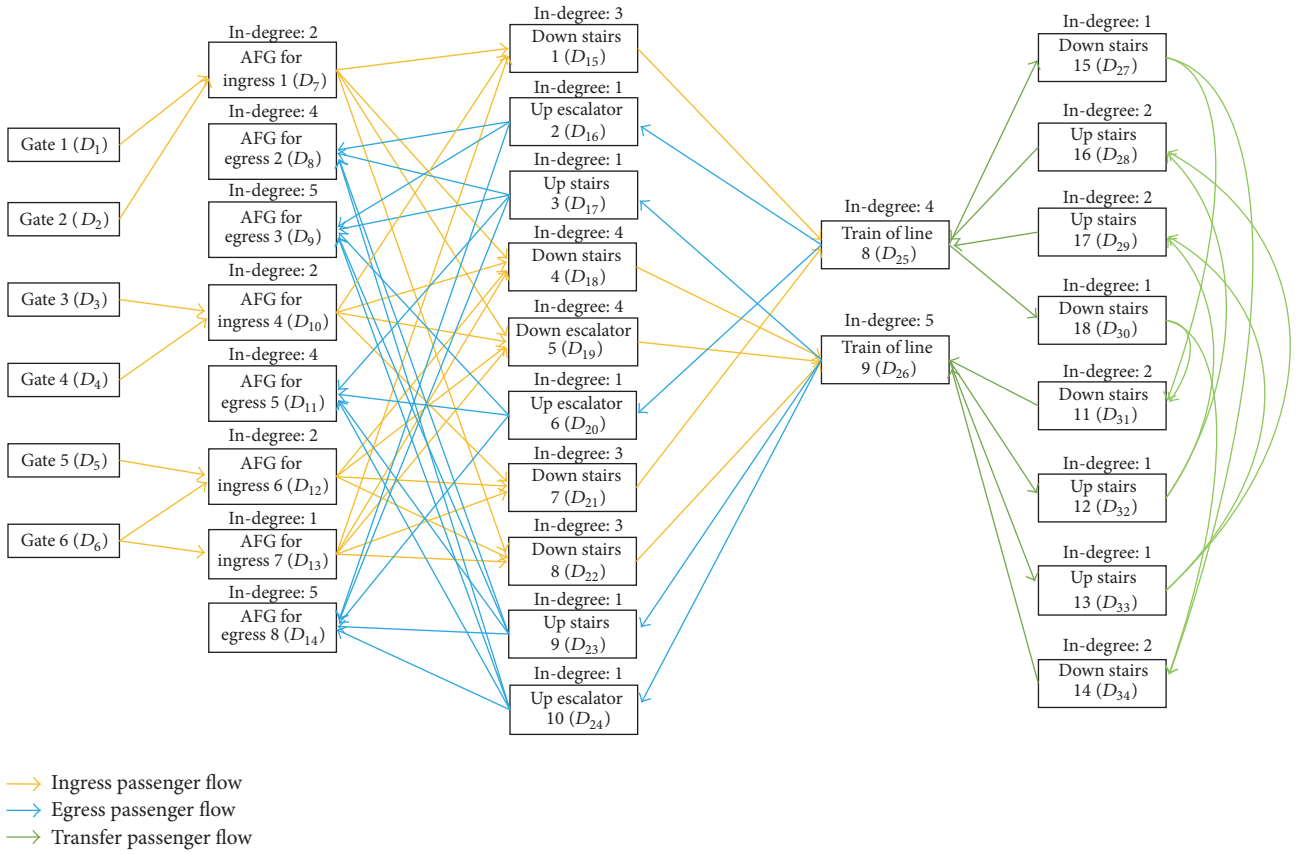


Figure 5: The association network of facilities in LJB station.

4. Algorithm

The Frank-Wolfe algorithm [23] is an effective method to solve the user-equilibrium model. In this algorithm, however, passenger flow assignment under condition of variable demand will lead to the exponential growth of computational complexity. In order to improve the calculating efficiency of station load capacity, attack strategy in communication network is introduced in this section to assist in approaching the verge of facility crash.

4.1. Node Attack Strategy. The core principle of node attack strategy is attacking crucial node in priority. Taking topological properties as the reference to evaluation index, node degree (ND) is used to quantify the node function and influence on the network. Considering that the association network of facilities in URT station is a directed network, node degree should be classified into in-degree and out-degree. The in-degree of nodes is accepted as the evaluation index in this paper, and the higher the in-degree rises with edges pointing to the node, the more significant it is in the network.

4.2. Solution Algorithm. The association network of facilities can reflect passenger flow's motion in the station, while there is a problem in circumstance of transfer stations. Different types of passenger flow have to share some facilities in station, which makes it hard to distinguish the impact on one facility from separate passenger flows. Thus we optimize the input of passenger flow by setting proportion on passenger types.

Based on the Frank-Wolfe algorithm and node attack strategy, the algorithm procedure for load capacity of URT station is as follows (shown in Figure 3).

Step 1. Simplify the association network of facilities, extract the subnetwork which is independent of the whole, and then remove edges with large initial impedance.

Step 2. Quantify all node degrees (in-degree) in the current subnetwork.

Step 3. Target the node with the highest in-degree (denoted as objective node) in a directed chain and launch attacks until it crashes. According to Item 4 in Section 2.2, if passengers are allocated to all out-direction edges and total outflow is less than the maximum service capacity, the node inflow volume equals the total outflow volume. Otherwise, the inflow volume of the node is the loading limit.

Step 4. Allocate the inflow of objective node in reverse direction in order to ascertain the outflow volume from those nodes which point to the objective node.

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21	D22	D23	D24	D25	D26	D27	D28	D29	D30	D31	D32	D33	D34
D1	-																																	
D2		-																																
D3			-																															
D4				-																														
D5					-																													
D6						-																												
D7	38.9	31.4					-																											
D8								-								18.4	24.1						28.5	27.3										
D9									-							31.1	16.2			24.3			30.1	30										
D10		49.9	12							-																								
D11											-						25.8			3.6			39.9	40.4										
D12				23.5	40							-																						
D13					27.1								-																					
D14														-		30.3	31.7			27.2			15.6	15.1										
D15							11			46.1			41.7		-																			
D16																-									6									
D17																	-									6.5								
D18							33.7			6.8		33.6	41.1					-																	
D19							35.1			5.5		32	40.9						-																
D20																				-					6									
D21										23		5.4	25.4								-													
D22							28.5					31.1	7.3									-												
D23																							-			6.5								
D24																								-		6.5								
D25															6						6				-									
D26																		6.5	6.5			6.5				-				6.5				6.5
D27																									6		-							
D28																												-			9.8	7.3		
D29																													-		7.6	9.5		
D30																									6					-				
D31																											5.8			9.9	-			
D32																										6.5						-		
D33																										6.5							-	
D34																											8.7			6.4				-

Figure 6: The distance matrix of association network of facilities.

Step 5. If the outflow volume exceeds the maximum service capacity, go to Step 8. Otherwise, go to Step 6.

Step 6. Attack adjacent nodes which connect with the objective node. The outflow volume of new objective node is equivalent to its maximum service capacity except when the out-degree is 1 (skip this node).

Step 7. Keep the flow volume of crash nodes invariant and assign the rest of the passenger flow by the Frank-Wolfe algorithm.

Step 8. Judge on whether network can keep offering service to all kinds of passenger flow. If all links for one kind of passenger flow turn to be infeasible, the load capacity of station is equivalent to the current passenger flow on network. If not, keep on implementing the node attack strategy, redoing Steps 4 to 8.

5. Case Study

In this section, we illustrate the application and evaluate the effect of the method to calculate the load capacity of URT station on real-world instance. The example is based on a URT station in Shanghai and all data are collected from Shanghai Metro Operation Co., Ltd., in November 2016.

5.1. Basic Scenario. Lujiabang (LJB) station is a transfer station for Metro Line 8 and Line 9, sharing the station hall on underground floor as shown in Figure 4(a). Staircases and elevators arranged in north-south position lead to the platform of Line 8 on underground two, shown in Figure 4(b), while those arranged in east-west position lead to the platform Line 9 on underground three, shown in Figure 4(c). Meanwhile, Figure 4(d) depicts the intersection

Table 2: The Limit capacity of nodes.

Node	Type	Limit Capacity (ped/h)
D_7	AFG	$9000 + 3 \times 25 = 9075$
D_8	AFG	$7200 + 3 \times 20 = 7260$
D_9	AFG	$12600 + 3 \times 35 = 12705$
D_{10}	AFG	$10800 + 3 \times 30 = 10890$
D_{11}	AFG	$7200 + 3 \times 20 = 7260$
D_{12}	AFG	$10800 + 3 \times 30 = 10890$
D_{13}	AFG	$10800 + 3 \times 30 = 10890$
D_{14}	AFG	$14400 + 3 \times 40 = 14520$
D_{15}	Staircase	$10080 + 3 \times 12 = 10116$
D_{16}	Escalator	$6720 + 3 \times 8 = 6744$
D_{17}	Staircase	$5460 + 3 \times 7 = 5481$
D_{18}	Staircase	$5460 + 3 \times 7 = 5481$
D_{19}	Escalator	$10080 + 3 \times 7 = 10101$
D_{20}	Escalator	$6720 + 3 \times 8 = 6744$
D_{21}	Staircase	$10080 + 3 \times 12 = 10116$
D_{22}	Staircase	$5460 + 3 \times 7 = 5481$
D_{23}	Staircase	$5460 + 3 \times 7 = 5481$
D_{24}	Escalator	$10080 + 3 \times 8 = 10104$
D_{25}	Train	$10560 + 3 \times 80 = 10800$
D_{26}	Train	$14880 + 3 \times 80 = 15120$
D_{27}	Staircase	$12025 + 3 \times 16 = 12093$
D_{28}	Staircase	$13650 + 3 \times 16 = 13698$
D_{29}	Staircase	$13650 + 3 \times 16 = 13698$
D_{30}	Staircase	$12025 + 3 \times 16 = 12093$
D_{31}	Staircase	$13650 + 3 \times 16 = 13698$
D_{32}	Staircase	$12025 + 3 \times 16 = 12093$
D_{33}	Staircase	$12025 + 3 \times 16 = 12093$
D_{34}	Staircase	$13650 + 3 \times 16 = 13698$

staircases for directly transferring from one platform to another.

(1) *Station network*: with analyzing the service chain in LJB station, we build the association network of facilities and equipment and calculate in-degree for all nodes, which is shown in Figure 5.

(2) *Data of passenger flow and train*: firstly, the passenger flow proportion of Line 8 to Line 9 is about 6 : 4. Secondly, the interval time of trains for two lines is 3 minutes in each operational direction, and the number of alighting passengers in LJB station is approximately 20 percent of train seating capacity.

(3) *Limit capacity of nodes*: level F in LOS ($3\,\text{ped/m}^2$) is regarded as the criterion of node failure, and the limit capacity of nodes is calculated, given in Table 2.

(4) *Distance of edges*: distance information of facilities and equipment was collected from CAD design drawings of LJB station, denoted as the matrix in Figure 6.

5.2. Results and Analysis. The association network shown in Figure 4 can be divided into two subnetworks. Subnetwork N_1 is for egress flow and the other one N_2 is for ingress and transfer flow.

Figure 7 depicts that attack was launched first on nodes D_9 and D_{14} in N_1, showing the assignment of current flow volume. Obviously, the outflow of nodes D_{16} and D_{20} equals 7383 ped/h and 8951 ped/h, respectively, which are beyond their maximum capacity. Thus these two nodes crash firstly, resulting in the cascading failure of egress service chain for Line 8. The load capacity of N_1 is equivalent to the sum of D_{16} and D_{20} limit capacities, namely, 13488 ped/h.

Figure 8 illustrates the similar process that happened in N_1. After node D_{26} was attacked at first, none of the allocated outflow volume on nodes D_{18}, D_{19}, D_{22}, D_{31}, and D_{34} exceeded their maximum capacity. Meanwhile, out-degrees of those nodes equal 1, indicating that the crash of node D_{26} would not cause congestion on others. Then attack was launched on node D_{25} whose out-degree is 4, while the outflows of related nodes D_{15}, D_{21}, D_{28}, and D_{29} were less than their maximum capacities. The crash of node D_{25} did not lead to cascading failure either. Therefore, the load capacity of

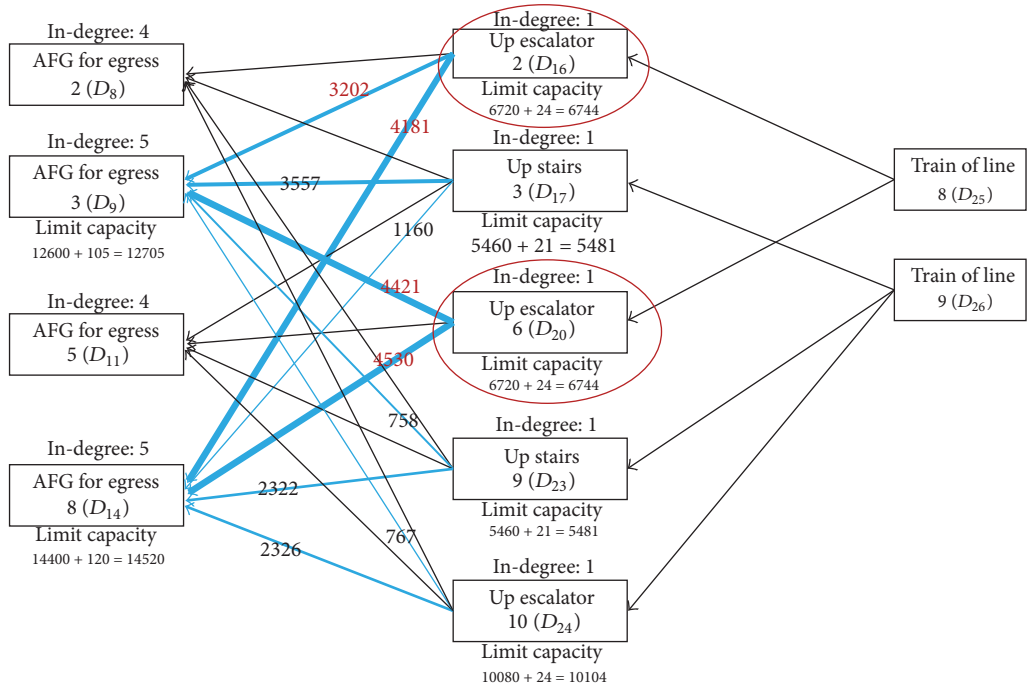


FIGURE 7: The cascading failure in subnetwork N_1.

N_2 is equivalent to the sum of D_{25} and D_{26} limit capacities; that is, alighting passengers reach 10800 ped/h for Line 8 or 15120 ped/h for Line 9.

In order to verify the accuracy of results, the microscopic simulation software StaPass is used in this paper. StaPass specializes in simulating the motor process of passengers specifically in URT station. It is developed by Tongji University and has been successfully applied in station design projects in Shanghai, Guangzhou, Nanjing, and so on.

Set calculation results on the input of the specific scenario. And after one-hour simulation, the density map of passenger flow is shown in Figure 9.

In Figure 9, facilities in dark yellow and red zone are those under F level of service. This is approximately consistent with the calculation result, except that

(1) in simulation, the density of node D_{17} (denoted as "Outlier") is under F level while it is not one of the collapsed nodes in algorithm. But we find that node D_{17} is the next target to be attacked after the crash of nodes D_{16} and D_{20}. It is simulation time that leads to the high density of node D_{17};

(2) the density of collapsed nodes in simulation is much higher than the recommended F value of LOS. In simulation environment, the overlap of different passenger flows in one area will increase the density, while motor progress is simplified in the algorithm. However, the result does not matter.

6. Conclusion

In this paper, we analyzed the service chain of passenger flows in URT station, which could reflect the coupling relationship between separate types of facility. On the basis of that, an association network was built up. Cascading failure theory was introduced to elaborate the influence mechanism of three elements: the motion of passenger flow, the capacity of single piece of equipment, and the layout of facilities. Then we proposed dynamic calculation model for station load capacity based on the user-equilibrium assignment principle. As to the algorithm, the Frank-Wolfe algorithm is a traditional approach for flow assignment, and node attack strategy of complex network was presented to lower the computational complexity. In the case study, we took Lujiabang station in Shanghai Metro as an example to demonstrate the performance of the method and algorithm. In comparison with the result of pedestrian simulation, the vulnerable nodes to facility network and the load capacity of station deduced from this approach are verified to be correct.

The proposed method can be considered as a step towards the quantification of station load capacity. This paper could thus stimulate further research to expend application for more complex stations in urban rail transit system. Moreover, security is a particular problem drawing more attention nowadays. Events related to station security like safety inspection are not considered in this method. The issue on how the interaction of different passenger flows affects load capacity should be studied further. All these will be addressed in future research.

Disclosure

This work has been presented at Transportation Research Board 97th Annual Meeting, but not for publication.

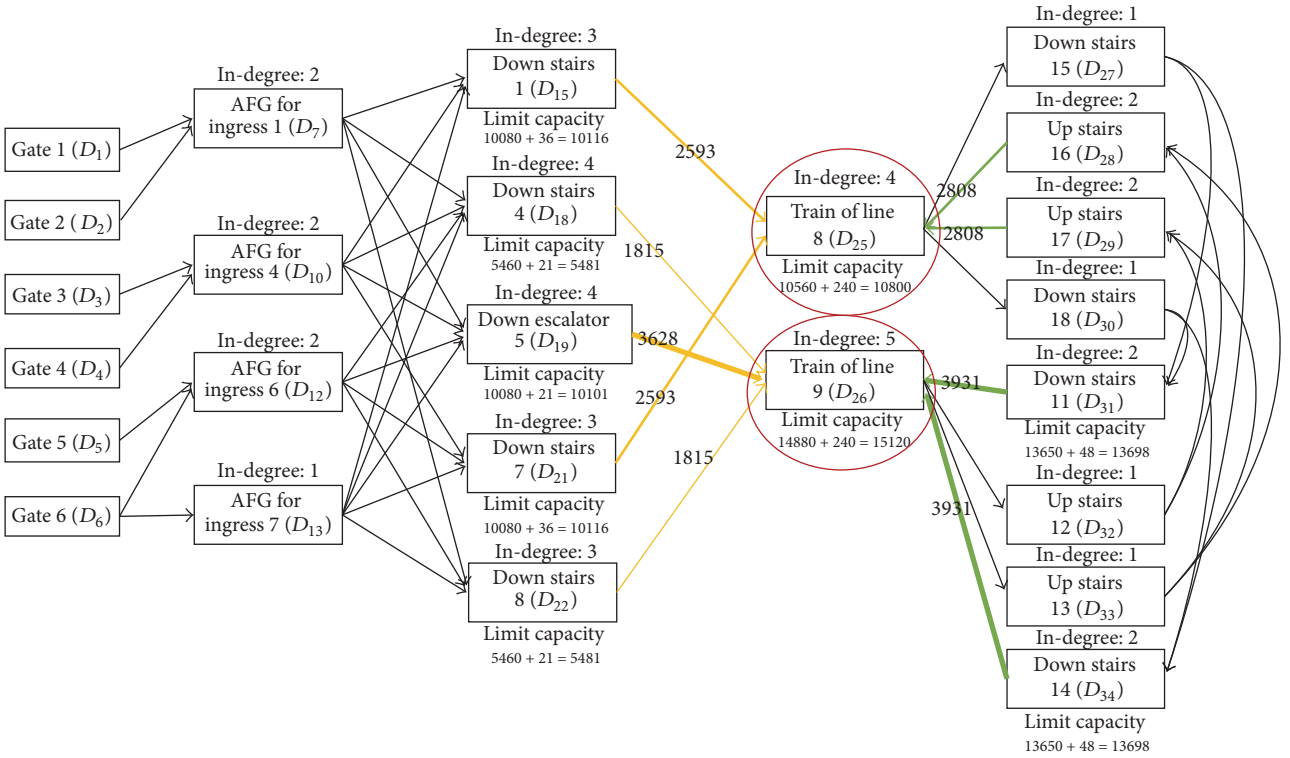


Figure 8: The cascading failure in network N_2.

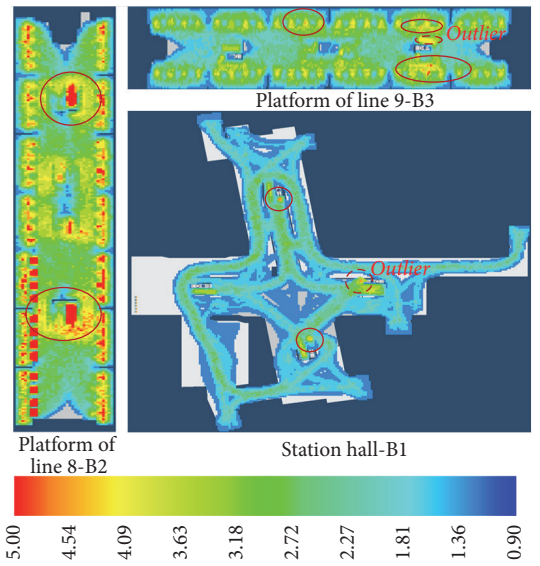


Figure 9: The density map exported from StaPass.

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 Natural Science Foundation of China (Grant no. 71271153).

References

[1] Ministry of Housing and Urban-Rual Development of the People's Republic of China, *The Code for Metro Design (GB50157-2003)*, China Planning Press, China, 2003.

[2] Underground London, *Station Planning Standards and Guidelines*, London Underground Limited, London, UK, 2005.

[3] F. R. Cruz and J. MacGregor Smith, "Approximate analysis of M/G/c/c state-dependent queuing networks," *Computers & Operations Research*, vol. 34, no. 8, pp. 2332–2344, 2007.

[4] F. R. B. Cruz, J. M. G. Smith, and R. O. Medeiros, "An M/G/C/C state-dependent network simulation model," *Computers & Operations Research*, vol. 32, no. 4, pp. 919–941, 2005.

[5] F. R. B. Cruz, T. van Woensel, J. MacGregor Smith, and K. Lieckens, "On the system optimum of traffic assignment in M / G / c / c state-dependent queueing networks," *European Journal of Operational Research*, vol. 201, no. 1, pp. 183–193, 2010.

[6] X.-Y. Xu, J. Liu, H.-Y. Li, and Y.-F. Zhou, "An analytical method to calculate station evacuation capacity," *Journal of Central South University*, vol. 21, no. 10, pp. 4043–4050, 2014.

[7] R. Khalid, M. A. Baten, M. K. M. Nawawi, and N. Ishak, "Analyzing and optimizing pedestrian flow through a topological network based on M/G/C/C and network flow approaches," *Journal of Advanced Transportation*, vol. 50, no. 1, pp. 96–119, 2016.

[8] F. Kaakai, S. Hayat, and A. El Moudni, "Simulation of railway stations based on hybrid petri nets," in *Proceedings of the 2nd IFAC Conference on Analysis and Design of Hybrid Systems, ADHS'06*, vol. 39, pp. 50–55, June 2006.

[9] F. Kaakai, S. Hayat, and A. El Moudni, "A hybrid Petri nets-based simulation model for evaluating the design of railway transit stations," *Simulation Modelling Practice and Theory*, vol. 15, no. 8, pp. 935–969, 2007.

[10] Q. M. Hu, *Passenger Carrying Capacity Evaluation and Simulation of Rail Transit*, Beijing Jiaotong University, Beijing, China, 2011.

[11] Y. Yang, *Layout Optimization of Station Facilities and Evacuation Leaders Based on Analysis of Traffic Capacity*, Beijing Jiaotong University, 2016.

[12] F. Xue, W. N. Fang, and B. Y. Guo, "Rail Transit Station Passenger Flow Evolution Algorithm Based on System Dynamics," *Journal of The China Railway Society*, vol. 36, no. 2, p. 10, 2014.

[13] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, "Simulation of pedestrian dynamics using a two-dimensional cellular automaton," *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3-4, pp. 507–525, 2001.

[14] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[15] J. Barraquand, B. Langlois, and J.-C. Latombe, "Numerical potential field techniques for robot path planning," *The Institute of Electrical and Electronics Engineers Systems, Man, and Cybernetics Society*, vol. 22, no. 2, pp. 224–241, 1992.

[16] D. Helbing and P. Mukerji, "Crowd disasters as systemic failures: Analysis of the love parade disaster," *EPJ Data Science*, vol. 1, no. 1, p. 7, 2012.

[17] P. Gao and R. H. Xu, "Event-driven simulation model for passenger flow in urban mass transit station," *Systems Engineering-Theory & Practice*, vol. 30, no. 11, pp. 2121–2128, 2010.

[18] J. J. Fruin, "Pedestrian planning and design," *Metropolitan Association of Urban Designers and Environmental Planners*, vol. 206, 1971.

[19] F. Xie, S. Q. Cheng, D. Q. Chen et al., "Cascade based attack vulnerability in complex networks," *Journal of Tsinghua University (Sci & Tech)*, vol. 51, no. 10, pp. 1252–1257, 2011.

[20] J. G. Wardrop, "Road paper. Some theoretical aspects of road traffic research," in *Proceedings of the Institution of Civil Engineers*, vol. 1, pp. 325–362, 1952.

[21] M. Beckmann, C. B. McGuire, and B. Winsten, *Studies in the Economics of Transportation*, Yale University Press, New Haven, CT, 1955.

[22] J. F. Zheng, *Studies on Complex Network Modeling and Dynamical Processes in Typical Networks*, Beijing Jiaotong University, Beijing, China, 2010.

[23] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval Research Logistics Quarterly*, vol. 3, pp. 95–110, 1956.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 7284815, 11 pages
https://doi.org/10.1155/2018/7284815

Research Article

PULSim: User-Based Adaptable Simulation Tool for Railway Planning and Operations

Yong Cui,[1] **Ullrich Martin**,[1] **and Jiajian Liang**[2]

[1]*Institut fuer Eisenbahn- und Verkehrswesen der Universitaet Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany*
[2]*China Academy of Railway Sciences, Daliushu Road 2, Haidian District, Beijing 100081, China*

Correspondence should be addressed to Yong Cui; yong.cui@ievvwi.uni-stuttgart.de

Received 13 October 2017; Revised 12 January 2018; Accepted 5 February 2018; Published 22 April 2018

Academic Editor: Lingyun Meng

Simulation methods are widely used in the field of railway planning and operations. Currently, several commercial software tools are available that not only provide functionality for railway simulation but also enable further evaluation and optimisation of the network for scheduling, dispatching, and capacity research. However, the various tools are all lacking with respect to the standards they utilise as well as their published interfaces. For an end-user, the basic mechanism and the assumptions built into a simulation tool are unknown, which means that the true potential of these software tools is limited. One of the most critical issues is the lack of the ability of users to define a sophisticated workflow, integrated in several rounds of simulation with adjustable parameters and settings. This paper develops and describes a user-based, customisable platform. As the preconditions of the platform, the design aspects for modelling the components of a railway system and building the workflow of railway simulation are elaborated in detail. Based on the model and the workflow, an integrated simulation platform with open interfaces is developed. Users and researchers gain the ability to rapidly develop their own algorithms, supported by the tailored simulation process in a flexible manner. The productivity of using simulation tools for further evaluation and optimisation will be significantly improved through the user-adaptable open interfaces.

1. Introduction

In order to improve the efficiency and effectiveness of railway planning and operations, the relationships between the critical components of railway systems and the behaviour of the studied system must be investigated. A specific investigated problem usually presents many possible alternatives for design. In order to enable evaluation and comparison among the various possible alternatives, their respective effects and outputs will be predicted through experiments. However, it is not practical to examine a design alternative directly based upon an actual system, because either such a system is not existing or experimentation upon it is prohibitively expensive.

For a simple railway network with a limited number of train runs, an analytical approach is practical in order to evaluate the output of a design alternative and even to find an optimal solution in a closed-form expression. For a large-scale network with a high density of train movements, the computational complexity of such an analytical model becomes considerably high. In this case, the simulation approach is often applied.

Several simulation tools are available for railway planning and operations purposes. The application of personal computer based simulation approaches for railway systems began in the 1990s, at which time most tools were originally developed as laboratory versions. Along with continuous development and applications, these tools have been widely accepted for research and commercial purposes. Meanwhile, a large number of data and models of railway systems have become available. In this paper, the existing simulation tools and applications will be presented in Section 2.

However, the potential of simulation approaches has not yet been sufficiently utilised. The limits of the current simulation tools, especially their lack of capability for extension and customisation, are discussed in Section 2. To overcome these limits, a user-based, adaptable simulation tool called PULSim has been developed in recent years. In Section 3,

the model, the workflow, and the main features of PULSim are introduced. The software's capability to provide users and third-party applications with an open interface for dynamic interaction and flexible extension is presented in Section 4, along with several use cases and scenarios. Finally, the perspective of further development is discussed in Section 6.

2. Applications of Railway Simulation and the Limitations of Existing Simulation Tools

For railway planning and operations, simulation approaches have been widely used by researchers, railway infrastructure companies, and railway operating companies. In German-speaking countries and Europe in general, the simulation software RailSys (developed by Rail Management Consultants GmbH, [1]), OpenTrack (from OpenTrack Railway Technology Ltd., [2]), and LUKS (by VIA Consulting & Development GmbH, [3]) are very practical for railway planning and operations.

The potential areas of application of simulation tools include technical feasibility studies, determination of conflicts, evaluation of the quality of a timetable, capacity research, and dispatching. Simulation tools can be used, for example, to build a test environment for the evaluation of the technical feasibility and the benefits of using the European Train Control System (ETCS) [4]. To investigate potential conflicts, a timetable simulation can be carried out, and the resulting waiting time can be determined as well [5, 6]. In certain situations, various optimisation functions can be added to a timetable simulation in order to reduce hindrances and waiting time of train runs. As a deterministic process in principle, the possible influence and variances during railway operations are not considered in a timetable simulation. If the robustness and stability of a timetable are matters of concern, a so-called operational simulation can be carried out, in which stochastic influences will be introduced. The robustness and stability of the investigated timetable can be derived according to statistical indicators, which are calculated based on the outputs of several rounds of simulation with different randomly generated influences [7, 8]. In the area of capacity research, system performance can be evaluated through simulating randomly generated timetables for a certain operating program, a process which can, in addition, identify bottlenecks and determine the quality of an operating program [9]. The approach of capacity research is implemented in the software tool PULEIV, developed by the IEV (Institut für Eisenbahn- und Verkehrswesen der Universitaet Stuttgart) for flexible analysis, evaluations and reports [10]. Simulation approaches can not only be used to identify conflicts, but also to generate a feasible dispatching timetable to resolve conflicts [11, 12], as well as to study the relationship between systems performance and dispatching measurements [13, 14].

At the beginning of development of a certain simulation tool, its further possible applications are often unknown by its developers. Therefore, it is difficult to foresee and to provide specifically desired functions and outputs, even if the simulation tool would theoretically able to fulfil said requirements. Hence, an open interface that enables customisation for user-based, adaptable simulation processes is a vital factor for the success of a simulation tool. Unfortunately, currently available simulation tools for railway planning and operations lack, in large part, transparent standards and open interfaces. Some internal workflows, such as the applied dispatching algorithm and the implementation of signalling systems, are not sufficiently documented for users and third party individuals and, as such, it is difficult for the end-user to understand the internal assumptions and simplifications made within these tools. The outputs of various simulation tools are insufficient and not standardised with insufficient documentation, which prevents users from carrying out further evaluation and optimisation. In addition, it is almost impossible to organise user-defined workflows and to integrate new functions into the existing simulation tools, which leads to constraints in their usability and applicability. The high efforts required for further evaluation decrease the efficiency and effectiveness of these simulation approaches. In addition, some well-known issues, such as deadlock problems in synchronous simulations, have not been addressed sufficiently. Once deadlocks take place, the simulation process must be cancelled or be solved manually, which greatly limits the usability of the simulation tool as well.

A user-based, adaptable simulation tool, PULSim, has been developed in recent years. In [15, 16], the tool was known as DoSim. The name has been changed to PULSim in order to conform to the naming convention of software tools in the IEV. PULSim provides a platform for railway simulation with a unified model, a transparent workflow, and open interfaces. Critically, PULSim provides the possibility for third-party applications to be flexibly integrated into the software. The functionality of PULSim can be extended thanks to its open interface, and the software has been tested with several railway networks in Germany. At the moment, the software is available for download from GitHub (https://github.com/herrcui/RailView/wiki), with the instructions for installing and using it provided with demo data. The user manual and the open interfaces are planned to be comprehensively documented and published in further development. The Continuous Integration (CI) for open access and updates will be implemented as well. In this paper, the model, the workflow, and the Graphical User Interface (GUI) are presented. Three special features of PULSim are outlined in particular:

(i) Dispatching mechanism

(ii) Deadlock avoidance

(iii) Open interfaces for user-based adaptable simulation.

An introduction of the model, the workflow, and the GUI of PULSim is provided in Section 3.1. In Sections 3.2 and 3.3, the design of the dispatching mechanism and deadlock avoidance are explained. The open interface for user-based adaptable simulation and its applications are introduced in Sections 4 and 5.

3. Introduction of Simulation Tool PULSim

3.1. The Model, the Workflow, and the GUI of PULSim. In PULSim, the infrastructure, rolling stocks, and railway

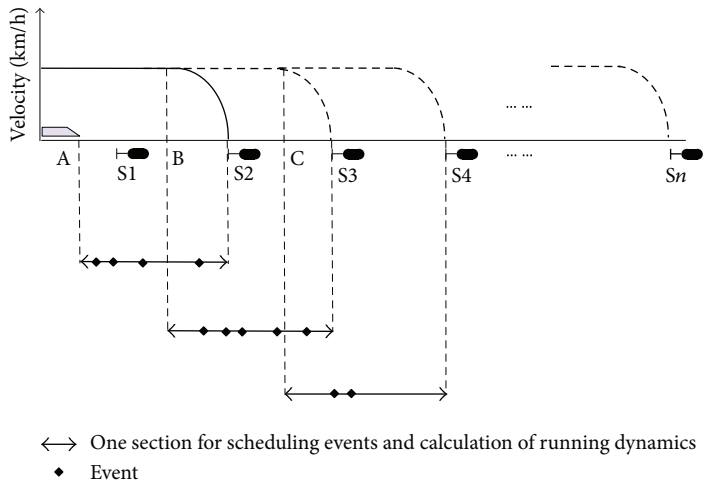


Figure 1: An example of event scheduling and calculation of running dynamics.

operations are modelled as components of railway systems. Attributes are used to describe the static and the dynamic information. The static information refers to the invariant values of the railway systems, for example, the length of a track, the configuration of a train, and the scheduled departure time of a train run. The dynamic information will vary along the duration of the simulation. For example, the aspect of a signal, the running dynamics of a train, and the occupancy situation of a track are updated continuously during the simulation. The model applied in PULSim provides a basis for multiscale simulation on the microscopic, mesoscopic, and macroscopic levels [17].

An internal class model of railway systems comes built-in with PULSim. This model is the further development and extension from the IEV core model [18]. The publication of the detailed design of this model, and its workflow, is planned for 2018. It will also consist of the details of the open interfaces for interacting with the model along with the software. Particularly, the basic structures and the path components based on the infrastructure element model are developed in the core model, so that the macroscopic, mesoscopic, and the microscopic models are fully integrated (the infrastructure element model is a microscopic model to model the basic infrastructure elements including tracks, turnouts, crossings, and single and double slips. A basic structure is defined as a basic occupancy element in which all parts should be equally occupied, regardless of the design of the operating program. A path component is a directed edge inside a block section used for train runs, which can be released separately as a directed occupancy element. Basic structures and path components are modelled for the mesoscopic level). This model can be used for various purposes, including capacity research, bottleneck analysis, and train dispatching [19]. It is independent of a concrete data format, which allows users to concentrate on the workflow and the business logic of railway planning and operations without requiring knowledge relating to a certain data format. Additional parsers are required to convert different data formats into the class model. At the moment, the data format used by RailSys [1] is supported by PULSim. The development of the parser to support railML [20] is in progress.

The workflow for an event-driven simulation is applied in PULSim. An event is defined as an occurrence that may change the attributes of the system at a certain point in time. During the simulation process, a series of time points for the occurrence of an event are identified and continuously updated. The system evolves over the series of time points. During a simulation process, the attributes (dynamic information) of the system (e.g., the position of a train) may be changed continuously. In PULSim, only that change of attribute, which causes interactions between two or more components or triggers another event, is considered a discrete event. Otherwise, the changes of the running dynamics, for example, the continuous variation of train velocity, will be handled at each discrete time point by the train itself, without requiring an explicit event. Hence, the event-driven process in PULSim can be organised efficiently with a limited number of events.

During the simulation process, a running time calculation is carried out for each individual train run. The change of position and velocity of each train will not be treated as an event. Instead, the events will be scheduled with respect to the signalling system. An example of organising an event-driven simulation is shown in Figure 1. After a train has received a Movement Authority (MA) at point A, the complete speed profile will be calculated until the End of Authority (EOA) at main signal S2. The corresponding events from point A to EOA S2 are scheduled. At point B, a new MA ending at main signal S3 is granted. The train movement from B to S3 will be calculated. Meanwhile, the events already scheduled between B and S2 will be discarded, since the running dynamics from B to S2 will be updated with the newly calculated speed profile. The events from B to S3 will be rescheduled and updated again.

Once a request of infrastructure resources is sent, the current operational situation will be observed by the dispatching module. The requested infrastructure resources can only be granted at a conflict-free and deadlock-free situation.

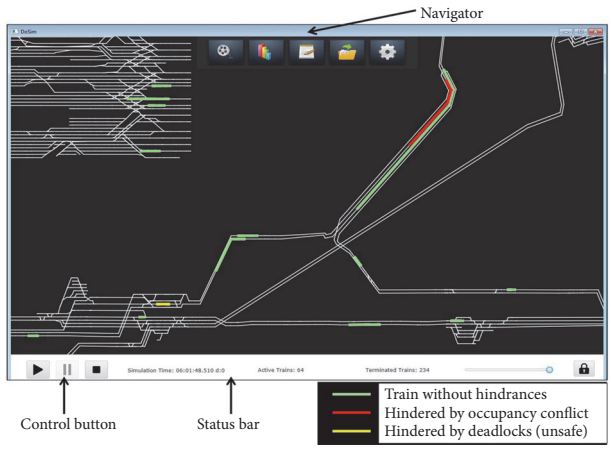


Figure 2: Screenshot of "Network/Train Run" view in PULSim.

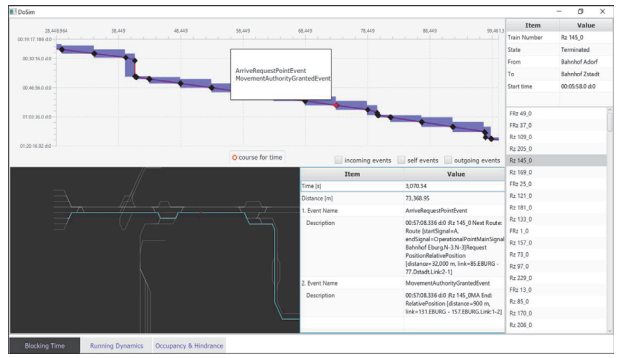


Figure 3: Screenshot of tracing events and train run.

A simple First Come First Served (FCFS) algorithm is provided by PULSim as a default module to solve occupancy conflicts. It can be replaced with a customised dispatching module provided by the user (see Section 3.2). A standard solution for deadlock avoidance, combined with the approach of searching for feasible resources to reduce false-positive situations, is provided in PULSim as well (see Section 3.3).

The design of the GUI in PULSim aims to provide the user with an insightful view during the process of railway simulation and to support efficient user interaction. Therefore, the concepts of visualisation and interaction are considered critical areas of focus. Various user interfaces are organised in different views. The movement of train runs in the railway network (the view of "Network/Train Run"), the information behind the scenes ("Analyse and Evaluation"), and the interaction between the simulation tool and users ("Open Interface") are provided.

In Figure 2, the screenshot of the "Network/Train Run" view is presented. This is a standard function provided by almost all railway simulation tools. A control panel can start, pause, or stop a simulation. Trains are marked in green, red, or yellow, which represent a train without hindrances, with hindrances due to occupancy conflicts, or with hindrances due to deadlocks, respectively. In order to enable a quick switch among different views, a floating navigator is set on top of the screen.

The function of tracing the occurrence of events is provided by PULSim. Within this view, users can observe the entire workflow of the simulation within the diagram of the blocking time stairway and running dynamics for each train run (Figure 3). Other evaluation views for the illustration of occupancy and hindrance of the running simulation are also provided. In addition, an "Open Interface" view to achieve a user-based, adaptable simulation is built into PULSim (see Section 4).

3.2. Dispatching Mechanism of PULSIM. With conventional dispatching approaches [21], the dispatching algorithm is activated once one or more potential conflicts are identified. However, it might be too late to start train dispatching at the time of the occurrence of conflicts. Furthermore, it is necessary to check for potential conflicts within the conventional dispatching systems periodically. The timing for identification of conflicts is critical. On one hand, some conflicts may be overlooked using a long time interval between two rounds of checks; on the other hand, overly frequent checks might lead to performance issues.

In PULSim, a new conflict identification/resolution mechanism is implemented. In contrast with other conventional dispatching approaches, the activation of a designed algorithm for train dispatching is not initiated at the time of facing conflicts, but at an earlier time, at which a train has the chance to occupy the requested infrastructure resources. For example, if the action to occupy the requested resources would produce significant hindrances on other trains, the train will give up the chance of occupancy. Therefore, potential conflicts from potential hindrances can be prevented in advance. In addition, this mechanism suits the workflow of event-driven simulation. A periodical examination of conflicts in a fixed time interval is not necessary; any potential conflicts can be identified and resolved at the time a Movement Authority is able to be granted.

In order to better illustrate this phenomenon, Figure 4 provides an example of giving up the opportunity of occupancy [15]. Within it, Train T1 is requesting occupancy of the route from signals S1 to S3. This route is not currently occupied by other trains (conflict-free), and its occupancy will not cause deadlocks (deadlock-free). Deadlocks may, indeed, be regarded as a type of conflict in train dispatching. In order to differentiate between occupancy conflicts and deadlocks, the term conflict-free and deadlock-free are used in this paper. There is a train T0 occupying the route from S3 to S5. Its occupancy time along this route is assumed to be very high due to a technical failure of train T0. If train T1 occupies the route from S1 to S3, trains T2 and T3, which will take the route from S2 to S4, must wait until the technical failure of T0 has been resolved. It would be more efficient if train T1 were to give up the chance of occupancy and wait before signal S2. Therefore, potential hindrances and conflicts can be reduced or avoided in advance.

The workflow of the dispatching mechanism is shown in Figure 5. As usual, a request will be pended in case of conflicts or deadlocks. A special feature for train dispatching

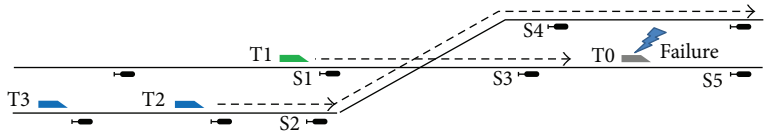


Figure 4: Example of giving up of occupancy [15].

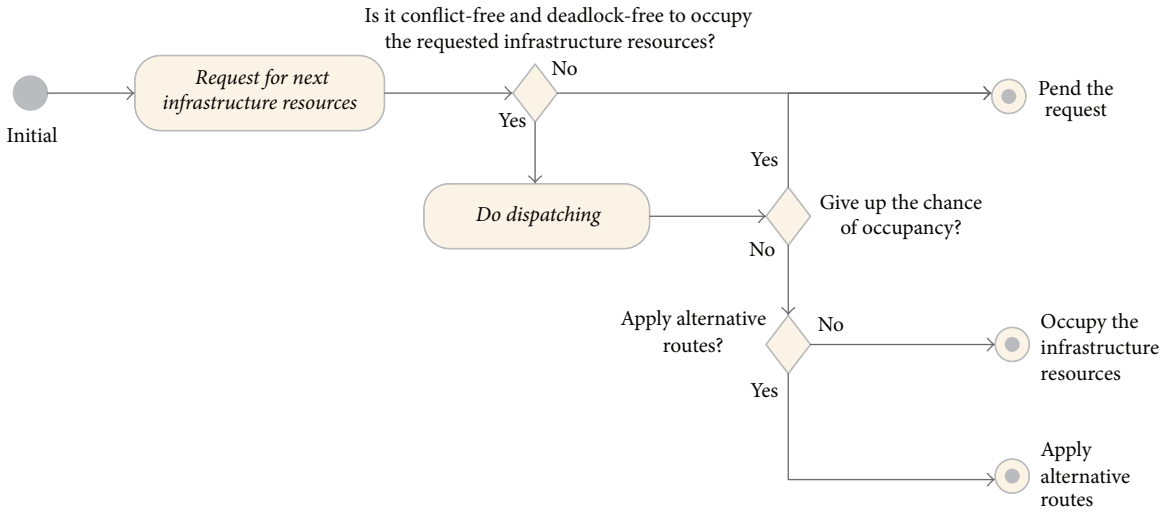


Figure 5: The dispatching mechanism in PULSim.

is designed when a request for infrastructure resources is conflict-free and deadlock-free. In this case, the train still has to evaluate the operational situation before it occupies the requested infrastructure resources. The decision of whether or not to give up the chance of occupancy depends on the applied dispatching algorithm, which is illustrated as the activity "do dispatching" (see Figure 5). On the basis of a FCFS strategy, the train will always take the chance to occupy the requested infrastructure resources. The FCFS strategy is implemented in PULSim by default, and if a dispatching strategy other than FCFS is applied, the train may give up the chance of occupancy according to certain dispatching objectives (e.g., to reduce total hindrances). The request will be then pended for the next round of dispatching. In addition, the dispatching mechanism also supports the implementation of applying alternative routes.

Although FCFS is currently the only strategy implemented in PULSim, users can customise and integrate their own dispatching algorithms into the dispatching process. This is achieved through the open interface and a standard workflow (see Section 4). Inside the workflow, users can define their own implementation to calculate the predicted effects for all possible dispatching actions (to occupy, to give up, or to choose an alternative route). The action with the best effects (e.g., with the minimal waiting time) will be chosen as the dispatching decision.

An example of using user-customised dispatching algorithms is published on GitHub (https://github.com/herrcui/RailView/blob/master/PULSimReleases/scripts/railapp.dispatching.services.ExternalDispatchingService.py). Within this example, the delays experienced by conflicting trains are retrieved from the simulation model by comparing the current departure/arrival times with the scheduled departure/arrival times. The example utilises a simple form of logic in order to illustrate the usage of the open interface. That is, the train experiencing the longest delay will receive the highest priority.

A Java interface is defined for users to provide customised train dispatching. Through the interface, a list of resulting states according to the generated possible actions will be given from PULSim as input. A state is a multidimensional list, which represents the resulting occupancy situation of an action for each train. The predicted occupancy situation includes the lists of the occupied infrastructure resources, as well as the start and end of the blocking time for each infrastructure resource. It should be noted that a state is not conflict-free. Users can implement their own algorithm to identify and to resolve conflicts. As a basis, all trains predict their further occupancy situation as if other trains do not exist. The predicted occupancy situation will then be shifted or changed according to a certain action. For the action to occupy the requested infrastructure resources by a train T, the blocking time of the infrastructure resources for other trains should be shifted until the infrastructure resources are released the train T. Similarly, if a train T gives up the chance of occupancy to wait for another train T', the blocking time of other related trains should also be shifted, until the infrastructure resources are released by T'. Hence, the current operational situation and the action to be taken are represented in the form of the resulting states.

Users can implement the method "determineAction" to determine a dispatching action. The method returns an

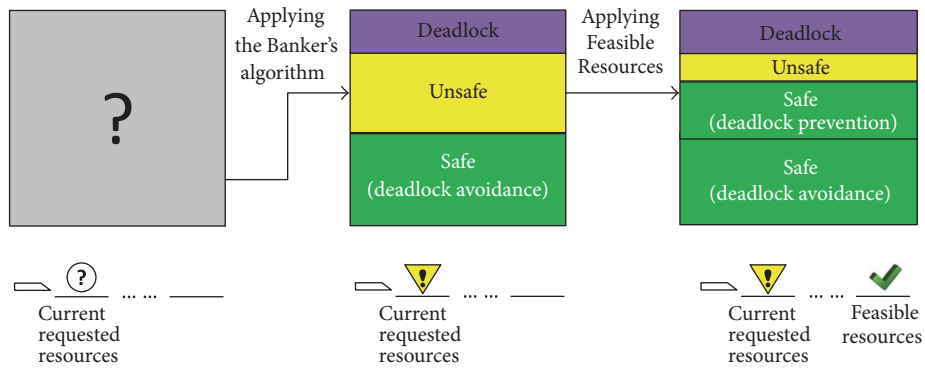


Figure 6: Searching feasible resources to reduce false-positive situations with the Banker's algorithm [15].

index of the given actions, which indicates the chosen action according to a specific dispatching objective (e.g., with minimal waiting time). At the moment, an algorithm based on reinforcement learning for train dispatching is being developed. A preliminary investigation to evaluate the effects for a certain state and a given action is described in Section 5.

3.3. Deadlock Avoidance Based on Banker's Algorithm. Deadlock problems [22] are well-known in the area of synchronous simulation. Some approaches of deadlock avoidance are proposed in [23–26]. PULSim applies a method based on the Banker's algorithm [27], which was initially used for the purposes of railway simulation in [11]. Once a train requests to occupy one or several infrastructure resources, a deadlock-free test will be carried out based on the Banker's algorithm. The workflow of this deadlock-free test is straightforward; if all the required infrastructure resources along the path of a train are available (in other words, not occupied by other trains), the train will be identified as a passed train. All the resources occupied by the passed train will be then returned to the system. The deadlock-free test will be executed iteratively for every train, until all trains have been identified as passed trains. In this case, a deadlock-free situation can be always guaranteed, since the feasibility of resource allocation has been proven.

If there are some trains remaining, whose required infrastructure resources are not available, the situation will be regarded as being in an unsafe state. It should be noted that an unsafe state does not necessarily result in a deadlock situation. A false-positive situation is defined in [11] as a deadlock-free test result that is read as positive but is actually negative. False-positive situations should be reduced to avoid unnecessary waiting time due to the overly strict rules in deadlock-free tests.

A method to search for feasible resources for reducing false-positive situations with the Banker's algorithm is implemented in PULSim (Figure 6). Feasible resources are defined as one or more connected and not yet granted infrastructure resources in the further path of a train. If the train occupies the feasible resources, a circular wait situation should not take place. The circular wait situation is one of the necessary conditions for deadlocks [25]. Once the circular wait situation is eliminated, the feasible resources can be granted to the train without deadlocks, even if the train has not passed the deadlock-free test through the Banker's algorithm. In [15], the approaches to identify feasible tracks as well as the effects of reducing false-positive situations are presented in detail.

In this method, a request for infrastructure resources will be at first analysed in accordance with the Banker's algorithm. Upon passing this test, the requested infrastructure resources can be safely granted to the requesting train. In case a request fails to pass the test with the Banker's algorithm, the method of searching for feasible resources will commence. If feasible resources are indeed available, the train's request may still be approved.

The aforementioned combination of the Banker's algorithm and the method of searching for feasible resources can ensure a high applicability and efficiency for deadlock avoidance. This method has been proven in PULSIM for several practical applications and case studies. For example, a real railway network in Germany is used for the case studies, consisting of 129 stations, including a large terminus station, and 2,388 train runs. Even with randomly shuffled departure times of train runs, deadlocks can be successfully avoided. Another extreme example is also available on GitHub. For a highly congested network with a high potential of deadlocks, the software PULSim is still able to simulate all the train runs without deadlocks.

4. User-Based Adaptable Simulation

4.1. Open Interface of Railway Simulation Tools. A simulation tool serves as the basis for further investigations into railway planning and operations, in which the output of a simulation can be used as the input of other applications. Evaluation of the simulation output is the most popular use case of these applications. Therefore, the interaction between a simulation tool and other applications for evaluation of simulation output should be enabled.

The interactions for evaluation can be categorised as either offline or online evaluations. In the case of offline evaluation, the data generated by the simulation tool will be initially stored and will be analysed afterwards. This process is suitable for a middle-term or long-term evaluation with a very large amount of data. The saved log files can be reused in the future for other purposes of evaluation without requiring

an additional run of simulation. With an online evaluation, the evaluation between a simulation tool and third-party applications can be achieved efficiently through exchanging real-time data. A built-in package in the simulation tool or an open interface supporting direct data access is required for online evaluation. Sometimes, the output of the third-party applications will be fed back to the simulation tool. In order to exchange information for further evaluation and analysis, open interfaces between a simulation tool and other third-party applications are provided in PULSim.

When conducting offline evaluation, it is very common to define the format and the schematic structure of the saved log files as an open interface. Popular formats of the log files include Extensible Markup Language (XML) files, Comma Separated Values (CSV), files or plain text files. As an open interface, the format and the schematic definition of the output files should be provided and published in advance. At the time that a simulation tool is developed, its possible usage cases may still be unknown. Therefore, it is a challenge to ensure that the tool provides open interfaces for further possible applications at an early stage. Currently, an open and standardised interface for the output of railway simulation is not available, with each simulation tool possessing its own definition for the output of simulation. Attempts to unify various outputs generated from different simulation tools require considerable effort. Certain simulation tools output log files without a published interface, which can be used in internal testing for various special purposes. It is not recommended to use these data due to the lack of documentation and support, although the information may be inferred from the contents of the files.

For online evaluation, the output of simulation can be retrieved and analysed directly through an open interface. The process of railway planning and operations can be evaluated and optimised by tuning the parameters of the simulation tools in real time. The open interface can be provided either from an open Application Programming Interface (API) or in the form of a scripting language.

An open API enables a dynamic data exchange via directly accessing the simulation tool from other third-party applications. In the simulation tool OpenTrack, the open API is provided in the form of web services [28], from which a third-party application can send messages as commands to OpenTrack and retrieve the output as status messages back to the application. The commands and the status messages are transferred in a machine readable format, according to the Simple Object Access Protocol (SOAP) over Hypertext Transfer Protocol (HTTP). An example use case of using the OpenTrack API for train dispatching is presented in [28]. New dispatching algorithms and prototypes are tested in a simulation environment, in which the realistic situation is replaced by the simulation tool OpenTrack. The commands and status messages in the same format used in reality are exchanged between the dispatching software and OpenTrack. Hence, the newly developed dispatching algorithms can be evaluated in an inexpensive experimental environment.

The system performance and efforts required for the development and maintenance of an open API solution should be considered. The amount of data in railway planning and operations is usually very large. Taking the example of using a web service with SOAP, the enveloped and transferred messages via XML data will cause a high overhead for communication, which will impact the performance of the system. Large software vendors and tools are often required for the implementation and development of web service solutions, which makes such a situation infeasible for researchers and institutes who are interested in rapid development with lightweight solutions. In addition, the complexity of learning the web service-based system and the induced efforts for debugging, testing, and maintaining the entire system are considerable.

Another option for an open interface is to use scripting language, which can combine the advantages of both offline evaluation and online evaluation. A scripting language is a programming language in the form of a series of imperative commands, which are executed in a certain run-time environment. These commands are interpreted during the run-time without needing to be compiled. Hence, sophisticated processes can be customised by users in an adaptable way. With scripting languages, the output of simulation can be either exported as log files for offline evaluation, or be accessed online via the provided interfaces (commands). Upon the demands and the required system performance, a user can flexibly decide the mode of interaction. A scripting language hides the internal structure and the implementation details of a complex system, which enables the user to concentrate on the core functions and the integration of the functions. The complexity of learning the system and the efforts for development, deployment, and maintenance are therefore significantly reduced.

Using open interfaces with scripting language can provide additional advantages in balancing the requirements of transparency and the complexity in development. It is important for users to understand the mechanism and the assumptions of a simulation tool. The open-source method would be an option to promote transparency and open collaboration. However, an excessive amount of implementation details of the complete source code will increase the complexity in comprehension. It should be noted that most users of railway simulation tools are interested principally in the domain logics related to railway planning and operations. Therefore, open interfaces will provide a moderate level of details for researchers to enable rapid development and integration of their own algorithm. As a precondition, the mechanism of the simulation tool should be published, so that users can customise their own logic based on the utilised mechanism. Hence, the internal assumptions and mechanism can be naturally understood by users according to the well-specified open interfaces. For example, the internal mechanism of dispatching systems should be specified in advance (see Section 3.2). As long as a customised dispatching algorithm is implemented, the mechanism of train dispatching is understood by users as well. In the future, the mechanism of running time calculations and signalling systems will be published in the form of open interfaces to users as well.

Other simulation tools also provide some open access to their internals. For example, the parameters used in RailSys and OpenTrack can be viewed and set through system

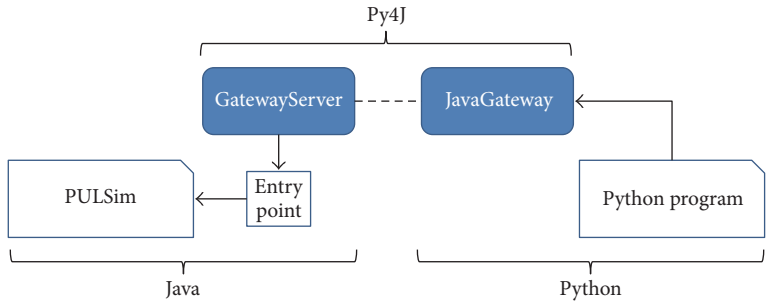


Figure 7: Py2J: accessing Java objects and services from Python.

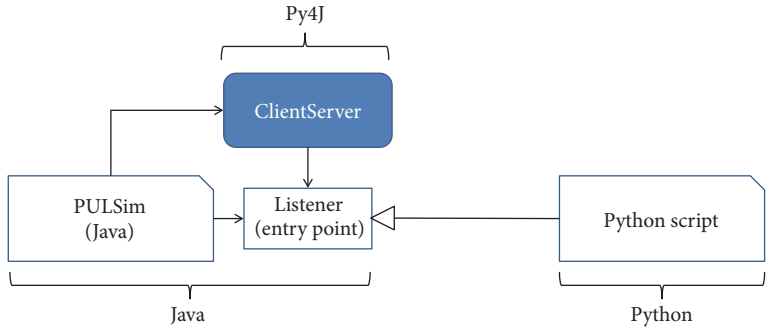


Figure 8: J2Py: accessing Python script from Java.

configuration. Users can also interact with OpenTrack through web services. PULSim provides additional support for customised simulation workflow and integration of their own algorithm. In Sections 4.2 and 5, the implementation of an open interface with scripting language and the applications of using the open interface provided by PULSim are presented.

4.2. The PULSim Implementation. In PULSim, a user-based adaptable interface for the simulation, evaluation, and optimisation for different applications is built using scripting language. Scripting languages can be categorised as either domain-specific languages or general-purpose languages. A domain-specific language is specifically designed for a certain application and platform, and an external parser is needed to interpret its commands. Additional efforts are required for users to learn the language. With general-purpose languages, and plenty of support, libraries and documentation are available, which usually allows users to obtain sufficient knowledge of general-purpose languages with lower learning efforts. In addition, most general-purpose languages are easily extended for certain special purposes.

Popular general-purpose scripting languages include Python, Perl, and Ruby. Among others, Python has been widely used for rapid and efficient development with plenty of extensions. Particularly, most implementations of machine learning in recent years have been written in Python. Therefore, Python is utilised as the scripting language in PULSim, and the open-source framework Py4J [29] is integrated within the software. Py4J enables Java programs to access Python code and also provides interfaces for Python programs to access Java objects. It works in a client-server mode for both directions.

In order to access Java objects from a Python program, the class GatewayServer should be at first initiated and started in Java code, taking an entry class as an entry point. The Python program will call the methods provided by the entry point through JavaGateway. Thereby, the Python program can access Java objects and services provided by PULSim. The mechanism of accessing Java objects and services from Python (Py2J) is shown in Figure 7.

To enable the Java code running in PULSim to access a Python script, a ClientServer class is initialised to start the Python script, which implements a listener interface defined in PULSim. Hence, the listener will be used as an entry point for Java code to access the functions provided by Python. The mechanism of accessing Python script from Java (J2Py) is shown in Figure 8.

The examples and the applications of using this bidirectional communication for user-based, adaptable simulation are described in Section 5.

5. Applications of User-Based Adaptable Simulation

A very common requirement for user-based, adaptable simulation is the ability to organise and customise several rounds of simulation as well as the desired output of simulation. For this purpose, PULSim provides an open interface for timetable simulation for Python. A simulator for timetable simulation can be obtained through the interface, in which users can designate the investigated data of infrastructure,

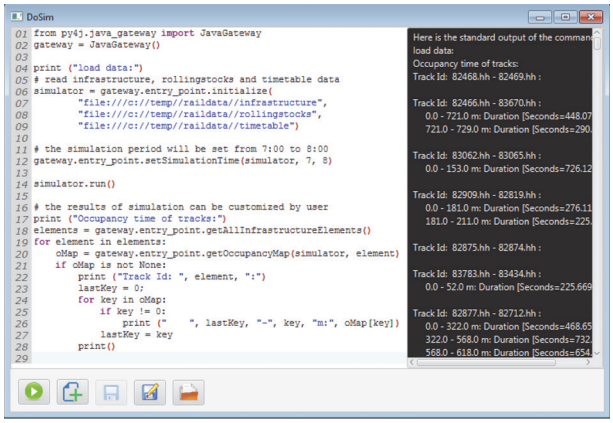


Figure 9: User interface for accessing Java services from Python in PULSim (Py2J).

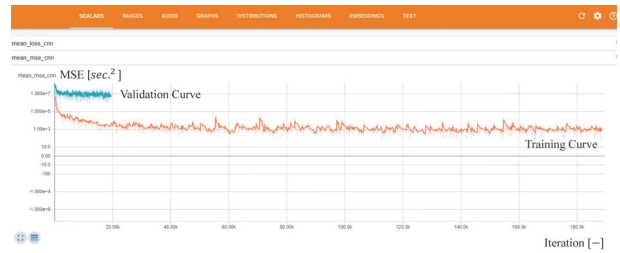


Figure 10: Prediction of total waiting time with TensorFlow (J2Py).

rolling stocks, and the timetable in the Python environment. Other configurations, for example, the time period of the simulation, can also be optionally specified for the simulator. With a few lines of code, a timetable simulation can be carried out in the Py2J mode (Figure 7).

The results of the simulation can be further analysed. For example, the occupancy and hindrance values during the simulation are matters of interest for purposes of capacity research. In Figure 9, the user interface for a timetable simulation with the outputs of occupancy and hindrance is shown. The editor for the Python program, the Py4J libraries, and the output window of the running Python program are integrated into PULSim as well.

In addition to being saved in log files for an offline evaluation, the output can also be directly retrieved by the Python program for further calculation and online evaluation. This is especially useful in order to customise a complicated workflow with many rounds of simulation. For example, to derive the recommended area of traffic flow within the scope of capacity research (source [10]) automatically, it is necessary to seamlessly integrate the software for capacity research with the simulation tools. Otherwise users would have to shift between different software tools manually to simulate several densified timetable variants. The advantage of the user-based, adaptable simulation can also be gained in the process of calibration of the parameters for operational simulation. In [16], an automatic process for calibration is implemented by the authors using PULSim. It is possible in this case, since the developers for the calibration system are also the developers of the simulation tool. However, if developers from a third party want to implement the calibration process, it is difficult if they do not have access and understanding of the source code of the simulation tools. A Py2J interface can provide the developer with the opportunity to adapt the disturbance parameters through many rounds of operational simulation iteratively, without the additional effort of interacting with the simulation tool.

Another advantage of user-based, adaptable simulation is the enhancement of the capabilities of the simulation tool through third-party software and algorithms. Today, machine learning plays an increasingly important role in learning and making predictions about data. Nowadays, many software packages for machine learning are implemented in Python, for example, Scikit-Learn [30], Theano [31], and Google TensorFlow [32]. PULSim is able to access these available Python-based software tools and to perform a learning process through J2Py communication (Figure 8). The desired functions for machine learning can be initially defined in a listener in PULSim. A Python script will implement the listener interface in order to carry out the learning process. PULSim will provide the training data for the Python-based software tools, which enables the learning process to be carried out either offline or online.

The authors are carrying out a case study, which integrates Google TensorFlow with PULSim to predict the total waiting time of a dispatching action. The case study is taken from a real railway network in Germany with 71 stations. There are in total 1350 trains running in the entire investigated day. The simulated timetable is not conflict-free; a dispatching action has to be taken for a certain state with occupancy conflicts. For a given state, there are several possible dispatching actions to be taken. The task of the dispatching system is to decide on a certain action as the dispatching decision, which results in the minimum possible waiting time. Therefore, the resulting total waiting time for a given state and a certain action has to be predicted exactly. This can be achieved through supervised learning. The object of learning is to minimise the Mean Squared Error (MSE) between the real waiting time and the predicted waiting time. During the simulation process, the states and the actions to be taken are recorded in the form of resulting states by PULSim as input (see Section 3.2), and the resulting total waiting time are learned as output. The total waiting time is learned and predicted by the Python script within the framework of TensorFlow.

With the case study, 12,000 resulting states and the total waiting time are fed into a Convolutional Neural Network (CNN) as training data, and 1,000 resulting states and the total waiting time are used as validation data. The CNN is at first trained with the training data and then validated with the validation data. The average value of the real waiting time for the training and validation data is 44.63 minutes. In Figure 10, the results are shown in Google TensorBoard, which is a tool provided for the visualisation of the learning process. The training curve represents the change of MSE for the training data along iterations, and the validation curve represents the MSE for the validation data along

iterations. After ten iterations of training on batch data, an epoch of validation is carried out. Hence, the iterations for training are ten times the iterations of validation. At the moment, the mean error of the predicted total waiting time from the training data is around 1 to 2 minutes, and the mean error of the predicted total waiting time from the validation data is around 8 minutes. The accuracy should be further improved through tuning the hyperparameter of the model.

6. Conclusion and Perspectives

In this paper, a user-based, adaptable simulation tool PULSim for railway planning and operations is presented. Supported by its open interface, users and researchers can rapidly develop their own algorithms to be integrated with PULSim. The productivity of using simulation tools for evaluation and optimisation of railway planning and operations will be improved significantly.

More interfaces will be provided in the further developments of PULSim. Special focus will be placed on the integration of varied signalling systems, dispatching algorithms, and the flexible configuration of operational simulation. The further developed interfaces will be built according to the demands and the feedback from users of the software. For example, in order to find an optimised setting of block sections, an interface to allow flexible placement of signals can be provided. The position of signals can be adjusted dynamically, so that the effects for different variants can be evaluated and optimised by the user-defined optimisation algorithm. In further development, additional dispatching actions, including the extension or shortening of train paths, along with the reordering or cancellation of train runs, should be integrated into PULSim for both railway simulation and real operations control.

Having the large amount of data generated by PULSim, various machine learning algorithms can be further applied to reveal and predict the system behaviour of complex railway systems. PULSim enables a simple integration of the simulation platform and popular machine learning software. Furthermore, it is a growing trend to carry out machine learning combined with the technology of big data. Since PULSim is developed in Java, it can be naturally integrated with Apache Hadoop [33], which is a Java-based software framework for managing and processing large amounts of data. The efficiency and the effectiveness of railway planning and operations will be continuously improved through integrating cutting-edge technologies with PULSim.

Conflicts of Interest

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

Acknowledgments

During the work, Mr. Kai Chen made his contribution to GUI programming. The text and the English have been proofread and improved by Mr. Syed Murtaza Hasan.

References

[1] RMCON, "RailSys 7 User Manual. Preparations for Multiple Simulation," 2010.

[2] D. Hürlimann, *Open Track Betriebssimulation von Eisenbahnnetzen*, Zürich, Switzerland, 1.6 edition, 2010.

[3] VIA Con, LUKS Handbuch Version 2.1, 2011.

[4] B. Kogel, N. Nießen, and T. Büker, *Influence of the European Train Control Sys-tem (ETCS) on the Capacity of Nodes*, International Union of Railways, Paris, France, 2010.

[5] M. Luethi, D. Huerlimann, and A. Nash, "Understanding the timetable planning process as a closed control loop," in *Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis*, I. A. Hansen, F. M. Dekking, R. M. P. Goverde, B. Heidergott, and L. E. Meester, Eds., Delft, Netherlands, 2005.

[6] A. Radtke, *EDV-Verfahren zur Modellierung des Eisenbahnbetriebs*, vol. 64, Eurailpress Tetzlaff-Hestra (Wissenschaftliche Arbeiten für den Schienenver-kehr, 64), Hamburg, Germany, 2005.

[7] A. Radtke and J. Bendfeldt, "Handling of railway operation problems with RailSys," in *Proceedings of the 5th World Congress on Rail Research*, Cologne, Germany, 2001.

[8] F. Corman, A. D'Ariano, and I. A. Hansen, "Evaluating disturbance robustness of railway schedules," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 18, no. 1, pp. 106–120, 2014.

[9] U. Martin, Y. Cui, F. Hantsch, Z. Chu, and X. Li, "Knotenkapazität - Bewertungs-verfahren für das Mikroskopische Leistungsverhalten und die Engpasserkennung im Spurgeführten Verkehr (RePlan)," in *VWI Neues verkehrswissenschaftliches Journal – Band 8*, Books on Demand GmbH Norderstedt, Norderstedt, Deutschland, 2014.

[10] U. Martin, C. Schmidt, and Z. Chu, "PULEIV Anwendungsleitfaden zur PULEIV-Version 2.1," *Anleitung zur Ermittlung des Leistungsverhaltens von Eisenbahninfra-struktur mithilfe von Simulationsprogrammen*, 2011.

[11] Y. Cui, "Simulation-Based Hybrid Model for a Partially-Automatic Dispatching of Railway Operation," *Norderstedt: Books on Demand GmbH (Neues verkehrswissenschaftliches Journal, 4)*, 2010.

[12] A. D'Ariano, M. Pranzo, and I. A. Hansen, "Conflict resolution and train speed coordination for solving real-time timetable perturbations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 208–222, 2007.

[13] I. A. Hansen and J. Pachl, *Railway Timetabling & Operations: Analysis - Modelling - Optimisation - Simulation - Performance Evaluation*, DW Media Group, Hamburg, Germany, 2014.

[14] J. Liang, *Metaheuristic-based Dispatching Optimization Integrated in Multi-scale Simulation Model of Railway Operation*, Institut für Eisenbahn- und Verkehrswesen der Universität Stuttgart, Stuttgart, Germany, 2017.

[15] Y. Cui, U. Martin, and J. Liang, "Searching feasible resources to reduce false-positive situations for resolving deadlocks with the Banker's algorithm in railway simulation," *Journal of Rail Transport Planning and Management*, vol. 7, no. 1-2, pp. 50–61, 2017.

[16] Y. Cui, U. Martin, and W. Zhao, "Calibration of disturbance parameters in railway operational simulation based on reinforcement learning," *Journal of Rail Transport Planning & Management*, vol. 6, no. 1, pp. 1–12, 2016.

[17] Y. Cui and U. Martin, "Multi-scale Simulation in Railway Planning and Operation," *PROMET - Traffic & Transportation*, vol. 23, no. 6, 2011.

[18] M. Wörner and Y. Cui, *Begriffslexikon - Datenmodell*, Institut für Eisenbahn- und Verkehrswesen der Universität Stuttgart, Stuttgart, Germany, 2008.

[19] U. Martin and J. Liang, "The Influence of Dispatching on the Relationship between Capacity and Operation Quality of Railway Systems (DFG Research Project MA 2326/15-1)," in *VWI Neues verkehrswissenschaftliches Journal – Band 20*, Books on Demand GmbH, Norderstedt, Deutschland, 2017.

[20] A. Nash, D. Huerlimann, J. Schütte, and V. P. Krauss, "RailML A Standard Data Interface for Railroad Applications," *WIT Transactions on The Built Environment*, vol. 74, 2004.

[21] W. Fang, S. Yang, and X. Yao, "A Survey on Problem Models and Solution Approaches to Rescheduling in Railway Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2997–3016, 2015.

[22] E. G. Coffman, M. Elphick, and A. Shoshani, "System Deadlocks," *ACM Computing Surveys*, vol. 3, no. 2, pp. 67–78, 1971.

[23] G. Mills and P. Pudney, "The Effects of Deadlock Avoidance on Rail Network Capacity and Performance," in *Proceedings of the 2003 Mathematics-in-Industry Study Group*, Brisbane, Australia, 2003.

[24] E. R. Petersen and A. J. Taylor, "Line block prevention in rail line dispatch and simulation models," *Information Systems and Operations Research*, vol. 21, no. 1, pp. 46–51, 1983.

[25] J. Pachl, "Deadlock Avoidance in Railroad Operations Simulations," in *Proceedings of the 90th Annual Meeting of Transportation Research Board*, Washington, DC, USA, 2011.

[26] R. Mittermayr, J. Blieberger, and A. Schöbel, "Kronecker algebra-based deadlock analysis for railway systems," *Promet - Traffic - Traffico*, vol. 24, no. 5, pp. 359–369, 2012.

[27] E. W. Dijkstra, "The mathematics behind the Banker's algorithm," in *Selected Writings on Computing: A Personal Perspective*, pp. 308–312, Springer-Verlag, New York, NY, USA, 1982.

[28] B. Seybold and D. Huerlimann, "OpenTrack - Simulation of railway systems. Presentation of the OpenTrack API. OpenTrack," http://www.opentrack.ch/opentrack/downloads/OpenTrack.API.pdf 2016.

[29] Py4J, "Py4J – A Bridge between Python and Java," https://www.py4j.org/, 2007.

[30] F. Pedregosa, G. Varoquaux, and A. Gramfort, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] J. Bergstra, O. Breuleux, P. Lamblin et al., "Theano: Deep learning on gpus with python," in *NIPS 2011, Big Learning Workshop*, vol. 3, Granada, Spain, 2011.

[32] M. Abadi, A. Agarwal, P. Barham et al., "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," https://arxiv.org/abs/1603.04467, 2016.

[33] J. Nandimath, E. Banerjee, A. Patil, P. Kakade, and S. Vaidya, "Big data analysis using Apache Hadoop," in *Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration, IEEE IRI 2013*, pp. 700–703, San Francisco, CA, USA, August 2013.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 5983250, 15 pages
https://doi.org/10.1155/2018/5983250

Research Article

Defining Reserve Times for Metro Systems: An Analytical Approach

Luca D'Acierno,[1] **Marilisa Botte**,[1] **Mariano Gallo**,[2] **and Bruno Montella**[1]

[1]*Department of Civil, Architectural and Environmental Engineering, Federico II University of Naples, Via Claudio 21, 80125 Naples, Italy*
[2]*Department of Engineering, University of Sannio, Piazza Roma 21, 82100 Benevento, Italy*

Correspondence should be addressed to Luca D'Acierno; luca.dacierno@unina.it

Received 17 November 2017; Revised 7 February 2018; Accepted 7 March 2018; Published 16 April 2018

Academic Editor: Andrea D'Ariano

The aim of this paper is to provide an analytical approach for determining operational parameters for metro systems so as to support the planning and implementation of energy-saving strategies. Indeed, one of the main targets of train operating companies is to identify and implement suitable strategies for reducing energy consumption. For this purpose, researchers and practitioners have developed energy-efficient driving profiles with the aim of optimising train motion. However, as such profiles generally entail an increase in travel times, the operating parameters in the planned timetable need to be appropriately recalibrated. Against this background, this paper develops a suitable methodology for estimating reserve times which represent the main rate of extra time needed to put ecodriving strategies in place. Our proposal is to exploit layover times (i.e., times spent by a train at the terminus waiting for the next trip) for energy-saving purposes, keeping buffer times intact in order to preserve the flexibility and robustness of the timetable in case of delays. In order to show its feasibility, the approach was applied in the case of a real metro context, whose service frequency was duly taken into account. In particular, after stochastic analysis of the parameters involved for calibrating suitable buffer times, different operating schemes were simulated by analysing the relationship between layover times, number of convoys, and feasible headway values. Finally, some operation configurations are analysed in order to quantify the amount of energy that can be saved.

1. Introduction

Rail transport systems are undoubtedly the best solution to satisfy high travel demand in densely populated areas: they offer high transport capacity, are unaffected by traffic congestion, and are able to reduce, capturing users from road transport, greenhouse gas (GHG) and air pollutant emissions. The importance of rail transport systems has increased in recent decades since environmental issues assumed a crucial role in transport policy at all territorial levels, whether worldwide, continental, country-wide, or at the more local scale. For instance, most of the European Union's [1] transport policy targets environmental issues and especially transport emissions. Indeed, to achieve the EU's objective to reduce transport emissions by 60% by the year 2050, two main strategies are involved: promotion of rail transport for both passengers and goods and improvement in vehicle energy efficiency.

Since 72.8% of GHG emissions in the transportation sector are produced by road traffic [2], it is important to increase the attractiveness of rail transport. This can be partly achieved by promoting intermodality between trains and buses [3–5]. Other strategies for promoting the use of rail systems are to increase service quality [6–8], reliability [9–11], and competitiveness with personal cars, by optimising service schedules and operations [12–15].

Energy efficiency in rail systems can be achieved by operating on vehicles, engines, and/or driving styles. In this paper we refer to driving styles that affect and interact with schedules: an energy-efficient driving style (or eco-driving) may need to modify running times and schedules.

As will be clarified in the next section, eco-drive strategies are mainly based on available reserve times on each railway section (between two stations) or, in some cases, at the terminus. Indeed, eco-driving increases train running times

and, if the scheduled timetables are to be maintained, only the available reserve time can be used.

When a timetable is designed, reserve times are usually defined according to the expected regularity of the service, without considering their (possible) utility in energy-efficient driving strategies. The aim of this paper is to provide an analytical approach for defining operational parameters for metro systems, including reserve times, considering explicitly their utility for implementing energy-saving strategies.

The paper is organised as follows: in Section 2, railway energy-efficient driving is examined; in Section 3 the analytical approach of our proposal is described; in Section 4 the methodology is applied in the case of a real metro line; finally, conclusions and research prospects are summarised in Section 5.

2. Railway Energy-Efficient Driving

Reducing the energy consumption of railway systems is currently considered an important objective in the transportation sector. Several approaches can be adopted to achieve this result: the design of low-consumption engines, the provision of efficient energy recovery systems, and the optimisation of both schedules, so as to reduce peaks in energy demand and train diving styles.

Energy-efficient driving, or eco-driving, has been widely studied in the literature. Most approaches are based on the application of optimal control theory in the case of railway contexts [16–18].

The effectiveness of energy-efficient driving strategies was underlined in [19, 20]. Interesting results were obtained in [21], where driving profiles were optimised by considering as constraints the operational requirements of railways. In [22], an optimisation framework based on a genetic algorithm was proposed for minimising energy consumption.

Other papers focusing on this issue are [23–26]; real-time optimisation was studied in [27, 28]. In [29], the potential effects of energy-efficient driving profiles on consumption were studied with a parametric approach in an Italian region; the results showed that up to 20% of total traction energy can be saved by optimising speed profiles and up to 35% if trains are equipped with braking energy recovery systems. Simulation-based methods can be found in [30–33].

The impact and integration of such procedures within the more complex problem of optimal train scheduling were studied in [34, 35]. Finally, in [36] energy-saving driving methods were applied to freight trains.

The starting point of any analysis is to define the mechanical kinetic energy required to move a rail convoy during a time interval which may be calculated as follows:

$$E(\Delta t) = \int_0^{\Delta t} dE(\tau) = \int_0^{\Delta t} P(\tau) \cdot d\tau = \int_0^{\Delta t} T(v(\tau)) \cdot v(\tau) \cdot d\tau, \quad (1)$$

where $E(\Delta t)$ is the kinetic energy required during time interval Δt; Δt is the generic time interval; $dE(\tau)$ is the increase in kinetic energy at time τ; τ is the generic time instant; $P(\tau)$ is the instantaneous power at time τ; $d\tau$ is the generic infinitesimal time interval; $v(\tau)$ is the instantaneous speed at time τ; and $T(\cdot)$ is the tractive effort (i.e., tractive force) at rail wheels which depends on instantaneous speed $v(\cdot)$.

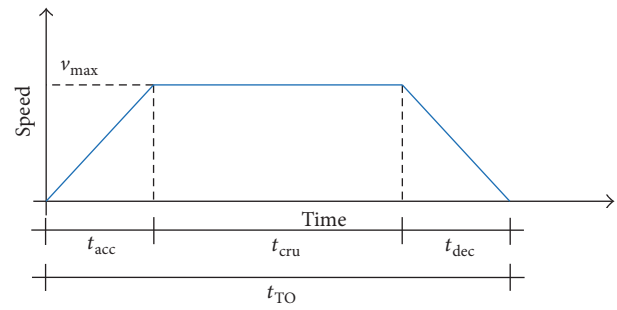


Figure 1: Speed profile in the case of a *Time Optimal* (TO) strategy.

Most approaches proposed in the literature (see, e.g., [37–39]) are based on the definition of a reference scenario, indicated as the *Time Optimal* (TO) scenario, which consists in considering the movement of the rail convoy in the case of maximum performance: the train first accelerates with maximum acceleration to reach the maximum speed *(acceleration phase)*, then travels at constant maximum speed *(cruising phase)*, and finally brakes at maximum deceleration *(deceleration phase)*. This scenario provides the lowest travel time but the maximum energy consumption.

Figure 1 provides a generic speed profile in the case of a TO strategy with the simplifying assumption of instantaneous variation of acceleration (i.e., the jerk value equal to $+\infty$), while in real cases, the jerk value is a finite quantity.

The total travel time between two successive stops (i.e., stations or red signals) in the case of strategy TO may be formulated as follows:

$$t_{TO} = t_{acc} + t_{cru} + t_{dec}, \quad (2)$$

where t_{TO} is the travel time in the case of a TO strategy; t_{acc} is the time duration of the acceleration phase; t_{cru} is the time duration of the cruising phase; and t_{dec} is the time duration of the deceleration phase.

Since parameter fluctuations are generally stochastic (e.g., a driver may adopt different values of acceleration, speed, and deceleration), when designing timetables UIC [40] suggests increasing the *total travel time* (i.e., the minimum running time) by a *running time reserve* of about 3–8% in order to provide a reserve in the case of reduced performance, thereby avoiding the onset of delays. Most *Energy-Saving* (ES) strategies proposed in the literature are based on adopting different speed profiles which consume less energy but require more time. The additional time may be obtained by consuming a part of the reserve time. In particular, these strategies may be classified according to two approaches:

(i) The first, where between the cruising and deceleration phases an additional phase is inserted, referred to as the *coasting phase*, where the engine is stopped and the train moves by inertia

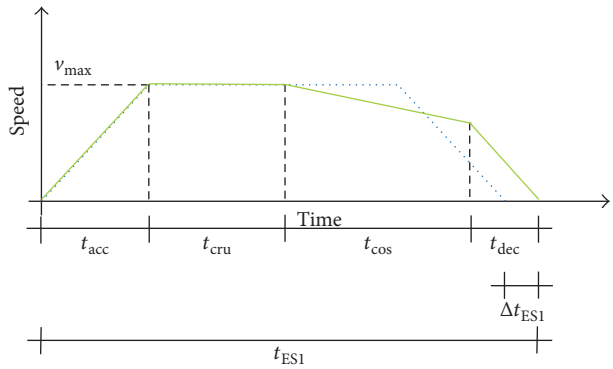


Figure 2: Speed profile in the case of *Energy-Saving* (ES) strategy 1.

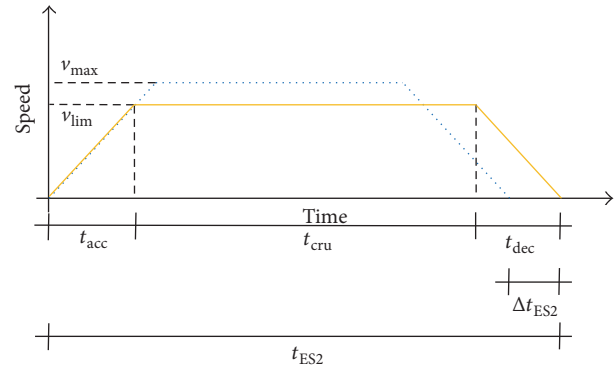


Figure 3: Speed profile in the case of *Energy-Saving* (ES) strategy 2.

(ii) The second where a lower maximum speed is adopted.

Figure 2 provides a generic speed profile in the case of the first ES strategy with the same assumption of jerk values adopted in the case of strategy TO. In this case, the total travel time between two successive stops (i.e., stations or red signals) may be formulated as

$$t_{ES1} = t_{acc} + t_{cru} + t_{cos} + t_{dec} \tag{3}$$

or equivalently

$$t_{ES1} = t_{TO} + \Delta t_{ES1}, \tag{4}$$

where t_{ES1} is the travel time in the case of the first ES strategy; t_{cos} is the time duration of the coasting phase; and Δt_{ES1} is the increase in travel time in the case of the first ES strategy with respect to strategy TO.

The increase in travel time due to implementing the first ES strategy may be calculated by imposing the constancy of the section length (i.e., the travel distance is independent of the implemented strategy); that is,

$$\Delta s = \int_0^{t_{TO}} v_{TO}(\tau) \cdot d\tau = \int_0^{t_{ES1}} v_{ES1}(\tau) \cdot d\tau, \tag{5}$$

where Δs is the track length between the two successive stops analysed, $v_{TO}(\cdot)$ is the speed profile in the case of the TO strategy as shown in Figure 1, and $v_{ES1}(\cdot)$ is the speed profile in the case of the first ES strategy as shown in Figure 2. Hence, by replacing (4) in (5), we obtain

$$\int_{t_{TO}}^{t_{TO}+\Delta t_{ES1}} v_{ES1}(\tau) \cdot d\tau = \int_0^{t_{TO}} \left(v_{TO}(\tau) - v_{ES1}(\tau)\right) \cdot d\tau, \tag{6}$$

where the increase in travel time, that is, term Δt_{ES1}, represents the unknown variable to be determined.

Likewise, Figure 3 provides a generic speed profile in the case of the second ES strategy with the same assumption of jerk values adopted in the case of the TO strategy. In this scenario, the total travel time between two successive stops may be formulated as

$$t_{ES2} = t_{acc} + t_{cru} + t_{dec} \tag{7}$$

or equivalently

$$t_{ES2} = t_{TO} + \Delta t_{ES2}, \tag{8}$$

where t_{ES2} is the travel time in the case of the second ES strategy and Δt_{ES2} is the increase in travel time in the case of the second ES strategy with respect to the TO strategy.

Also in the case of the second ES strategy, the increase in travel time may be calculated by imposing the constancy of the section length; that is,

$$\Delta s = \int_0^{t_{TO}} v_{TO}(\tau) \cdot d\tau = \int_0^{t_{ES2}} v_{ES2}(\tau) \cdot d\tau, \tag{9}$$

where $v_{ES2}(\cdot)$ is the speed profile in the case of the second ES strategy, as shown in Figure 3.

In this case, by replacing (8) in (9), we obtain

$$\int_{t_{TO}}^{t_{TO}+\Delta t_{ES2}} v_{ES2}(\tau) \cdot d\tau = \int_0^{t_{TO}} \left(v_{TO}(\tau) - v_{ES2}(\tau)\right) \cdot d\tau, \tag{10}$$

where the increase in travel time, that is, term Δt_{ES2}, represents the unknown variable to be determined.

Adoption of the first or second approach depends on the technological level of the rail system. Indeed, the first approach requires the need to communicate to the train the precise point at which the coasting phase begins. By contrast, the second approach requires simply communicating a different speed limit and hence is more straightforward to implement. However, most of the contributions in the literature are based on considering the implementation of ES strategies between two successive stops in order to preserve arrivals and departures at each station.

Since a metro system may be considered a frequency service [41, 42] where it is necessary to observe the average headway between two successive convoys rather than a timetable generally unknown to users, we propose to apply the ES strategies by considering as mainstay the arrival and departure times only at the terminus and not at the intermediate stations. Hence, we suggest implementing any strategy by considering the entire outward and return trip. Obviously, implementation of any optimisation approach for ensuring lower energy consumption requires the knowledge and estimation of reserve times according to the formulation proposed in the paper.

3. The Analytical Approach

The aim of this paper is to determine the value of reserve times in a metro system in order to quantify expendable time resources for reducing energy consumption. The use of an analytical approach arises from the need to have a system of equations which allows a priori exclusion of all unfeasible operative service configurations, as shown below. However, since both ES strategies described in the previous section are based on the extension of travel time, the following analytical formulation can be applied in both cases. Moreover, in the case of more complex layouts, such as in the case of rail systems, the proposed approach may be easily integrated with suitable train timetabling optimisation techniques.

Generally, we may define the minimum cycle time, indicated as CT_{min}, as the minimum time required by a convoy to complete the outward trip and the following return trip and achieve the initial condition. It can be calculated as follows:

$$CT_{min} = \sum_{lot} tt_{lot} + \sum_{sot} dt_{sot} + it_{ot} + \sum_{lrt} tt_{lrt} + \sum_{srt} dt_{srt} + it_{rt}, \tag{11}$$

where tt_{lot} and tt_{lrt} are the travel times associated, respectively, with links lot and lrt; lot and lrt are the generic links (i.e., track sections) associated, respectively, with the outward trip (ot) and return trip (rt); dt_{sot} and dt_{srt} are the dwell times associated, respectively, with platforms sot and srt; sot and srt are the generic platforms of station s for, respectively, the outward trip (ot) and return trip (rt); and it_{ot} and it_{rt} are the inversion times (i.e., preparation times for the subsequent trip) associated, respectively, with the outward trip (ot) and return trip (rt).

However, in the ordinary operations to avoid delay propagation, the adopted cycle time, indicated as CT_{ord}, may be obtained by increasing the minimum value as follows:

$$CT_{ord} = CT_{min} + bt_{ot} + bt_{rt}, \tag{12}$$

where bt_{ot} and bt_{rt} are the buffer times used for recovery of delays, respectively, in the case of the outward trip (ot) and return trip (rt) due to variation in travel, dwell, and inversion times.

Finally, in order to satisfy timetable requirements, the planned cycle time, indicated as CT_{plan}, may be calculated as follows:

$$CT_{plan} = CT_{ord} + lt_{ot} + lt_{rt}, \tag{13}$$

where lt_{ot} and lt_{rt} are the layover times (i.e., times spent by a train at the terminus waiting for the subsequent trip) associated, respectively, with the outward trip (ot) and return trip (rt).

In this context, in order to maintain a constant headway between two successive convoys, it is necessary to impose the following constraints:

$$0 \le bt_{ot} + lt_{ot} \le H \tag{14}$$

$$0 \le bt_{rt} + lt_{rt} \le H, \tag{15}$$

where H is the headway between two successive rail convoys.

The number of convoys, indicated as NC, required to perform the service may be calculated as

$$NC = \frac{CT_{plan}}{H}. \tag{16}$$

By substituting (13) into (16), we obtain

$$lt_{ot} + lt_{rt} = NC \cdot H - CT_{ord}. \tag{17}$$

Likewise, since the layover times must be non-negative, combining (14) and (15) with the non-negativity condition produces

$$0 \le lt_{ot} + lt_{rt} \le 2 \cdot H - \left(bt_{ot} + bt_{rt}\right). \tag{18}$$

Finally, by substituting (17) into (18), we obtain

$$\frac{CT_{ord}}{H} \le NC \le 2 + \frac{CT_{ord}}{H} - \frac{bt_{ot} + bt_{rt}}{H}. \tag{19}$$

Therefore, since NC must be an integer value, we obtain the following constraints:

$$NC_{min} = \text{int}\left(\frac{CT_{ord}}{H}\right) + 1$$
$$NC_{max} = \text{int}\left(2 + \frac{\left(CT_{ord} - \left(bt_{ot} + bt_{rt}\right)\right)}{H}\right), \tag{20}$$

where NC_{min} and NC_{max} are, respectively, the minimum and the maximum number of rail convoys required to perform the service.

A number of convoys lower than NC_{min} are unable to ensure a headway of H due to lack of rolling stock. A number of convoys higher than NC_{max} provide an interference in terms of movements among convoys which prevents them from maintaining their cycle time.

According to the definition of reserve time provided by UIC [40], in our approach it is the sum of two quantities, the buffer time and the layover time; that is,

$$rt_{ot} = bt_{ot} + lt_{ot}$$
$$rt_{rt} = bt_{rt} + lt_{rt}, \tag{21}$$

where rt_{ot} and rt_{rt} are the reserve times for, respectively, the outward and return trips.

Since the layover times (i.e., terms lt_{ot} and lt_{rt}) represent quantities that a train loses to maintain the headway, our proposal consists in adopting these times as storage tanks for incrementing travel times without affecting service frequencies (expressed in terms of H) so as to implement suitable ES strategies. Hence, we may define the total usable reserve time, that is, the term turt, which represents the sum of the two layover times; that is,

$$turt = lt_{ot} + lt_{rt} \tag{22}$$

since in our approach we assume that buffer times (i.e., bt_{ot} and bt_{rt}) used for recovery of delays are kept intact and are not used to implement ES strategies.

In this context, (17) may be expressed as

$$\text{turt} = \text{NC} \cdot H - \text{CT}_{\text{ord}}. \tag{23}$$

Theoretically, the turt value could be split arbitrarily between the outward and the return trip. Hence, we may define a new parameter, indicated as α, which expresses the partition rate as follows:

$$\text{lt}_{\text{ot}} = \alpha \cdot \text{turt} \tag{24}$$

$$\text{lt}_{\text{rt}} = (1 - \alpha) \cdot \text{turt} \tag{25}$$

with $\alpha \in [0; 1]$.

By substituting (24) into (14), we obtain

$$0 \leq \alpha \leq \frac{(H - \text{bt}_{\text{ot}})}{\text{turt}}. \tag{26}$$

Likewise, by substituting (25) into (15), we obtain

$$1 - \frac{(H - \text{bt}_{\text{rt}})}{\text{turt}} \leq \alpha \leq 1. \tag{27}$$

Since parameter α has to satisfy jointly (26) and (27), and it is not possible to state a priori whether $(H - \text{bt}_{\text{ot}})/\text{turt}$ is greater or lower than $1 - (H - \text{bt}_{\text{rt}})/\text{turt}$, we need to identify two feasible conditions which are mutually exclusive (i.e., disjoint):

(i) Condition 1: $1 - (H - \text{bt}_{\text{rt}})/\text{turt} \leq (H - \text{bt}_{\text{ot}})/\text{turt}$

(ii) Condition 2: $(H - \text{bt}_{\text{ot}})/\text{turt} < 1 - (H - \text{bt}_{\text{rt}})/\text{turt}$.

In the first case, it is possible to identify the feasibility set of parameter α as follows:

$$\max\left\{0; 1 - \frac{H - \text{bt}_{\text{rt}}}{\text{turt}}\right\} \leq \alpha \leq \min\left\{\frac{H - \text{bt}_{\text{ot}}}{\text{turt}}; 1\right\}, \tag{28}$$

while, in the second case, the feasibility set of parameter α is an empty set since (26) and (27) identify disjointed sets. However, by reductio ad absurdum, we may state that the second condition never occurs. Indeed, condition 2 may be formulated as

$$\text{turt} > 2 \cdot H - (\text{bt}_{\text{ot}} + \text{bt}_{\text{rt}}). \tag{29}$$

By substituting (23) into (29), we obtain

$$\text{NC} \cdot H - \text{CT}_{\text{ord}} > 2 \cdot H - (\text{bt}_{\text{ot}} + \text{bt}_{\text{rt}}). \tag{30}$$

Moreover, by substituting (12) into (30), we obtain

$$\text{NC} > 2 + \left(\frac{\text{CT}_{\min}}{H}\right). \tag{31}$$

Since, by substituting (12) and (13) into (16), we obtain

$$\text{NC} = \frac{(\text{CT}_{\min} + (\text{bt}_{\text{ot}} + \text{bt}_{\text{rt}}) + (\text{lt}_{\text{ot}} + \text{lt}_{\text{rt}}))}{H}, \tag{32}$$

it is possible to substitute (32) into (31), thus obtaining

$$\frac{(\text{bt}_{\text{ot}} + \text{bt}_{\text{rt}}) + (\text{lt}_{\text{ot}} + \text{lt}_{\text{rt}})}{H} > 2. \tag{33}$$

Combining (14) and (15) results in the following:

$$0 \leq \frac{(\text{bt}_{\text{ot}} + \text{bt}_{\text{rt}}) + (\text{lt}_{\text{ot}} + \text{lt}_{\text{rt}})}{H} \leq 2 \tag{34}$$

which expresses a contradiction between condition 2 (which provides (33)) and the constant headway constraint (which provides (34)). Therefore, this contradiction implies that condition 2 never occurs.

In conclusion, although parameter α belongs to set $[0; 1]$, its value cannot be fixed arbitrarily but has to satisfy condition (28). Hence, we define a minimum, $\alpha_{\min}$, and a maximum, $\alpha_{\max}$, value of parameter α, that is,

$$\begin{aligned} \alpha_{\min} &= \max\left\{0; 1 - \frac{H - \text{bt}_{\text{rt}}}{\text{turt}}\right\} \\ \alpha_{\max} &= \min\left\{\frac{H - \text{bt}_{\text{ot}}}{\text{turt}}; 1\right\}. \end{aligned} \tag{35}$$

Importantly, having fixed the infrastructure and the signalling system, the minimum headway between two successive convoys depends on inversion times and the main features of the signalling system, that is,

$$H_{\min} = \max\left\{\text{ts}_{\text{ot}}^{\text{inv}}; \text{ts}_{\text{rt}}^{\text{inv}}; \text{ts}_{\text{ot}}^{\text{ls}}; \text{ts}_{\text{rt}}^{\text{ls}}; \text{ts}_{\text{min-ss}}\right\}, \tag{36}$$

where $H_{\min}$ is the minimum value of H; $\text{ts}_{\text{ot}}^{\text{inv}}$ is the time spacing due to the inversion of the rail convoy at the final terminus of the outward trip; $\text{ts}_{\text{rt}}^{\text{inv}}$ is the time spacing due to the inversion of the rail convoy at the final terminus of the return trip; $\text{ts}_{\text{ot}}^{\text{ls}}$ is the time spacing of the rail convoy at the last section of the outward trip (to access the first station of the return trip); $\text{ts}_{\text{rt}}^{\text{ls}}$ is the time spacing of the rail convoy at the last section of the return trip (to access the first station of the subsequent trip which is an outward trip); and $\text{ts}_{\text{min-ss}}$ is the minimum time spacing allowed by the signalling system along the line which has to take into account dwell times at stations and circulation rules (such as the criterion of station releasing).

Term $\text{ts}_{\text{ot}}^{\text{inv}}$ (or equivalently $\text{ts}_{\text{rt}}^{\text{inv}}$) and term $\text{ts}_{\text{ot}}^{\text{ls}}$ (or equivalently $\text{ts}_{\text{rt}}^{\text{ls}}$) are strictly related to the framework of the infrastructure, signalling system, and the service organisation. Indeed, these values depend on the place where trains stop for the buffer and layover times. For instance, if these times are spent on the inversion track, terms $\text{ts}_{\text{ot}}^{\text{inv}}$ and $\text{ts}_{\text{rt}}^{\text{inv}}$ have to include travel times along inversion tracks, stop times for waiting for technical inversion operations (such as the transfer of the driver from the previous head to the new head which was the previous tail), the buffer time for recovery delays, the layover time for achieving the planned headway, and, finally, the times related to the signalling system functioning (such as the release time for unlocking the block system). Alternatively, if buffer and layover times are spent at the first station of the subsequent trip, these times have to be included in the definition of terms $\text{ts}_{\text{ot}}^{\text{ls}}$ and $\text{ts}_{\text{rt}}^{\text{ls}}$, together with the dwell time planned for that station.

Therefore, in both cases, the allocation of layover times between the outward and return trips (i.e., the value of

parameter α) affects the definition of the minimum headway (i.e., $H_{\min}$) by means of (36).

The feasibility of a planned headway H may be verified by means of the following condition:

$$H \geq H_{\min}. \tag{37}$$

Since the value of H and the adopted number of convoys NC provide univocally the term turt according to (23) and the allocation of turt between the outward and return trips (i.e., the value of parameter α) affects the definition of the minimum headway (i.e., $H_{\min}$) by means of (36), formally (36) may be expressed as follows:

$$H_{\min} = H_{\min}\left(\text{turt}\left(H, \text{NC}\right), \alpha\right) \tag{38}$$

subject to

$$\text{NC} \in \left[\text{NC}_{\min}\left(H\right); \text{NC}_{\max}\left(H\right)\right] \tag{39}$$

$$\alpha \in \left[\alpha_{\min}\left(H\right); \alpha_{\max}\left(H\right)\right], \tag{40}$$

where (39) concisely expresses constraints (20) and (40) concisely expresses (35).

In this context, in order to verify the feasibility of a generic headway H to be planned, for any value of NC which satisfies condition (39), we may identify an optimal value of α, indicated as α_{opt}, which minimises $H_{\min}$, that is,

$$\alpha_{\text{opt}} = \arg\min_{\alpha} H_{\min}\left(\alpha\right) \tag{41}$$

with

$$\alpha \in \left[\alpha_{\min}; \alpha_{\max}\right] \tag{42}$$

and we can verify whether condition (37) is satisfied.

However, if trains spend buffer and layover times on the inversion track, variation in parameter α

(i) does not affect the value of terms $\text{ts}_{\text{ot}}^{\text{ls}}$, $\text{ts}_{\text{rt}}^{\text{ls}}$, or $\text{ts}_{\text{min-ss}}$;

(ii) affects term $\text{ts}_{\text{ot}}^{\text{inv}}$, according to a linear strictly increasing function;

(iii) affects term $\text{ts}_{\text{rt}}^{\text{inv}}$, according to a linear strictly decreasing function.

Alternatively, if trains spend buffer and layover times at the first station of the subsequent trip, variation in parameter α

(i) does not affect the value of terms $\text{ts}_{\text{ot}}^{\text{inv}}$, $\text{ts}_{\text{rt}}^{\text{inv}}$, or $\text{ts}_{\text{min-ss}}$;

(ii) affects term $\text{ts}_{\text{ot}}^{\text{ls}}$, according to a linear strictly increasing function;

(iii) affects term $\text{ts}_{\text{rt}}^{\text{ls}}$, according to a linear strictly decreasing function.

In both cases, function $H_{\min}(\alpha)$ is convex, that is,

$$\begin{aligned} &H_{\min}\left(\lambda \cdot \alpha_1 + (1-\lambda) \cdot \alpha_2\right) \\ &\quad \leq \lambda \cdot H_{\min}\left(\alpha_1\right) + (1-\lambda) \cdot H_{\min}\left(\alpha_2\right) \\ &\qquad \forall \alpha_1, \alpha_2 \in \left[\alpha_{\min}; \alpha_{\max}\right] \ \ \forall \lambda \in [0;1]. \end{aligned} \tag{43}$$

Moreover, if

$$\max\left\{\text{ts}_{\text{ot}}^{\text{inv}}; \text{ts}_{\text{rt}}^{\text{inv}}\right\} > \max\left\{\text{ts}_{\text{ot}}^{\text{ls}}; \text{ts}_{\text{rt}}^{\text{ls}}; \text{ts}_{\text{min-ss}}\right\} \quad \forall \alpha \in \left[\alpha_{\min}; \alpha_{\max}\right] \tag{44}$$

in the case of buffer and layover times spent at the inversion track, or

$$\max\left\{\text{ts}_{\text{ot}}^{\text{ls}}; \text{ts}_{\text{rt}}^{\text{ls}}\right\} > \max\left\{\text{ts}_{\text{ot}}^{\text{inv}}; \text{ts}_{\text{rt}}^{\text{inv}}; \text{ts}_{\text{min-ss}}\right\} \quad \forall \alpha \in \left[\alpha_{\min}; \alpha_{\max}\right] \tag{45}$$

in the case of buffer and layover times spent at the first station of the subsequent trip, function $H_{\min}(\alpha)$ is strictly convex, that is,

$$\begin{aligned} &H_{\min}\left(\lambda \cdot \alpha_1 + (1-\lambda) \cdot \alpha_2\right) \\ &\quad < \lambda \cdot H_{\min}\left(\alpha_1\right) + (1-\lambda) \cdot H_{\min}\left(\alpha_2\right) \\ &\qquad \forall \alpha_1, \alpha_2 \in \left[\alpha_{\min}; \alpha_{\max}\right] \ \ \forall \lambda \in [0;1]. \end{aligned} \tag{46}$$

The convexity (or the strict convexity) of function $H_{\min}(\alpha)$ allows us to solve the constrained optimisation problem (41) subject to (42) by adopting traditional solution algorithms for convex objective function problems.

However, the proposed approach, being considered a decision support system, has to be applied in planning phases, that is, by considering ordinary conditions. Obviously, an extension to the disruption conditions may be easily obtained by affecting operating values with the corresponding disrupted values.

4. Application of the Proposed Methodology

In order to validate the proposed formulation, we applied it to the case of a real metro line: Line 1 of the Naples metro system in southern Italy. This line, which is about 18 km long, connects the suburbs with the city centre, and its outward and return routes are strongly asymmetric in terms of elevations (Figure 4), which also entails asymmetry in terms of energy consumption (as shown in Table 1).

The first step of the application was to calibrate the commercial software OpenTrack® [43] in order to provide a mathematical model which allows simulation of all phases of the metro service (i.e., travel, dwell, and inversion times) and also exploration of different operational configurations without the need to apply them physically (i.e., on the real line). In particular, all infrastructure, rolling stock, and signalling system data were appropriately adopted for tuning the model (details on the calibration techniques may be found in [33, 44, 45]).

Travel, dwell, and inversion times were obtained by implementing a deterministic simulation of the metro service (see Table 1).

Likewise, by means of 200 stochastic simulations (for details on the use of stochastic simulations see, e.g., [46]), we were able to determine the statistical distribution of all service parameters so as to provide buffer times and related

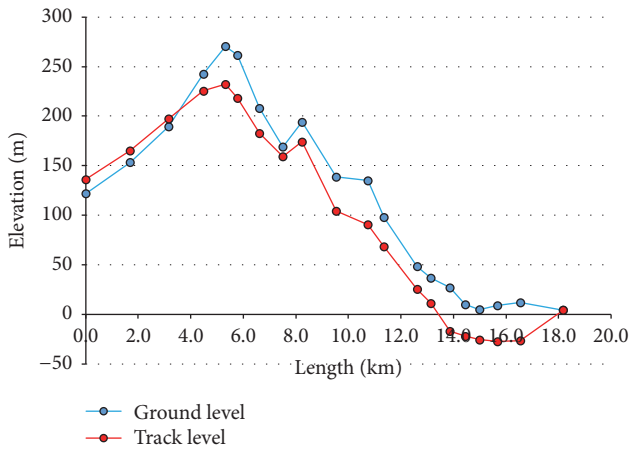


Figure 4: Elevation profile of Line 1 (from Piscinola to Garibaldi).

Table 1: Operational parameters of Line 1.

Parameter	Value	
	Piscinola-Garibaldi direction	Garibaldi-Piscinola direction
Travel distance	18.791 km	18.616 km
Total travel time	1,463 s (24.4 min)	1,485 s (24.8 min)
Total dwell time	400 s (6.7 min)	400 s (6.7 min)
Inversion time	307 s (5.1 min)	268 s (4.5 min)
Buffer time [90th percentile]	116 s (1.9 min)	103 s (1.7 min)
Buffer time [95th percentile]	131 s (2.2 min)	116 s (1.9 min)
Buffer time [99th percentile]	159 s (2.7 min)	141 s (2.4 min)
Adopted cycle time CT_{ord} [90th percentile]	4,542 s (75.7 min)	
Adopted cycle time CT_{ord} [95th percentile]	4,570 s (76.2 min)	
Adopted cycle time CT_{ord} [99th percentile]	4,623 s (77.1 min)	
Minimum time spacing along the line $ts_{min\text{-}ss}$	110 s (1.83 min)	
Energy consumption	279.01 kWh	386.21 kWh

cycle times as functions of an assumed confidence level. Indeed, differences in performance between the stochastic and deterministic travel times may be calculated as follows:

$$\delta_{ot}^{i} = \left(\sum_{lot} tt_{lot}^{i,STOC} + \sum_{sot} dt_{sot}^{i,STOC} + it_{ot}^{i,STOC} \right) - \left(\sum_{lot} tt_{lot}^{DET} + \sum_{sot} dt_{sot}^{DET} + it_{ot}^{DET} \right)$$

$$\delta_{rt}^{i} = \left(\sum_{lrt} tt_{lrt}^{i,STOC} + \sum_{srt} dt_{srt}^{i,STOC} + it_{rt}^{i,STOC} \right) - \left(\sum_{lot} tt_{lrt}^{DET} + \sum_{srt} dt_{srt}^{DET} + it_{rt}^{DET} \right), \quad (47)$$

where δ_{ot}^{i} and δ_{rt}^{i} represent, respectively, the difference in the case of outward and return trips at the ith stochastic simulation; $X^{i,STOC}$ represents the value of variable X (where X represents the travel time, dwell time, or inversion time) in the case of the ith stochastic simulation; and X^{DET} represents the value of variable X in the case of a deterministic simulation.

Since stochastic simulations are based on reductions in train performance, we may assume that

$$X^{i,STOC} \geq X^{DET} \quad \forall i \; \forall X \quad (48)$$

which implies that

$$\delta_{ot}^{i} \geq 0 \quad \forall i$$
$$\delta_{rt}^{i} \geq 0 \quad \forall i. \quad (49)$$

Hence, if we assume that δ_{ot}^{i} and δ_{rt}^{i} are distributed according to a normal (i.e., Gaussian) distribution, we can calibrate function parameters (i.e., mean and variance) so as to reproduce the observed data by solving the following minimisation problems:

$$\left[\widehat{\mu}_{ot}, \widehat{\sigma}_{ot}^{2}\right] = \underset{\mu_{ot},\sigma_{ot}^{2}}{\arg\min} \, Z_{ot}\left(\delta_{ot}^{i}, \mu_{ot}, \sigma_{ot}^{2}\right)$$
$$\left[\widehat{\mu}_{rt}, \widehat{\sigma}_{rt}^{2}\right] = \underset{\mu_{rt},\sigma_{rt}^{2}}{\arg\min} \, Z_{rt}\left(\delta_{rt}^{i}, \mu_{rt}, \sigma_{rt}^{2}\right) \quad (50)$$

with

$$\sigma_{ot}^{2} \geq 0,$$
$$\sigma_{rt}^{2} \geq 0, \quad (51)$$

where μ_{ot} and μ_{rt} are the means of the normal distributions in the case of the outward trip (ot) and return trip (rt); $\widehat{\mu}_{ot}$ and $\widehat{\mu}_{rt}$ are optimal values of μ_{ot} and μ_{rt}; σ_{ot}^{2} and σ_{rt}^{2} are the variances of the normal distributions in the case of the outward trip (ot) and return trip (rt); $\widehat{\sigma}_{ot}^{2}$ and $\widehat{\sigma}_{rt}^{2}$ are the optimal values of σ_{ot}^{2} and σ_{rt}^{2}; Z_{ot} is an objective function which expresses the distance between the cumulative distribution of observed values δ_{ot}^{i} and the cumulative distribution of the normal function of parameters μ_{ot} and σ_{ot}^{2}; Z_{rt} is an objective function which expresses the distance between the cumulative distribution of observed values δ_{rt}^{i} and the cumulative distribution of the normal function of parameters μ_{rt} and σ_{rt}^{2}.

The results of the calibration phases (i.e., solution of minimisation problems (50)) are shown in Table 2, and comparisons between the cumulative distribution of observed values

Table 2: Normal distribution parameters.

μ_{ot}	σ^2_{ot}	μ_{rt}	σ^2_{rt}
64.114	40.803	56.922	36.247

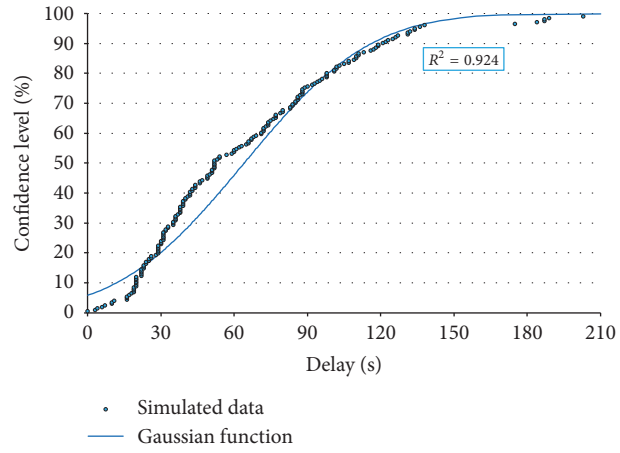


Figure 5: Comparison between cumulative distributions in the case of the outward trip.

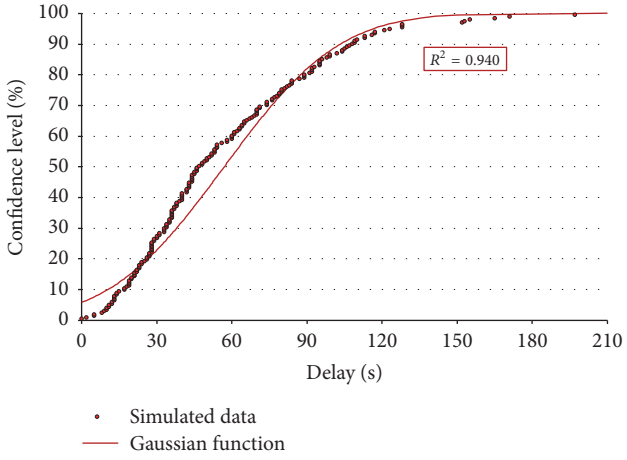


Figure 6: Comparison between cumulative distributions in the case of the return trip.

and that of corresponding normal functions are proposed in Figures 5 and 6.

These values were adopted to determine buffer times in the case of confidence levels equal to 90th, 95th, and 99th percentiles (values are shown in Table 1). Obviously, the adopted cycle times, that is, CT_{ord}, were calculated by assuming three different confidence levels associated with buffer time calculation, as shown in Table 1. In particular, the adoption of buffer times calculated by means of the proposed statistical approach allows any feasible fluctuation in operating service parameters to be handled.

The aim of the second test consisted in analysing and testing different operative schemes, and verifying their feasibility analytically. Indeed, since the planned headway H affects the value of the minimum headway of the line (see (38)), it is necessary to verify whether condition (37) is satisfied.

First, having fixed an operative scheme based on trains spending buffer and layover times at the first station of the subsequent trip, the lower bound of H_{min} (i.e., the minimum headway) may be identified as the maximum among constant terms, that is, ts^{inv}_{ot}, ts^{inv}_{rt}, and $ts_{min\text{-}ss}$. This preliminary analysis provided a lower bound value of 5.1 min due to the inversion time of the outward trip (i.e., at Garibaldi station), as shown in Table 1.

Hence, for any planned headway H greater than 5.1 minutes, it is possible to fix a corresponding number of convoys NC satisfying constraints (20) and then calculate turt in a closed form formulation by means of (23). Since the adoption of different split rates for layover times affects the value of H_{min} (as shown by (38)), we solved the optimisation problem (41) subject to (42) for identifying the minimum value of H_{min} to be compared with the planned headway H so as to verify the feasibility test (37).

The results of the proposed procedure are summarised in Tables 3–5, where the italic values indicate the unfeasible operating schemes (i.e., combinations of planned headway H and adopted number of convoys NC) which do not satisfy condition (37). However, tests were performed by considering different confidence levels in the buffer time estimation.

The results show that the increase in total buffer time (i.e., the sum of bt_{ot} and bt_{rt}) was 0.47 min from the 90th to 95th confidence level, 0.88 min from the 95th to 99th confidence level, and 1.34 min from the 90th to 99th confidence level.

Moreover, numerical results show that, for any operating scheme, if the sum of layover times was higher than buffer time increases, the increase in buffer time was offset by the reduction in layover time such that their sum (i.e., total reserve time according to UIC [40] or equivalently the sum of (21)) and the minimum headway was kept constant. Obviously, if the turt value was lower than the increase in buffer times, their sum could not be kept constant and the configuration would be unfeasible (see, e.g., headway values of 5.5 and 7.0 in the case of the 99th confidence level).

Since, in real cases, service frequencies may be different during a day (e.g., the headway during the peak hours may be lower than that during off-peak hours), the feasibility combination of planned headway H and adopted number of convoys NC has to be verified for any frequency configuration.

Finally, in order to show the utility of the proposed approach, we calculated reductions in energy consumption in the case of some operative schemes.

First of all, we assumed that the ES strategy was implemented by imposing a reduction in maximum speed during the outward and return trips. Indeed, this approach, which is described as ES strategy 2 in Section 2 (see Figure 3), may be easily implemented by affecting signalling system features (e.g., from the operative centre) without the awareness, knowledge, and/or preparation of drivers or changes in the ground-train communication system (i.e., information related to the beginning of the coasting phase or the new start-breaking point).

Table 3: Feasible configuration calculation in the case of the 90th percentile.

H [min]	NC_{min}	NC_{max}	NC	turt [min]	α_{min}	α_{max}	α_{opt}	H_{min} [min]	Test
5.5	**14**	**15**	**14**	**1.30**	**0.0%**	**100.0%**	**41.0%**	**5.12**	**OK**
5.5	*14*	*15*	*15*	*6.80*	*44.4%*	*52.5%*	*48.3%*	*6.87*	*NO*
6.0	**13**	**14**	**13**	**2.30**	**0.0%**	**100.0%**	**44.9%**	**5.12**	**OK**
6.0	*13*	*14*	*14*	*8.30*	*48.4%*	*49.0%*	*48.6%*	*7.62*	*NO*
7.0	**11**	**12**	**11**	**1.30**	**0.0%**	**100.0%**	**41.0%**	**5.12**	**OK**
7.0	*11*	*12*	*12*	*8.30*	*36.3%*	*61.0%*	*48.6%*	*7.62*	*NO*
8.0	**10**	**11**	**10**	**4.30**	**0.0%**	**100.0%**	**47.3%**	**5.62**	**OK**
8.0	*10*	*11*	*11*	*12.30*	*48.9%*	*49.3%*	*49.1%*	*9.62*	*NO*
9.0	**9**	**10**	**9**	**5.30**	**0.0%**	**100.0%**	**47.8%**	**6.12**	**OK**
9.0	*9*	*10*	*10*	*14.30*	*49.1%*	*49.4%*	*49.2%*	*10.62*	*NO*
10.0	**8**	**9**	**8**	**4.30**	**0.0%**	**100.0%**	**47.3%**	**5.62**	**OK**
10.0	*8*	*9*	*9*	*14.30*	*42.1%*	*56.4%*	*49.2%*	*10.62*	*NO*
12.0	**7**	**8**	**7**	**8.30**	**0.0%**	**100.0%**	**48.6%**	**7.62**	**OK**
12.0	*7*	*8*	*8*	*20.30*	*49.3%*	*49.6%*	*49.4%*	*13.62*	*NO*
14.0	**6**	**7**	**6**	**8.30**	**0.0%**	**100.0%**	**48.6%**	**7.62**	**OK**
14.0	*6*	*7*	*7*	*22.30*	*44.9%*	*54.1%*	*49.5%*	*14.62*	*NO*
15.0	**6**	**6**	**6**	**14.30**	**7.1%**	**91.4%**	**49.2%**	**10.62**	**OK**

Table 4: Feasible configuration calculation in the case of the 95th percentile.

H [min]	NC_{min}	NC_{max}	NC	turt [min]	α_{min}	α_{max}	α_{opt}	H_{min} [min]	Test
5.5	**14**	**15**	**14**	**0.83**	**0.0%**	**100.0%**	**34.0%**	**5.12**	**OK**
5.5	*14*	*15*	*15*	*6.33*	*43.7%*	*52.4%*	*47.9%*	*6.87*	*NO*
6.0	**13**	**14**	**13**	**1.83**	**0.0%**	**100.0%**	**42.7%**	**5.12**	**OK**
6.0	*13*	*14*	*14*	*7.83*	*48.1%*	*48.7%*	*48.3%*	*7.62*	*NO*
7.0	**11**	**12**	**11**	**0.83**	**0.0%**	**100.0%**	**34.0%**	**5.12**	**OK**
7.0	*11*	*12*	*12*	*7.83*	*35.3%*	*61.5%*	*48.3%*	*7.62*	*NO*
8.0	**10**	**11**	**10**	**3.83**	**0.0%**	**100.0%**	**46.5%**	**5.62**	**OK**
8.0	*10*	*11*	*11*	*11.83*	*48.7%*	*49.2%*	*48.9%*	*9.62*	*NO*
9.0	**9**	**10**	**9**	**4.83**	**0.0%**	**100.0%**	**47.2%**	**6.12**	**OK**
9.0	*9*	*10*	*10*	*13.83*	*48.9%*	*49.3%*	*49.0%*	*10.62*	*NO*
10.0	**8**	**9**	**8**	**3.83**	**0.0%**	**100.0%**	**46.5%**	**5.62**	**OK**
10.0	*8*	*9*	*9*	*13.83*	*41.7%*	*56.5%*	*49.0%*	*10.62*	*NO*
12.0	**7**	**8**	**7**	**7.83**	**0.0%**	**100.0%**	**48.3%**	**7.62**	**OK**
12.0	*7*	*8*	*8*	*19.83*	*49.2%*	*49.5%*	*49.3%*	*13.62*	*NO*
14.0	**6**	**7**	**6**	**7.83**	**0.0%**	**100.0%**	**48.3%**	**7.62**	**OK**
14.0	*6*	*7*	*7*	*21.83*	*44.7%*	*54.1%*	*49.4%*	*14.62*	*NO*
15.0	**6**	**6**	**6**	**13.83**	**5.5%**	**92.7%**	**49.0%**	**10.62**	**OK**

Our proposal was based on two different speed limits between the outward and the return trips, indicated, respectively, as v_{ot}^{LIM} and v_{rt}^{LIM}. Hence, for any feasible operative scheme (i.e., bold values in Tables 3–5) and for any assumed confidence levels (i.e., 90th, 95th, and 99th), we calculated the maximum reductions in energy consumption in the case of three values of parameter α (i.e., split rates):

(i) α_{min}, which represents the minimum layover time for the outward trip or, equivalently, the maximum layover time for the return trip (see (35))

(ii) α_{opt}, which represents the layover times providing the minimum headway (see (41))

(iii) α_{max}, which represents the maximum layover time for the outward trip or, equivalently, the minimum layover time for the return trip (see (35)).

Obviously, in the case of $\alpha_{min} = 0$, no ES strategy can be defined for the outward trip (i.e., it is not possible to define term v_{ot}^{LIM}). Likewise, in the case of $\alpha_{max} = 100\%$, it is not possible to perform an ES strategy for the return trip (i.e., it is not possible to define term v_{rt}^{LIM}).

Table 5: Feasible configuration calculation in the case of the 99th percentile.

H [min]	NC_{min}	NC_{max}	NC	turt [min]	α_{min}	α_{max}	α_{opt}	H_{min} [min]	Test
5.5	*15*	*15*	*15*	*5.45*	*42.2%*	*52.3%*	*47.1%*	*6.87*	*NO*
5.5	*15*	*15*	*16*	*10.95*	*71.2%*	*26.0%*	*48.6%*	*9.62*	*NO*
6.0	**13**	**14**	**13**	**0.95**	**0.0%**	**100.0%**	**33.3%**	**5.12**	**OK**
6.0	*13*	*14*	*14*	*6.95*	*47.5%*	*48.2%*	*47.7%*	*7.62*	*NO*
7.0	*12*	*12*	*12*	*6.95*	*33.1%*	*62.6%*	*47.7%*	*7.62*	*NO*
7.0	*12*	*12*	*13*	*13.95*	*66.7%*	*31.2%*	*48.9%*	*11.12*	*NO*
8.0	**10**	**11**	**10**	**2.95**	**0.0%**	**100.0%**	**44.6%**	**5.62**	**OK**
8.0	*10*	*11*	*11*	*10.95*	*48.4%*	*48.9%*	*48.6%*	*9.62*	*NO*
9.0	**9**	**10**	**9**	**3.95**	**0.0%**	**100.0%**	**46.0%**	**6.12**	**OK**
9.0	*9*	*10*	*10*	*12.95*	*48.6%*	*49.0%*	*48.8%*	*10.62*	*NO*
10.0	**8**	**9**	**8**	**2.95**	**0.0%**	**100.0%**	**44.6%**	**5.62**	**OK**
10.0	*8*	*9*	*9*	*12.95*	*40.9%*	*56.8%*	*48.8%*	*10.62*	*NO*
12.0	**7**	**8**	**7**	**6.95**	**0.0%**	**100.0%**	**47.7%**	**7.62**	**OK**
12.0	*7*	*8*	*8*	*18.95*	*49.1%*	*49.3%*	*49.2%*	*13.62*	*NO*
14.0	**6**	**7**	**6**	**6.95**	**0.0%**	**100.0%**	**47.7%**	**7.62**	**OK**
14.0	*6*	*7*	*7*	*20.95*	*44.4%*	*54.2%*	*49.2%*	*14.62*	*NO*
15.0	**6**	**6**	**6**	**12.95**	**2.3%**	**95.4%**	**48.8%**	**10.62**	**OK**

Since, for any α value, it is possible to determine layover times for the outward and the return trips, the speed limits were fixed by solving the following optimisation problem in the case of the outward trip:

$$v_{ot}^{LIM} = \arg\min_{v_{ot}^*} \Delta t_{ot}\left(v_{ot}^*\right) \tag{52}$$

subject to

$$\begin{gathered}\int_0^{t_{ot}^{TO}} v_{ot}^{TO}(\tau)\cdot d\tau = \int_0^{t_{ot}^{TO}+\Delta t_{ot}} v_{ot}^{ES}(\tau)\cdot d\tau \\ \Delta t_{ot} \le \alpha \cdot \text{turt} \\ v_{ot}^{ES} \le v_{ot}^{LIM}\end{gathered} \tag{53}$$

or equivalently in the case of the return trip

$$v_{rt}^{LIM} = \arg\min_{v_{rt}^*} \Delta t_{ot}\left(v_{rt}^*\right) \tag{54}$$

subject to

$$\begin{gathered}\int_0^{t_{rt}^{TO}} v_{rt}^{TO}(\tau)\cdot d\tau = \int_0^{t_{rt}^{TO}+\Delta t_{rt}} v_{rt}^{ES}(\tau)\cdot d\tau \\ \Delta t_{rt} \le (1-\alpha)\cdot \text{turt} \\ v_{rt}^{ES} \le v_{rt}^{LIM},\end{gathered} \tag{55}$$

where v_{ot}^* and v_{rt}^* are the generic speed limit in the case of the outward trip (ot) and return trip (rt); Δt_{ot} and Δt_{rt} are the increase in total travel time (i.e., the sum of travel times and dwell times) due to the implementation of the ES strategy in the case of outward trip (ot) and return trip (rt); t_{ot}^{TO} and t_{rt}^{TO} are the total travel time (i.e., the sum of travel times and dwell times) in the *Time Optimal* (TO) condition in the case of outward trip (ot) and return trip (rt); v_{ot}^{TO} and v_{rt}^{TO} are the travel speeds in the *Time Optimal* (TO) condition in the case of outward trip (ot) and return trip (rt); and v_{ot}^{ES} and v_{rt}^{ES} are the travel speeds in the case of ES strategy implementation in the case of outward trip (ot) and return trip (rt).

Once the new speed limits have been fixed, it is possible to calculate the new energy consumption by (1) and hence the energy consumption reductions as follows:

$$\begin{aligned}\Delta E_{ot} &= E_{ot}^{ES} - E_{ot}^{TO}, \\ \Delta E_{rt} &= E_{rt}^{ES} - E_{rt}^{TO}.\end{aligned} \tag{56}$$

Obviously, in the case of an undetermined speed limit (i.e., $\alpha_{min} = 0$ in the case of an outward trip and $\alpha_{max} = 100\%$ in the case of a return trip), reductions in energy consumption are null.

The reduction in energy consumption for each cycle may be calculated as follows:

$$\Delta E_{TOT} = \Delta E_{ot} + \Delta E_{rt}. \tag{57}$$

Finally, in order to show the order of magnitude of the amount of energy saved, we calculated the daily reduction in energy consumption by adopting the following assumptions:

(i) There are 17 hours (i.e., from 6.00 to 23.00) of metro service in a day.

(ii) The service frequency is constant throughout the day.

(iii) There is lack of interference among operating trains and trains moving from or to the depot.

(iv) The transition phase in the morning (TP_m) to reach the planned headway H, since initially all convoys are at the depot, and the transition phase in the

Table 6: Energy variations in the case of the 90th percentile.

H [min]	NC	α	v_{ot}^{LIM}	v_{rt}^{LIM}	ΔE_{ot} [kWh]	ΔE_{rt} [kWh]	ΔE_{TOT} [kWh]	ΔE_{Daily} [kWh]	Reduction in energy consumption
5.5	14	0.0%	–	57	–	32.7	32.7	5,487.9	4.91%
5.5	14	41.0%	65	61	20.5	20.3	40.8	6,852.5	6.13%
5.5	14	100.0%	58	–	39.8	–	39.8	6,688.6	5.98%
6.0	13	0.0%	–	51	–	44.2	44.2	6,888.6	6.64%
6.0	13	44.9%	61	57	34.0	32.7	66.7	10,402.0	10.02%
6.0	13	100.0%	53	–	56.1	–	56.1	8,751.5	8.43%
7.0	11	0.0%	–	57	–	32.7	32.7	4,311.9	4.91%
7.0	11	41.0%	65	61	20.5	20.3	40.8	5,384.1	6.13%
7.0	11	100.0%	58	–	39.8	–	39.8	5,255.4	5.98%
8.0	10	0.0%	–	44	–	44.1	44.1	5,290.4	6.63%
8.0	10	47.3%	54	52	50.2	38.9	89.1	10,687.9	13.39%
8.0	10	100.0%	45	–	67.5	–	67.5	8,104.5	10.15%
9.0	9	0.0%	–	42	–	57.8	57.8	6,246.2	8.69%
9.0	9	47.8%	51	49	59.3	49.1	108.4	11,709.0	16.30%
9.0	9	100.0%	65	–	20.5	–	20.5	2,215.4	3.08%
10.0	8	0.0%	–	44	–	44.1	44.1	4,232.3	6.63%
10.0	8	47.3%	54	52	50.2	38.9	89.1	8,550.3	13.39%
10.0	8	100.0%	45	–	67.5	–	67.5	6,483.6	10.15%
12.0	7	0.0%	–	36		68.0	68.0	5,233.1	10.22%
12.0	7	48.6%	45	44	67.5	44.1	111.6	8,595.1	16.78%
12.0	7	100.0%	37	–	95.2	–	95.2	7,328.9	14.31%
14.0	6	0.0%	–	36	–	68.0	68.0	4,485.5	10.22%
14.0	6	48.6%	45	44	67.5	44.1	111.6	7,367.2	16.78%
14.0	6	100.0%	37	–	95.2	–	95.2	6,281.9	14.31%
15.0	6	7.1%	61	30	34.0	90.1	124.1	7,447.5	18.66%
15.0	6	49.2%	39	38	80.3	70.7	150.9	9,056.9	22.69%
15.0	6	91.4%	31	57	108.0	32.7	140.7	8,439.3	21.14%

evening (TP_e) to lead all convoys to the depot may be calculated as follows:

$$TP_m = TP_e = (NC - 1) \cdot H. \quad (58)$$

It is worth noting that since in the considered line the depot is located next to Piscinola station (the suburb terminus) and is connected to the line by means of two different tracks (one for each direction), the introduction, variation, and exit of convoys do not provide any interference to the service.

Obviously, the daily energy saving could be easily calculated also by removing some of the above assumptions (such as the variation in service frequency during the day).

Tables 6–8 provide the energy variation for any feasible operating scheme and for the three confidence levels adopted in the computation of buffer times. The results showed that the greater the confidence level, the lower the energy saving since there is a reduction in total usable reserve time.

Moreover, since an increase in headway (frequency reduction) provides, on the one hand, an increase in energy saving for each convoy, but, on the other, a reduction in the number of convoys required for performing the service, it is not possible to provide a monotonic trend of the energy-saving function.

5. Conclusions and Research Prospects

In this paper, we proposed an analytical methodology for determining all operating parameters (including reserve times) in the case of metro systems. In particular, calculating the amount of reserve time is a fundamental step for implementing suitable *Energy-Saving* (ES) strategies. Indeed, as shown in the literature, reduction in energy consumption may be achieved by reducing performance of convoys (e.g., by reducing the maximum travel speed). Obviously, in order to preserve service quality, it is necessary to compensate the reduction in performance by consuming dead times (i.e., times spent by a train in a stop condition) such as buffer and layover times.

Generally, buffer times are adopted for recovery of delays, layover times for waiting for the subsequent trip. Our proposal consists in consuming layover times for energy-saving

Table 7: Energy variations in the case of the 95th percentile.

H [min]	NC	α	v_{ot}^{LIM}	v_{rt}^{LIM}	ΔE_{ot} [kWh]	ΔE_{rt} [kWh]	ΔE_{TOT} [kWh]	ΔE_{Daily} [kWh]	Reduction in energy consumption
5.5	14	0.0%	–	61	–	20.3	20.3	3,406.3	3.05%
5.5	14	34.0%	69	63	13.5	16.4	30.0	5,036.9	4.51%
5.5	14	100.0%	62	–	32.9	–	32.9	5,523.1	4.94%
6.0	13	0.0%	–	54	–	34.4	34.4	5,374.1	5.18%
6.0	13	42.7%	62	59	32.9	24.6	57.4	8,958.8	8.63%
6.0	13	100.0%	55	–	49.8	–	49.8	7,771.6	7.49%
7.0	11	0.0%	–	61	–	20.3	20.3	2,676.4	3.05%
7.0	11	34.0%	69	63	13.5	16.4	30.0	3,957.5	4.51%
7.0	11	100.0%	62	–	32.9	–	32.9	4,339.6	4.94%
8.0	10	0.0%	–	45	–	48.1	48.1	5,768.5	7.23%
8.0	10	46.5%	56	53	46.4	42.0	88.4	10,602.6	13.28%
8.0	10	100.0%	46	–	68.4		68.4	8,204.8	10.28%
9.0	9	0.0%	–	43	–	56.4	56.4	6,086.2	8.47%
9.0	9	47.2%	53	50	56.1	40.7	96.8	10,457.5	14.56%
9.0	9	100.0%	43	–	79.1	–	79.1	8,546.4	11.90%
10.0	8	0.0%	–	45	–	48.1	48.1	4,614.8	7.23%
10.0	8	46.5%	56	53	46.4	42.0	88.4	8,482.0	13.28%
10.0	8	100.0%	46	–.	68.4	–	68.4	6,563.8	10.28%
12.0	7	0.0%	–	37	–	67.6	67.6	5,202.5	10.16%
12.0	7	48.3%	47	45	69.2	48.1	117.2	9,026.7	17.62%
12.0	7	100.0%	38	–	96.4	–	96.4	7,421.4	14.49%
14.0	6	0.0%	–	37	–	67.6	67.6	4,459.3	10.16%
14.0	6	48.3%	47	45	69.2	48.1	117.2	7,737.2	17.62%
14.0	6	100.0%	38		96.4	–	96.4	6,361.2	14.49%
15.0	6	5.5%	62	30	32.9	90.1	123.0	7,379.2	18.49%
15.0	6	49.0%	39	38	80.3	70.7	150.9	9,056.9	22.69%
15.0	6	92.7%	31	59	108.0	24.6	132.5	7,952.5	19.92%

purposes, keeping buffer times intact in order to preserve the flexibility and robustness of the timetable in case of delays.

In order to verify the utility and feasibility of the proposed analytical approach, we applied it in the case of an Italian metro line in order to

(i) calculate buffer times as a function of the adopted confidence levels;

(ii) verify the feasibility of the operating scheme having fixed the planned headway H and the adopted number of convoys NC;

(iii) calculate the amount of energy consumption reduction for three different allocations of the total usable reserve time between the outward and return trips.

The first numerical applications showed that the higher the planned headway H, the lower the number of convoys required. Moreover, the higher the planned headway H, the higher the total usable reserve time (turt). Obviously, since the allocation of turt between the outward and return trips directly affects the minimum headway between two successive trains, the feasibility of any operating scheme may be verified by means of the proposed approach.

Moreover, having fixed a feasible operating scheme, an increase in confidence level provides an increase in buffer time. Hence, if the sum of layover times (i.e., turt) is higher than buffer time increases, the increase in buffer time is offset by the reduction in layover time such that their sum and the minimum headway are kept constant. Obviously, if the turt value were lower than the increase in buffer times, their sum could not be kept constant and the configuration would be unfeasible.

Finally, numerical tests for quantifying the reduction in energy consumption showed that an increase in headway provides an increase in the term turt and hence an increase in energy saving for the single convoy. Higher values of energy reduction were identified in the case of split rates (i.e., parameter α) providing the minimum headway where reductions are up to 22.69%. Hence, for future research, we propose to improve these performances further by optimising the allocation of reserve times between the outward and return trips in order to minimise energy consumption. Moreover,

Table 8: Energy variations in the case of the 99th percentile.

H [min]	NC	α	v_{ot}^{LIM}	v_{rt}^{LIM}	ΔE_{ot} [kWh]	ΔE_{rt} [kWh]	ΔE_{TOT} [kWh]	ΔE_{Daily} [kWh]	Reduction in energy consumption
6.0	13	0.0%	–	59	–	24.6	24.6	3,830.2	3.69%
6.0	13	33.3%	68	62	16.5	20.7	37.2	5,802.0	5.59%
6.0	13	100.0%	61	–	34.0	–	34.0	5,306.1	5.11%
8.0	10	0.0%	–	49	–	49.1	49.1	5,894.9	7.38%
8.0	10	44.6%	58	55	39.8	36.3	76.1	9,133.8	11.44%
8.0	10	100.0%	50	–	60.7	–	60.7	7,285.0	9.13%
9.0	9	0.0%	–	45	–	48.1	48.1	5,191.6	7.23%
9.0	9	46.0%	55	52	49.8	38.9	88.7	9,581.7	13.34%
9.0	9	100.0%	45	–	67.5	–	67.5	7,294.0	10.15%
10.0	8	0.0%	–	49	–	49.1	49.1	4,715.9	7.38%
10.0	8	47.7%	58	55	39.8	36.3	76.1	7,307.0	11.44%
10.0	8	100.0%	50	–	60.7	–	60.7	5,828.0	9.13%
12.0	7	0.0%	–	38	–	70.7	70.7	5,443.3	10.63%
12.0	7	47.7%	48	46	65.5	49.9	115.4	8,887.8	17.35%
12.0	7	100.0%	39	–	80.3	–	80.3	6,179.7	12.06%
14.0	6	0.0%	–	38	–	70.7	70.7	4,665.7	10.63%
14.0	6	47.7%	48	46	65.5	49.9	115.4	7,618.1	17.35%
14.0	6	100.0%	39	–	80.3	–	80.3	5,296.9	12.06%
15.0	6	2.3%	69	31	13.5	78.1	91.6	5,497.0	13.77%
15.0	6	48.8%	40	39	81.2	54.6	135.8	8,146.4	20.41%
15.0	6	95.4%	32	63	108.5	16.4	125.0	7,499.8	18.79%

stochastic simulations may be adopted to verify benefits in more realistic cases. Finally, we propose a reduction in the confidence level adopted in the buffer time calculation in order to increase the amount of available reserve time.

Obviously, in the case of rail systems (or metro systems with more complex layouts), it is necessary to integrate the proposed approach with timetable optimisation techniques.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] European Commission, White Paper. Roadmap to a Single European Transport Area - Towards a competitive and resource efficient transport System, COM (2011), Brussels, Belgium, 2011.

[2] European Commission: Climate Action – Reducing emissions from transport. 2018, http://ec.europa.eu/clima/policies/transport/index_en.htm.

[3] S. Chien and Z. Yang, "Optimal feeder bus routes on irregular street networks," *Journal of Advanced Transportation*, vol. 34, no. 2, pp. 213–248, 2000.

[4] R. Garcia and A. Marin, "Network equilibrium with combined modes: models and solution algorithms," *Transportation Research Part B: Methodological*, vol. 39, no. 3, pp. 223–254, 2005.

[5] L. Deng, W. Gao, W. Zhou, and T. 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.

[6] L. Dell'Olio, A. Ibeas, and P. Cecin, "The quality of service desired by public transport users," *Transport Policy*, vol. 18, no. 1, pp. 217–227, 2011.

[7] J. De Oña and R. De Oña, "Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches," *Transportation Science*, vol. 49, no. 3, pp. 605–622, 2015.

[8] R. F. Abenoza, O. Cats, and Y. O. Susilo, "Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution," *Transportation Research Part A: Policy and Practice*, vol. 95, pp. 64–84, 2017.

[9] P. Pereira, R. P. Ribeiro, and J. Gama, "Failure prediction – an application in the railway industry," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 8777, pp. 264–275, 2014.

[10] A. Consilvio, A. Di Febbraro, and N. Sacco, "Stochastic scheduling approach for predictive risk-based railway maintenance," in *Proceedings of the 2016 IEEE International Conference on Intelligent Rail Transportation, ICIRT 2016*, pp. 197–203, Birmingham, UK, August 2016.

[11] L. D'Acierno, A. Placido, M. Botte, and B. Montella, "A methodological approach for managing rail disruptions with different perspectives," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 10, pp. 80–86, 2016.

[12] 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.

[13] F. Corman and L. Meng, "A review of online dynamic models and algorithms for railway traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1274–1284, 2015.

[14] Y. Gao, L. Kroon, M. Schmidt, and L. Yang, "Rescheduling a metro line in an over-crowded situation after disruptions," *Transportation Research Part B: Methodological*, vol. 93, pp. 425–449, 2016.

[15] S. Zhan, J. Zhao, and Q. Peng, "Real-time train rescheduling on high-speed railway under partial segment blockages," *Tiedao Xuebao/Journal of the China Railway Society*, vol. 38, no. 10, pp. 1–13, 2016.

[16] H. Strobel, P. Horn, and M. Kosemund, "Contribution to optimum computeraided control of train operation," in *Proceedings of the 2nd IFAC/IFIP/IFORS symposium on traffic control and transportation systems*, Monte Carlo, Monaco, 1974.

[17] P. Howlett, "The optimal control of a train," *Annals of Operations Research*, vol. 98, pp. 65–87 (2001), 2000.

[18] E. Khmelnitsky, "On an optimal control problem of train operation," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 45, no. 7, pp. 1257–1266, 2000.

[19] C. Sicre, A. P. Cucala, A. Fernández, and P. Lukaszewicz, "Modeling and optimizing energy-efficient manual driving on high-speed lines," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 7, no. 6, pp. 633–640, 2012.

[20] L. Yang, T. Liden, and P. Leander, "Achieving energy-efficiency and on-time performance with Driver Advisory Systems," in *Proceedings of the 2013 IEEE International Conference on Intelligent Rail Transportation, IEEE ICIRT 2013*, pp. 13–18, China, September 2013.

[21] T. Albrecht, C. Gassel, A. Binder, and J. Van Luipen, "Dealing with operational constraints in energy efficient driving," in *Proceedings of the IET Conference on Railway Traction Systems, RTS 2010*, April 2010.

[22] Y. V. Bocharnikov, A. M. Tobias, C. Roberts, S. Hillmansen, and C. J. Goodman, "Optimal driving strategy for traction energy saving on DC suburban railways," *IET Electric Power Applications*, vol. 1, no. 5, pp. 675–682, 2007.

[23] M. Miyatake and H. Ko, "Optimization of train speed profile for minimum energy consumption," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 5, no. 3, pp. 263–269, 2010.

[24] R. Chevrier, P. Pellegrini, and J. Rodriguez, "Energy saving in railway timetabling: a bi-objective evolutionary approach for computing alternative running times," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 20–41, 2013.

[25] T. Albrecht, "Energy-efficient railway operation," in *Railway Timetabling & Operations, Eurailpress*, A. Hansen and J. Pachl, Eds., Eurailpress, Hamburg, Germany, 2014.

[26] G. Corapi, V. De Martinis, A. Placido, and G. De Luca, "Impacts of energy saving strategies (ESSs) on rail services and related effects on travel demand," in *Proceedings of the 14th International Conference on Railway Engineering Design and Optimization, COMPRAIL 2014*, pp. 709–720, June 2014.

[27] A. D'Ariano and T. Albrecht, "Running time re-optimization during real-time timetable perturbations," in *Proceedings of the 10th International Conference on Computer System Design and Operation in the Railway and Other Transit Systems, COMPRAIL 2006, CR06*, pp. 531–540, July 2006.

[28] F. Mehta, C. Rößiger, and M. Montigel, "Latent energy savings due to the innovative use of advisory speeds to avoid occupation conflicts," in *Proceedings of the 12th International Conference on Computer System Design and Operation in the Railways and other Transit Systems, COMPRAIL 2010*, pp. 99–108, China, September 2010.

[29] M. Gallo, F. Simonelli, and G. De Luca, "The potential of energy-efficient driving profiles on railway consumption: a parametric approach," in *Transport Infrastructure and Systems*, G. Dell'Acqua and F. Wegman, Eds., pp. 819–824, Taylor & Francis Group, 2017.

[30] V. De Martinis and M. Gallo, "Models and methods to optimise train speed profiles with and without energy recovery systems: a suburban test case," *Procedia: Social and Behavioral Sciences*, vol. 87, pp. 222–233, 2013.

[31] V. De Martinis, U. Weidmann, and M. Gallo, "Towards a simulation-based framework for evaluating energy-efficient solutions in train operation," in *Proceedings of the 14th International Conference on Railway Engineering Design and Optimization, COMPRAIL 2014*, pp. 721–732, June 2014.

[32] N. Zhao, C. Roberts, S. Hillmansen, and G. Nicholson, "A Multiple Train Trajectory Optimization to Minimize Energy Consumption and Delay," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2363–2372, 2015.

[33] M. Ercolani, A. Placido, L. D'Acierno, and B. Montella, "The use of microsimulation models for the planning and management of metro systems," *WIT Transactions on the Built Environment*, vol. 135, pp. 509–521, 2014.

[34] S. Su, X. Li, T. Tang, and Z. Gao, "A subway train timetable optimization approach based on energy-efficient operation strategy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 883–893, 2013.

[35] X. Li and H. K. Lo, "Energy minimization in dynamic train scheduling and control for metro rail operations," *Transportation Research Part B: Methodological*, vol. 70, pp. 269–284, 2014.

[36] P. Lukaszewicz, "Energy saving driving methods for freight trains," *Advances in Transport*, vol. 15, pp. 901–909, 2004.

[37] M. Gallo, F. Simonelli, G. De Luca, and V. De Martinis, "Estimating the effects of energy-efficient driving profiles on railway consumption," in *Proceedings of the 15th International Conference on Environment and Electrical Engineering (EEEIC '15)*, pp. 813–818, IEEE, Rome, Italy, June 2015.

[38] T. Montrone, P. Pellegrini, P. Nobili, and G. Longo, "Energy consumption minimization in railway planning," in *Proceedings of the 16th International Conference on Environment and Electrical Engineering, EEEIC 2016*, Florence, Italy, June 2016.

[39] F. Simonelli, M. Gallo, and V. Marzano, "Kinematic formulation of energy-efficient train speed profiles," in *Proceedings of the AEIT International Annual Conference: A Sustainable Development in the Mediterranean Area, AEIT 2015*, Naples, Italy, October 2015.

[40] International Union of Railways (UIC), *UIC CODE 451-1 Timetable recovery margins to guarantee timekeeping – Recovery margins*. UIC, Paris, France, 2000.

[41] M. Cepeda, R. Cominetti, and M. Florian, "A frequency-based assignment model for congested transit networks with strict capacity constraints: characterization and computation of equilibria," *Transportation Research Part B: Methodological*, vol. 40, no. 6, pp. 437–459, 2006.

[42] Q. Fu, R. Liu, and S. Hess, "A review on transit assignment modelling approaches to congested networks: a new perspective,"

Procedia: Social and Behavioral Sciences, vol. 54, pp. 1145–1155, 2012.

[43] A. Nash and D. Huerlimann, "Railroad simulation using OpenTrack," *WIT Transactions on the Built Environment*, vol. 74, pp. 45–54, 2004.

[44] L. D'Acierno, M. Gallo, B. Montella, and A. Placido, "Analysis of the interaction between travel demand and rail capacity constraints," *WIT Transactions on the Built Environment*, vol. 128, pp. 197–207, 2012.

[45] L. D'Acierno, M. Botte, A. Placido, C. Caropreso, and B. Montella, "Methodology for Determining Dwell Times Consistent with Passenger Flows in the Case of Metro Services," *Urban Rail Transit*, vol. 3, no. 2, pp. 73–89, 2017.

[46] L. D'Acierno, A. Placido, M. Botte, M. Gallo, and B. Montella, "Defining robust recovery solutions for preserving service quality during rail/metro systems failure," *International Journal of Supply and Operations Management*, vol. 3, no. 3, pp. 1351–1372, 2016.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 8502819, 17 pages
https://doi.org/10.1155/2018/8502819

Research Article

The Planners' Perspective on Train Timetable Errors in Sweden

Carl-William Palmqvist,[1] **Nils O. E. Olsson**,[1,2] **and Lena Winslott Hiselius**[1]

[1]*Faculty of Engineering, Lund University, P.O. Box 118, 221 00 Lund, Sweden*
[2]*Norwegian University of Science and Technology, Høgskoleringen 1, 7491 Trondheim, Norway*

Correspondence should be addressed to Carl-William Palmqvist; carl-william.palmqvist@tft.lth.se

Received 20 November 2017; Revised 5 February 2018; Accepted 8 March 2018; Published 12 April 2018

Academic Editor: Lingyun Meng

Timetables are important for train punctuality. However, relatively little attention has been paid to the people who plan the timetables: the research has instead been more centred on how to improve timetables through simulation, optimisation, and data analysis techniques. In this study, we present an overview of the state of practice and the state of the art in timetable planning by studying the research literature and railway management documents from several European countries. We have also conducted interviews with timetable planners in Southern Sweden, focusing on how timetable planning relates to punctuality problems. An important backdrop for this is a large project currently underway at the Swedish Transport Administration, modernizing the timetable planning tools and processes. This study is intended to help establish a baseline for the future evaluation of this modernization by documenting the current process and issues, as well as some of the research that has influenced the development and specifications of the new tools and processes. Based on the interviews, we found that errors in timetables commonly lead to infeasible timetables, which necessitate intervention by traffic control, and to delays occurring, increasing, and spreading. We found that the timetable planners struggle to create a timetable and that they have neither the time nor the tools required to ensure that the timetable maintains a high quality and level of robustness. The errors we identified are (a) crossing train paths at stations, (b) wrong track allocation of trains at stations, especially for long trains, (c) insufficient dwell and meet times at stations, and (d) insufficient headways leading to delays spreading. We have identified eleven reasons for these errors and found three themes among these reasons: (1) "missing tools and support," (2) "role conflict," and (3) "single-loop learning." As the new tools and processes are rolled out, the situation is expected to improve with regard to the first of these themes. The second theme of role conflict occurs when planners must strive to meet the demands of the train operating companies, while they must also be unbiased and create a timetable that has a high overall quality. While this role conflict will remain in the future, the new tools can perhaps help address the third theme by elevating the planners from first- to double-loop learning and thereby allowing them to focus on quality control and on finding better rules and heuristics. Over time, this will lead to improved timetable robustness and train punctuality.

1. Introduction

Railways are an important part of the transport system. In Sweden, trains traveled 153 million km during 2015, which is an increase of 9% over five years [1]. Passenger traffic has increased by 16% over the same period, and in 2015 passenger trains made up 83% of all trains. While freight traffic in 2015 was at the lowest level since 1990, the freight tonnage transported by rail has risen slightly as the loads have increased. The capacity is most heavily utilised around the three major metropolitan areas of Stockholm, Gothenburg, and Malmö-Lund. A quarter of the metropolitan lines is utilised at levels associated with high sensitivity to delays, low average speeds, and little time for infrastructure maintenance [2]. On the rest of the network, only about five percent of the segments are utilised to the same extent. When measured during peak loads, these figures are higher across the board.

The punctuality of trains in Sweden has been close to 90% for the last several years [1], with punctuality measured as a maximum delay of five minutes at the final stop. This is considered too low by the industry, which has set a goal that it should be 95% by 2020. This ambition, to increase punctuality, is the background for our research and for this paper. Many factors influence punctuality, such as

weather [3–5], congestion [6], other operational factors [7, 8], and infrastructure [9]. Previous studies also indicate that properties of the timetable can have a large impact on delays and punctuality, that delays often occur at station stops, and that dwell times are systematically underestimated [10–13]. Thus, there is reason to believe that errors in the timetable may affect punctuality.

The interaction between infrastructure, capacity, and timetable planning is found on a strategic, tactical, and operational level. The strategic level is typically long term, over several years, and can be related to new infrastructure or new timetable structures, while the tactical level is related to producing a timetable implemented in a shorter perspective, typically one year. This paper mainly studies tactical timetable planning. Operational timetabling is related to making short-term changes to a timetable, often a few weeks or days in advance.

Timetable planners prepare timetables and other documentations related to planned changes for passenger and freight trains. Planners are often faced with the challenges of working with complex timetabling [14]. In addition to the complexity of the planning itself, they must be able to deal with different stakeholders in the railway sector and have conflict resolution skills. A final timetable must satisfy the needs of travelers and public and private parties, while maintaining the fairness, transparency, reliability, and safety of the railway system. The train service specifications are passed to the timetable planners, who produce the timetables. However, these specifications can be in violation of guidelines, or there can be conflicting needs of different train operators. Watson [14] found that the complexity of the timetabling and capacity allocation processes can hinder effectiveness, highlighting the conflicting nature of objectives for timetable planning, especially in the privatized railway.

While extensive research has been carried out on the human-machine interface in train traffic control in Sweden ([15–17]; see also [18] for dispatchers in the US) and some work has been done on the integration of timetable planning and traffic control [19], relatively little research has been done focusing on timetable planners and their tools. Watson [14, 20] covered timetable planning in the UK, which has many similarities to the Swedish context, and ONTIME [21] contains some expert judgment on the state of practice in Sweden. National interest in this topic has increased, as a large project is currently in progress at the Swedish Transport Administration, seeking to modernize the interface between train operating companies and the Transport Administration by developing new tools and routines for timetable planning. These tools will, among other things, support more flexible and optimal planning, improve capacity and punctuality, shorten planning lead times, and improve transparency and the handling of engineering works [22]. Since 2011, the situation with the deregulated passenger train market in Sweden is also new and rare in an international context, with the new divisions of responsibility resulting both in new role conflicts and in more collaborative decision-making between stakeholders. To learn more about this, it is useful to talk to the timetable planners.

As the timetable planning process and methods in Sweden are about to undergo significant change over the next few years, this paper (1) presents the state of practice in order to establish a baseline and (2) outlines the state of the art in research, which has inspired and influenced the development of the new tools and methods. It also (3) gives a description of the timetable planners current situation in Sweden and (4) identifies common errors in Swedish train timetables, which influence the punctuality, as well as the reasons behind them. Thereby, the paper helps to enable future studies looking to evaluate the effects and effectiveness of the new tools being developed and implemented for train timetable planning in Sweden and elsewhere.

2. Background

Timetabling has largely been studied from a technical and optimisation perspective (see, e.g., [24]). However, timetabling can also be studied from an organizational perspective, using other methods. For instance, Avelino et al. [25] study the politics of timetabling, comparing the Dutch and Swiss experiences and illustrating that "timetable planning is not merely an operational process to be left to engineers or economists" (pp. 20). Even though the Swiss have many more train operating companies than the Dutch, their federal government still takes a much more active and strategic role in ensuring an optimised travel time over the whole network. Watson [14] found that the privatization of British Rail had a negative effect on the timetabling processes. The problems were a result of poor planning and rushed implementation of new organization of the British railway sector. Since then, both the British and Swedish railways have gained experiences from the new structures with a division between infrastructure and train operation. However, the inherent characteristics of the divided structure remain.

The infrastructure manager supplies capacity on the railway, while train operating companies represent the demand for transport. Timetable planners are squeezed in between these needs, which sometimes conflict [26]. Watson [14] discusses timetable planning as a process by which a "demand" for rail transport (passenger and freight) is connected to the "supply side" constraints (especially available infrastructure capacity) in order to produce timetables that meet the demand. Through train planning, railway administrators seek to meet the needs of customers while utilising available resources as well as possible. Efficient and effective train planning is the key to getting the best possible performance on a railway network.

Timetabling is governed by several restrictions, such as safety requirements and organizational policies. A routine or heuristic approach can be applied to timetabling. Routines can be defined as "a repetitive, recognizable pattern of interdependent actions, involving multiple actors" [27, pp. 96]. Heuristics [28] are cognitive rules of thumb, or shortcuts, that people apply, consciously or unconsciously [29]. Argyris and Schön [30] present learning as understanding and eliminating the gap between the expected result and the actual result of an action. This gap can be eliminated either by making changes (taking corrective measures) within the

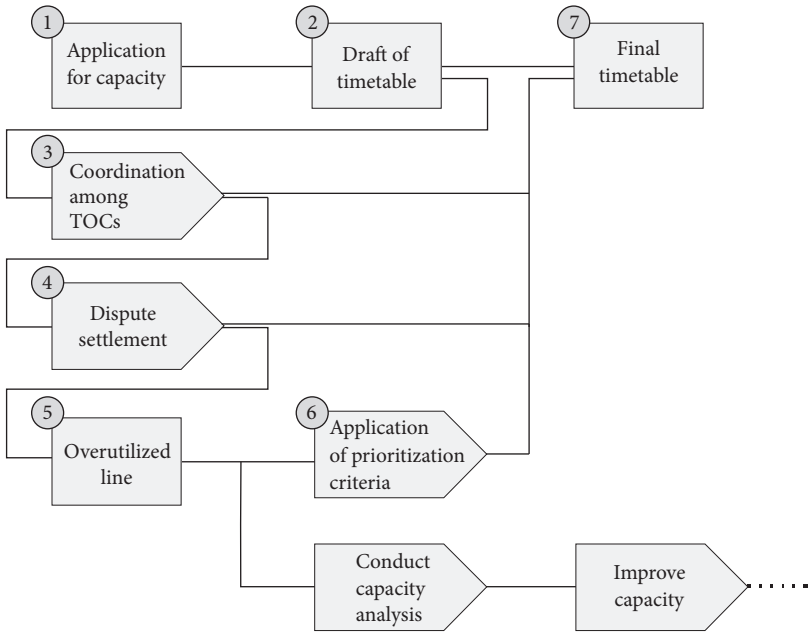


Figure 1: Capacity allocation process in Sweden. Adapted from the Swedish Transport Administration [23].

existing values and norms, or by changing the existing values and norms. The former is called single-loop learning and the latter is called double-loop learning. Single-loop learning is connected to doing things right, in accordance with the existing values and norms. Double-loop learning is about doing the right things, by questioning the existing values and norms. This is a concept we will return to throughout the paper.

According to Loock and Hinnen [31], organizational heuristics are the result of collective learning processes. They found that successful organizations refine their heuristics over time, as a result of feedback loops. Organizational heuristics can also interact with individual heuristics and with improvisation in the decision-making process [32]. Kirkebøen [33] shows that heuristics can be biased, such as a bias to rely on the information that is most available or a bias to search for information that confirms rather than contradicts a decision.

Managerial issues in railway planning have been the topic of a range of publications, notably Vuchic [34] and Profillidis [35]. Managerial aspects of railway planning include strategic, tactical, and operational issues, and timetabling is an important factor in all these perspectives. A strategic perspective on railway planning includes selecting major investments and positioning the railway within the overall transport system, as described by Harris et al. [36]. In a tactical perspective, timetabling is one step in the planning process. Ceder [37] describes the train scheduling process in four steps: network route design, setting timetables, scheduling vehicles, and assignment of crew. All of these planning steps have been of interest to researchers, especially from an optimisation perspective [38]. Operational management issues in railway planning have also been studied: Veiseth et al. [39] study how timetable improvements in a Total Quality Management perspective for operational improvements, and Samà et al. [40] provide an example of applying optimisation to support operational management. Roth et al. [18] and Tschirner [41] also studied aspects of operational management for train dispatching in traffic control centres, while Watson [14, 20] considered the contrasting needs and preferences of timetable planners and their managers.

2.1. State of Practice in Sweden. The following is an outline of the current state of practice in Swedish train timetabling, to give the reader a better understanding of the processes and tools involved. Rail traffic in Sweden has been gradually deregulated since competition for some tracks for passenger trains was allowed 1990, with open access with competition on all tracks since 2011, while freight services have been competing on the tracks since 1996 [42].

2.1.1. The Capacity Allocation Process. The annual capacity allocation process in Sweden is illustrated briefly in Figure 1, which is reconstructed from the Network Statement published by the Swedish Transport Administration. This corresponds to the tactical level of planning described above. The process as described here is based on the Network Statement published by the Swedish Transport Administration [23] and an excellent account in Hellström [19]. First the train operating companies send in requests for the capacity they want during the next year. The timetable planners at the Transport Administration combine these requests and, by following their rules and guidelines for how to plan timetables, come up with a draft that contains all the train timetables for one year. In case there are conflicting requests, such that not all trains can be run when and where the train operating companies desire, there is first a step where the parties are encouraged to coordinate among themselves. If

this coordination is unsuccessful, there is a process where the Transport Administration, together with the parties, tries to settle the dispute. If these attempts are also unsuccessful, parts of the infrastructure can be declared to be saturated, and the planners at the Transport Administration use prioritization criteria to determine which trains have priority, sometimes entirely rejecting the other requests. In yet another step, the train operating companies can appeal the decision of the planners to an administrative court, if they are unsatisfied with the planners' decisions.

The point of departure compared to other European countries is when the coordination process breaks down. The United Kingdom [43] has a more qualitative process with an overarching objective "to share capacity on the Network for the safe carriage of passengers and goods in the most efficient and economical manner in the overall interest of current and prospective users and providers of railway services." Along with this objective it has a list of twelve criteria, on which to assess the fulfilment of the objective. In Germany [44], priority is given to regular-interval or integrated network services, cross-border trains, and train paths for freight trains. If none of these criteria are sufficient, priority is given to the train paying the higher track charge. In the Netherlands [45], the track charges for the conflicted train path are raised if an agreement cannot be found by the parties, to the point that only one of the actors remains. The Dutch also emphasize the basic hourly patterns in their capacity allocation process, striving for a cyclic timetable, something that is not found in the other countries' process descriptions.

How trains are prioritized in the timetabling is regulated in the Network Statement [23]. In the request for capacity, the train operating companies must classify their trains according to previously set criteria. For passenger trains, the expected number of passengers, the share of passengers who are time-sensitive, the share of regional travelers, and demands for high speeds are the basis for this categorisation. Similar criteria are used for freight trains. There are 18 categories for passenger trains, shown in Table 1, and 15 for freight, which are not shown here. Trains can also have associations with other trains: these are categorised into five categories each for passenger and freight, based on the number of passengers or tonnage of freight, and three categories for the turnaround of vehicles. Each prioritization category is linked to a series of social cost estimates, which are based on the methodology in ASEK [46] and are used to find a total solution, which minimises the social welfare costs. Track charges to be paid by the train operating companies to the Swedish Transport Administration are determined using other methods and are not in any way involved in the prioritization of trains.

Sometimes train path requests, or associations, are not possible to fulfil, and they are instead excluded from the timetable. While there is no clear guidance from welfare economic theory on how to evaluate these exclusions [47], a cost is still assigned so that large numbers of trains are not simply excluded. The cost estimate has instead been calculated roughly in the following way: for each type of train (each prioritization criteria) an estimate has been made based on how much the travel time can be extended before it is no longer considered viable to operate the train at all. For commuter and regional trains, this has been set to 15%, so that if the travel time of a train had to be extended by more than 15%, it is assumed that no one would like to use that train (there is no documentation or discussion on how these limits have been determined). The cost of excluding a train is then set to be equal to the cost of such a train being run to that maximum limit, which results in a very high cost of excluded train paths, severely punishing solutions that deny requests for capacity. A minor detail in the calculation of exclusion costs is that a template is applied to estimate a reasonable level of margins for the trains, instead of using the actual size of margins in the train path. This is done so that companies do not act strategically in reducing margins in order to have their competitors' trains be excluded instead of their own.

2.1.2. Robust Timetables. Robustness in timetables is primarily achieved by modifying time supplements, headways, and dwell times (see, i.e., [48]). In Sweden, time supplements are added in several ways. The first way is included in the runtime calculation, which is calculated as if the train had a technical top speed 3% lower than it does. Practically speaking, this adds a uniformly distributed margin of 3%. This is often motivated by differences in train driver behaviour and is described as the primary source of margins in timetables but in fact only makes up a small fraction of the total margins [49].

The other explicit way to add time supplements is by assigning node supplements [50]. On each major railway line, between two and four stations have been designated as nodes, while minor lines instead use the first and last stations. The idea is then to add a number of minutes as time supplements for trains traveling between two nodes. For passenger trains, those that have maximum speeds above 180 km/h need four additional minutes between each pair of nodes, and those that have lower maximum speeds instead need three minutes for each pair of nodes. Trains that travel shorter distances on the major lines, not reaching a node, require two minutes. The airport express trains are given an exception, only needing node supplements of one minute. Freight trains require two minutes for each set of nodes, or one minute for shorter distances.

In addition to these two methods, timetable planners use their discretion in assigning time supplements. One common practice is to add seconds, so that arrival times occur at whole minutes. For instance, if a train would arrive at 12:44:27, the planner might add 33 seconds, so that the arrival instead occurs at 12:45:00. Over long journeys, this often adds up to considerable supplements. Supplements are also sometimes given for trains that are scheduled to stop at a platform which is not on the main track, because this takes slightly longer to get to, and because engineering works are being done on the track, requiring lower speeds for part of the journey. These are meant to correct cases where the runtime calculation is known to be wrong; however they are not really margins increasing the robustness.

Headways, the time separation between trains going in the same direction on the same track, are another important way to provide robustness. A short separation implies a

Table 1: Prioritization criteria for passenger trains and their associations, with social cost estimates. Reproduced from Appendix 4 B in the Swedish Network Statement for 2017 [23]. When alternative timetable solutions are compared to one another, the solution with the lowest total social cost estimate should be chosen. This entails first calculating the social cost for each train, using the estimates above value travel time, travel distance, phasing time (shifting departures from the times requested by the train operating companies), and associations between trains for both passengers and vehicles and then summing up the social cost across all trains in the timetable scenario. The cost estimates for rejected requests, trains that are not allowed to run in the timetable alternative, are estimated using a slightly different methodology outlined briefly in Section 2.1.1. Similarly, there is a threshold at which point an association is no longer considered viable, and instead of basing the cost on the time used, a fixed penalty is applied, as indicated in the column SEK/Assn. The train operating companies are required to (accurately, to the best of their knowledge) report the prioritization category of each of their trains along with the request for capacity, based on the identification criteria, and the computation of the social cost estimate is then done in Trainplan for each timetable scenario. Guidelines are also provided for how to handle new train concepts, where there is no previous knowledge of the number and type of passengers.

Description of category	Identification criteria for prioritization categories				Social cost estimate		
	Number of passengers	Share of time sensitive passengers	Share of regional passengers	Demand for high speed, few stops	SEK/min transport time	SEK/km transport distance	SEK/min phasing time
Heavy commuter	≥700	≥75%	≥75%		1,150	93	784
Regio-commuter	≥300	≥75%	≥75%		736	93	474
Regio-commuter	≥300	≥75%	≥75%		736	93	474
Regio-max	≥200	≥75%	≥75%		499	76	212
Regio-max	≥75	≥75%		High	499	76	212
Regio-standard	≥75	≥75%	≥75%		240	27	132
Regio-standard	≥25	≥25%		High	240	27	132
Regio-low	≥25	≥75%	≥75%		170	29	96
Regio-low	≥75		≥25%		170	29	96
Regio-low	≥25		≥25%		170	29	96
Regio-mini	≥0		≥25%		46	22	10
IC-express	≥200	≥75%		High	753	64	429
IC-standard	≥75	≥25%			484	41	291
IC-low	≥25	≥25%			253	38	125
IC-low	≥75				253	38	125
IC-mini	≥0				80	15	31
IC-mini	≥0				80	15	31
Unspecified					35	11	8
Association	Passengers				SEK/min	SEK/Assn.	
Conn. pas. max	≥125				647	55,300	
Conn. pas. high	≥75				304	26,000	
Conn. pas. std.	≥50				190	16,300	
Conn. pas. low	≥20				107	9,110	
Conn. pas. mini	≥0				30	2,600	
Turnaround high					0	37,300	
Turnaround std.					0	19,300	
Turnaround low					0	11,800	

high capacity and throughput but also increases the risk that delays spread from train to train. In Sweden, minimum headways are regulated in a document [51], varying from two to seven minutes, depending on the location. The normal range, applying to most of the network, is from three to five minutes. It is unclear from the documentation to what extent this is a technical minimum and what has been added to improve robustness, but if headways are higher than required by the regulations, robustness would be expected to improve.

The third key factor for robustness of timetables is dwell time. If a dwell time is too short, the departure of the train will be delayed. Correspondingly, if the scheduled dwell time is longer than required for the exchange of passengers or goods, the excess time can be used to make up for any previous delays. The guidelines [52] state that dwell times for passenger trains should, in general, be two minutes long. Sometimes longer times are required and other times, if the number of passengers is small and the train and station are prepared for

Table 2: Timetable planning standards for passenger trains in Sweden.

Robustness measure	Norm in the Swedish regulations
Running time supplement	+3% across the board, included in run time calculation
Node supplement	3-4 minutes per pair of nodes passed; 2 minutes if partial; 2–4 nodes exist per railway line
Dwell time at stations	2 minutes standard; 1 minute if the number of passengers is small
Minimum headway	2–7 minutes, most commonly 3–5 minutes

a speedier boarding process, one minute can be used instead. If the number of passengers is very small, it is possible to schedule a stop without dwell time, merely slowing the train down to a stop and then starting again immediately, but if this is done the runtime on the next line section should be extended, and if passenger numbers increase, the timetable should be redone and longer dwell times set.

The norms discussed above are summarised in Table 2. Vromans [53] compiled information on running time supplements used in the Netherlands, the United Kingdom, and Switzerland, and as required by the UIC which can be repeated here as a point of reference. While this information may appear dated, the timetabling norms have not changed significantly in Sweden since at least the 1990s, and there is reason to believe that the norms would be stable in other countries as well, even though the actual planning practice may well develop over time. According to Vromans [53], the Dutch use a running time supplement of approximately 7% across the board, for passenger trains. In the United Kingdom, runtime calculations are based on previous performance rather than physics-based methods, and the running time supplements are not explicitly defined [54]. The Swiss [55] use a supplement of 7% for passenger trains, on top of which they add one minute for every 30 minutes of runtime and additional supplements in some locations, usually at highly utilised nodes. As a final point of comparison, the UIC [56] recommends a combination of time and distance based supplements: between three and seven percentage points are added to the running time, to which one should add between one and one and a half minutes for every 100 km. Unfortunately, we have not found any norms or standards with regard to dwell times or headway times in other countries.

The complexity of traffic, changing conditions, and sheer number of decisions makes it difficult for planners to foresee the punctuality effects of individual decisions on the size and distribution of margins, headways, and dwell times. Because of this difficulty, there has not been a convergence around best practice in timetabling in Sweden, and there has not been a steady improvement in punctuality [1], which one might expect if planners were able to learn what works and what does not. Nonetheless, there has been a significant drift from the norms in how the timetables are scheduled in practice [49], as the running time supplements are routinely much larger in practice than in the norms, while the dwell times are significantly shorter.

2.1.3. Tools. The Swedish Transport Administration currently uses the tool Trainplan to create timetables, a tool which is also used in the UK railways [14] and is reviewed by Hammerton [57]. RailSys is increasingly being used to perform limited test runs of parts of the timetable using stochastic simulation, on a more detailed infrastructure model, and can address many of the issues presented in this paper. As in the UK, the group of trained users is small and used only as a complement to the main planning process. These software packages, their use, and their limits are discussed in depth by Watson [14]. One of the most important constraints is that Trainplan does not provide automatic conflict detection, meaning that the planners must check for these manually. As each planner plans for thousands or tens of thousands of trains per year, this is a recurring issue. New tools and processes, intended to improve the quality and efficiency of the timetable process roughly along the lines discussed in ONTIME [21], are under development at the Swedish Transport Administration and are to be rolled out gradually from 2018 through 2020.

With the introduction of information technology and increasingly powerful computers, simulation is gaining an increasingly important role in the railways, both in practice at infrastructure managers and in academia. In the early 2000s researchers at the Royal Institute of Technology began to model the Swedish railways in the simulation software RailSys [58]. After several years, the Swedish Transport Administration began to use this model to perform simulations in different aspects of its practice. This is now an established part of the process, and parts of the annual timetable are run through simulations in several iterations before they are finalised. Larger engineering works are also simulated regularly in order to estimate the effects on train traffic and to make appropriate changes to the timetable. This is being developed further, so that alternate plans for engineering works can be tested against each other [59].

However, the team of capacity analysts who use RailSys is still quite small and they are organized as a group of experts, which is involved in many projects besides the annual timetable planning. The timetable planning is still done in Trainplan, and the test runs in RailSys are limited in both number and scope. Thus, while some of the errors produced in the planning process are identified and corrected, the workflow of the planners is not really affected and many of the errors still go unnoticed.

2.2. State of the Art in Sweden. As a way to utilise the infrastructure capacity more effectively, the Swedish Transport Administration is funding research for simulation and optimisation tools for timetabling and rescheduling. Some of this work is outlined here, to give an idea of the research and development underway in Sweden, without intending to

give a comprehensive review. This work creates an important and interesting backdrop to the ongoing development and impending implementation of new tools and processes for timetable planning.

2.2.1. The Capacity Allocation Process. The capacity allocation process itself, the framework for timetable planning, is also being developed and improved spearheaded by researchers at the Swedish Institute for Computer Science. A summary and roadmap toward implementation is presented in Aronsson et al. [60]. The background for this is twofold. One issue is that, with deregulation and the existence of multiple, competing train operating companies, the demands put upon the capacity allocation process are fundamentally different than before. Kreuger et al. [61] list some typical requirements from train operating companies and their customers, and based on these they developed a mathematical framework for detecting and resolving conflicts. The other issue is that the current allocation process has very long lead times, is inflexible, and leads to an inefficient capacity utilisation of the infrastructure. This is described by Forsgren et al. [62], who found several mathematical opportunities and challenges following an alternative process, where the decision-making is postponed as far as possible into the future.

Gestrelius et al. [63] developed an outline for a more efficient capacity allocation process. In the current process, train paths are made and used for the entire year. Once finalised, they are not allowed to be modified; they must instead be cancelled and replaced by a new train path in the ad hoc-process, which is not allowed to disturb the surrounding train paths. This often leads to, for instance, a train stopping at a certain station every day of the week, to await a meeting train that only runs on Tuesdays. All other days of the week, these scheduled stops are entirely unnecessary. The suggestion for the new process is to only lock down certain key characteristics of train paths, such as departure times from some important stations, retaining greater flexibility in later planning, and operational stages without compromising the quality of the train paths.

Gestrelius et al. [63] also present a method to generate these key characteristics based on an annual timetable, using rolling horizon planning and a mixed integer programming model, which optimises the train paths for each individual day, using different delivery commitments for different operators. They then applied the model on a case study in Sweden and showed that this allows for a more efficient utilisation of the infrastructure. Aronsson et al. [64] continued this work by working to estimate the value of this uncovered capacity, showing that a large portion of the available capacity is hidden when using the current planning methods and scheduling rules.

Working further on improving the process, Svedberg et al. [65] developed a model to optimise the welfare cost of different timetable variants, containing competing train operators, and they applied the model to a part of the Swedish rail network. A model like this gives the infrastructure manager a more correct and transparent way to rule on which trains should get their requested timetables and which should be adjusted or excluded. This is increasingly important as the number of requested train paths is increasing, and especially as the requests are coming from competing actors. The welfare cost derived from this framework can also be used to find the optimal number of departures, find the best departure times, and estimate the economic value of different timetable variants.

2.2.2. Robust Timetables. One line of research is directed at creating timetables that are robust against minor disturbances. An important prerequisite is developing methods on how to measure and quantify this robustness, ideally before the timetables are put into operation. Andersson et al. [66] show that there is a clear mismatch between where margins are placed and where delays occur. They also suggest that punctuality should be measured at all stops, not only the final destination. Peterson [67] studied two Swedish train services, finding that dwell times are usually underestimated and not sufficiently compensated by margins on the line and that the precision of the train paths decreases linearly with the running time. Building on this finding, Khoshniyat and Peterson [68] modified timetables so that the assigned minimum time slot in the train path is increased linearly with the service's travel time. Based on numerical experiments on a double track segment of the Swedish Southern Mainline, they conclude that the modified timetables perform better when small disturbances are introduced.

Several new ex ante robustness measures have been proposed. Gestrelius et al. [69] propose that the number of possible alternative meeting locations between two trains is a robustness measure, as it gives flexibility for rescheduling. Khoshniyat [70] developed a headway-based method, which can improve robustness without imposing major changes in timetables, and proposes four new measures: Channel Width, Channel Width Forward, Channel Width Behind, and Track Switching. Andersson et al. [71] also propose a new robustness measure: Robustness in Critical Points (RCP), which is focused on points in the timetable that are particularly sensitive to delays. Warg and Bohlin [72] established a timetable performance index to evaluate the benefits and robustness of a timetable from a passenger perspective, by combining simulation and socioeconomic analyses.

Once robustness measures have been proposed, it is possible to optimise timetables around them. For instance, Andersson et al. [73] propose a model which reallocates existing margin time to increase the RCP. This is tested by simulation runs of an initial timetable and one with reassigned margins, while introducing small delays. In the adjusted case, total delays at the end station are 28% lower. Solinen et al. [74] also use an optimisation model to increase the RCP throughout a timetable and evaluate the results with simulation. They found that robustness increased locally, but that the relationship between ex post measures and RCP must be studied further.

2.2.3. Tools. One prototype tool based on optimisation, which has been developed and evaluated, is called the Maraca [75]. It is used for nonperiodic timetabling and minimises resource conflicts, thus enabling the user to focus on the

strategic decisions. Based on a trial of this prototype, Forsgren et al. [76] show how computer and optimisation based tools can provide valuable insights, even before full-scale implementation. Another tool, intended for marshalling yards, is designed by Gestrelius et al. [77] who design an integer programming model to schedule shunting tasks and to allocate tracks in both the arrival yard and the classification bowl. This tool has a planning period of four days and optimises characteristics like shunting work effort, the number or cost of tracks, and the shunting task start times.

Similarly, the rescheduling of trains in disturbed scenarios is amenable to methods and tools using optimisation. Krasemann [78] developed an algorithm, which effectively delivers good solutions within the permitted time, performing a depth-first search using an evaluation function to prioritize when conflicts arise, and then branches according to a set of criteria. Krasemann [79] then shows that the approach is feasible for practical problems, using the case of the Iron Ore Line in Northern Sweden and solving many different delay scenarios to optimality within one minute or less.

3. Methodology

To produce the material for this paper, we carried out semistructured interviews with timetable planners working at the Swedish Transportation Administration's office in Malmö. Each interview was approximately an hour long, recorded, and transcribed in full, which resulted in a written material of around 50 pages. The results were analysed by manual categorisation and concentration of meaning.

The Swedish Transport Administration employs about 20 long term timetable planners, who work chiefly in the annual timetabling process. In addition to these, there are short-term timetable planners who work in the ad hoc-process. The Swedish railway is divided into eight regions, and the southernmost region is planned from the office in Malmö by four timetable planners, all of whom we interviewed. Two of the planners were men and two women. All of them have worked in the industry for many years, at least since 2003 and going back as far as 1985, and with timetables for nine or more years.

This region is a sort of microcosm of the railway network in Sweden, and it contains a very diverse mix of railway lines, train traffic, and capacity utilisation. It includes the Southern Mainline, one of the most heavily used in the country, dense commuter systems around Malmö-Lund-Helsingborg, heavy freight traffic mixed with passenger trains on the single-track Scania Line, sparser passenger traffic around Ystad, Karlskrona, and Kalmar, and several nonelectrified lines with local, manual train dispatching and very low traffic volumes. Thus, while we have only interviewed planners in one region, those planners have been exposed to a wide variety of planning conditions and circumstances.

We used a qualitative method because this allowed us to effectively study the values and priorities of those involved. In this choice of method, we thus applied a qualitative approach on a topic that is typically studied using quantitative methods. We prepared an interview guide based on four areas which were identified before the interviews: (1) guidelines and support, (2) rules of thumb, (3) feedback loops, and (4) trade-offs, with a handful of guiding questions in each area.

The process of analysing the transcribed material was carried out in sequence. The first step was to sort the different interviewer-interviewee exchanges by area, rather than by chronology, what Kvale [80] and Hammersley and Atkinson [81] call categorisation. The second step was to concentrate the meaning of the answers [80] by cutting superfluous words and sometimes reformulating entire paragraphs into a few sentences. This was a necessary process, to make it feasible to get an overview of what was said, and the volume of text was reduced from 24,500 words to only 4,500. All these steps were performed manually.

The following has been translated from Swedish and provides an example of the concentration of meaning:

> The transcript
>
> "Unfortunately, the time is short, so we don't have time to quality control ourselves, rather it's like: now I've done that train I hope it's right. We have two occasions where Traffic Control go in and check, but they can't see everything either. So unfortunately, we can't do the kind of quality of work that we would like because the resources aren't enough, we have to focus on getting it done."
>
> can be condensed to
>
> "We focus on getting it done, and don't have time for quality control. The Traffic Control try to check, but can't see everything."

Having concentrated the interview answers, we made a more detailed sorting, and 16 new subareas were identified. Following this, we further summarised the contents, reducing the volume from 4,500 to 500 words. This made it manageable to get an overview of the contents. Section 4 contains translations of these summaries, grouped into the four areas of the interview guide. An example of this second step of summarising

> "Traffic Control gives daily feedback: insufficient meeting time, crossing train paths, stopped freight train before a slop, train stops on the wrong track."
>
> "Feedback: want the train on another track, crossing train paths, infeasible timetable. Adjust in ad hoc-timetable and in dialogue with train operating company, but not always possible."
>
> "Traffic control usually tells us: trains are too long, or always late. Less feedback about punctuality, but there can be problems around engineering works or ad hoc-trains that only run a single day and make a mess. We mostly focus on the train numbers that run more often."
>
> "A new group is looking at the code 'suspected error in the timetable', finding new errors: shouldn't have a meeting with zero dwell time in Mörrum [a small station], because then the trains lose two minutes."

> "Easy to miss crossing train paths, because our systems lack conflict management. You learn after a while how to handle different locations, but it's not written down anywhere."
>
> can be summarised into
>
> "Traffic control has a lot of feedback. Short meeting times, crossing train paths, bad track allocation, trains that are too long, and freight trains stopped in slopes."

The analysis in Section 5 is based on an alternate reading of the interview responses. We identified instances in the interviews where the planners described feedback from traffic control about errors in the timetable, which can be seen in the examples of translated and concentrated statements provided above, and we focused on the ones that were mentioned most frequently among the planners. We also identified several statements that could explain why errors sometimes occur in the timetable and condensed these into a list of eleven reasons. At this point, we looked for different ways to group and categorise the answers, looking for themes on a higher analytical level, and came up with the three following categories, which we use in the analysis: (1) "missing tools and support," (2) "role conflict," and (3) "single-loop learning" (see Argyris and Schön [30] and Section 2 above). These three categories are used to explain and discuss the reasons behind the errors more deeply.

4. Results

The questions in the interview guide were structured around four areas, and we will begin by reporting summaries of the responses according to each area. These summaries add to the contextual understanding of the timetable planning in Sweden and document issues in ways that could not be achieved by studying documents or guidelines and are a key part in establishing a baseline for later evaluations and studies, following the implementation of the new tools and processes. Further analysis and discussion of interview responses follows in Sections 5 and 6.

4.1. Learning and Feedback Loops. The planners explained that the timetabling work is split by lines, with some much-needed reinforcement at large stations. They state that they have a large individual responsibility, both in learning what is necessary and in performing quality control, that it takes a long time to learn the details of each railway line, and that there is no time for quality control. It is difficult to transfer accumulated experience. Even though there is some education and transfer of knowledge as a line is handed over, it is insufficient and quite short. The planning method has been largely the same for the last 20 years or so, but the work is continuously getting more difficult, due to the increasing number of trains and engineering works. Problems increasingly occur at the stations, where the capacity is insufficient. In the past, the reverse used to be true.

There is no systematic evaluation or quality control: it is up to the individual planners, and they do not have time to perform it. The feedback loop that exists is from operational traffic control, most frequently about insufficient meeting times, crossing train paths, poor track allocation, trains that are too long for the allocated tracks, and freight trains being stopped at inclines. However, there is no established system or routine to keep track of these comments, or to make use of them, other than making a note mentally or on paper and trying to remember this information until it can be used next year. The planners also feel that important preconditions change from year to year, which makes it difficult to draw comparisons and lessons between years, making it even more difficult to make use of the loose notes that they make based on the feedback.

4.2. Support and Guidelines. The planners describe a somewhat contradictory experience of the work environment: the women describe the atmosphere as open and helpful, while the male planners describe a lonelier experience. The more experienced planners work more based on discussions with the train operating companies than on a strict application of the guidelines. The one who was newest at the job and who had more often switched between railway lines stressed the importance of studying the geography and signalling systems extensively. The guideline document [52] was mentioned by all the interviewees, but it is interpreted liberally and was not described as helpful. A new version is said to be coming soon, but this has been said for several years. Trainplan is the main tool and while it does many things, the version used at the Swedish Transport Administration does not handle track allocation, manage conflicts, or provide topographical information. The train operating companies often agree among themselves and apply detailed timetables. The planner only adjusts when necessary, and this is done in dialogue with the train operating companies. As a planner, one must know which signal-box models are present and how they work, which is quite complicated without sufficient technical support.

4.3. Trade-Offs. Although the timetable planners know that trains are sometimes late, they report that they cannot plan a timetable based on the trains running late. Engineering works that move along the line during the year are difficult to schedule properly, as the location for the delay shifts over time. Creating new train paths for each scenario is challenging and discouraged by the capacity allocation process. The lack of capacity, especially at stations, is mentioned by several interviewees as the most difficult problem in their work. It appears to be a bottleneck-problem, rather than one of sheer volume, where past a certain threshold it becomes very difficult. The planners try to handle the lack of capacity in dialogue with the train operating companies, and they all describe different principles for doing so. The problem of congestion is exacerbated by the short-term planners bending the rules to fit in more trains. Negative margins are often used for local trains on single-track, so that the scheduled runtime is shorter than the calculated minimum runtime, at the request of train operating companies. The explanations for this vary from person to person, but the planners state that "it has always been like this."

4.4. Rules of Thumb. The planners express that node supplements are the primary way of assigning margins, but everyone describes a different methodology for assigning them. Some give descriptions that seem to run counter to the few rules that are written down. Another very common practice is to adjust arrival and departure times at control points so that they occur at whole minutes. At some places, this is described as required for technical reasons, but it appears popular even elsewhere. Usually, seconds are added up to the whole minute but sometimes subtracted. Phasing supplements are an important tool to make the timetables feasible. Only the most experienced planner adds a minute after a scheduled stop, as per an old unwritten rule going back decades, although others were aware of the practice. The train operating companies generally set the dwell times. The planners say that the standard is two minutes, but they give the impression that shorter times dominate. Local trains are often given the same arrival and departure times, with no scheduled dwell time, to avoid waiting unnecessarily in case the train is delayed or the number of passengers is small. Dwell times in excess of two minutes are primary for connections and phasing reasons and in rare cases for trains going into the mountains during holidays, when many people bring skis and similar equipment.

5. Analysis

Here we present the results of an alternate reading of the responses, focusing on the errors that occur in timetable planning, the reasons behind them, and three themes running through these reasons.

The Swedish timetable planners described receiving daily feedback from traffic control; see the excerpts in Section 3, which are primarily centred on four areas: (a) crossing train paths at stations, (b) wrong track allocation of trains at stations, especially long trains, (c) insufficient dwell and meet times at stations, and (d) insufficient headways leading to delays spreading. To give a rough idea of the relative frequency of these errors, throughout the transcripts the planners explicitly mention receiving feedback relating to (a) five times, (b) nine times, (c) six times, and (d) twice. The issue of crossing train paths is mentioned frequently as a difficult issue: 15 times throughout the transcripts, suggesting that the planners are working hard on finding and eliminating such errors, with partial success.

Based on the interviews, we have also identified eleven reasons why the timetables sometimes lack quality, allowing the occurrence of errors. These reasons were not explicitly stated or described as such but were identified by reading and sorting through the transcripts. We have also identified three common themes that run through the list: "missing tools and support," "role conflict," and "single- rather than double-loop learning" (trying to follow the established norms rather than trying to establish the right norms, see Argyris and Schön [30]). This is all summarised in Table 3.

The leftmost column in Table 3 contains an identifying number, used in the following sections. The second column from the left describes the reason for lacking quality, and the centremost column identifies which of the four common errors (a)–(d) is associated with this reason. These four errors are discussed further in the table text. The three rightmost columns illustrate how the eleven reasons map onto the three themes we have identified. Each of these themes are discussed in the following sections.

5.1. Missing Tools and Support. One theme running through the responses is that the proper tools to perform timetable planning are missing. This is most clearly illustrated by reasons (2)–(5) in Table 3. Based on the answers given, such tools would free up time and allow a shift in focus from the details to quality assurance and a better overview of the timetable. The planners state that they must keep track of which model of signalling control is present at each location and how it works. This is complicated and the planners described that there is very little support available.

The main tool used by the planners in Sweden also lacks functions for track planning and conflict management and does not provide topographical information. These functions were intended to be part of the current software tool, when it was procured in the early 2000s, but the implementation of these modules was cancelled, allegedly because the quality of infrastructure data was too poor. New tools for both planning and control of traffic are under development, and the planners hope that these important functions will be implemented in the coming years.

The guiding document for timetable planning was mentioned by all four planners but was not very helpful and is interpreted liberally. A new version is said to be arriving soon, but this has been said for several years. The planners also state that the problem of insufficient capacity is worsened by operational timetable planners being less rigorous in following the guiding documents, when inserting new trains into the gaps that do exist.

5.2. Role Conflict. Another underlying theme is the inherent role conflict of timetable planners, which is best illustrated by reasons (6)–(9) in Table 3; an overly liberal interpretation of the planning rules and guidelines is that train operating companies ask for shortcuts to fit more trains into the timetable, that train operating companies request short dwell times to avoid trains waiting, and that there is no clear strategy for the location and size of time supplements.

One example of this conflict is how the decision to deny a train request, because of capacity constraints, is described as difficult and controversial. Denied requests can be, and often are, challenged through the formal capacity allocation process described in Section 2.1.1, which leads to a lengthy and difficult legal process of showing that everything was done correctly and transparently. Producing a timetable, which cannot in practice be executed as scheduled, and which is likely to cause delays, is not subject to the same formal procedures or complaints.

Another example is how the more experienced planners focus more on discussions aimed at reaching a consensus with the train operating companies than on a strict application of the guidelines. In Sweden, the train operating companies mostly agree between themselves and apply with detailed timetables, which the planners only adjust when necessary

Table 3: Reasons why errors occur in timetables. The four types of associated errors are (a) crossing train paths at stations, (b) wrong track allocation of trains at stations, especially long trains, (c) insufficient dwell and meet times at stations, and (d) insufficient headways leading to delays spreading. (a)-(b) make the timetable infeasible without intervention from traffic control, are therefore considered the most critical, and can be described as inadvertent mistakes. (c)-(d) systematically lead to delays occurring, increasing, and spreading and are made intentionally to accommodate the train operating companies, even if the consequences are not fully understood. Note the high extent to which these errors are focused around stations.

Number	Reason for lacking quality, description	Associated errors	Theme 1. Missing tools and support	Theme 2. Role conflict	Theme 3. Single- rather than double-loop learning
(1)	Insufficient time for quality assurance of timetables	(a), (b), (c), and (d)			X
(2)	Too many issues to keep in mind manually for planners	(a), (b)	X		X
(3)	Work is becoming more difficult due to increasing congestion	(a), (b)	X		
(4)	Congestion on stations, especially large and complex stations	(a), (b), (c), and (d)	X		
(5)	Missing tools for track allocation and conflict management	(a), (b)	X		
(6)	Liberal interpretation of the planning rules and guidelines	(c), (d)		X	
(7)	Train operating companies ask for shortcuts to fit more trains into the timetable	(c), (d)		X	
(8)	Train operating companies want short dwell times to avoid trains waiting	(c)		X	
(9)	No clear strategy for the location and size of time supplements	(c), (d)		X	X
(10)	Poor feedback and evaluation of timetables; no routine for continuous improvement	(a), (b), (c), and (d)			X
(11)	Poor knowledge transfer to new planners; poor documentation	(a), (b), (c), and (d)			X

and then in dialogue with the companies, doing what they can to squeeze the trains in. For instance, insufficient dwell times and negative margins are often given to local trains on single tracks, on request from the train operating companies. The rationale for this differs from planner to planner, but "it has always been done like this."

5.3. Single-Loop Learning. The last theme is best illustrated by reasons (1)-(2) and (9)–(11) in Table 3 and refers to the concept described in Argyris and Schön [30] and Section 2. The timetable planners have a large amount of individual responsibility: in the number of timetables that they must create, in learning the relevant signalling control systems, in applying the rules and guidelines, in assigning margins, in creating robustness, and in controlling the quality. They show that it takes a long time to learn the geography and that there is no time for quality control. The task gets harder and harder because the number of trains and engineering works are increasing.

There is no systematic evaluation. This is up to each planner, and they say that they do not have the time. Important preconditions change from year to year, which makes it difficult to compare between different timetables and to transfer the notes made about feedback and errors from one year to the next one. This contributes to the conditions described in ONTIME [21, p. 37] which also commented on the timetabling process in Sweden: "the accumulated know-how of train dispatchers and train drivers is not fed back to the timetable construction process to any larger extent." It is also difficult to transfer accumulated experience: even though there is some education and transfer of knowledge to new planners, it is described as insufficient and too short.

The planning work in Sweden is centred on finishing the timetables in time, while keeping in mind all the technical details and constraints of different signalling control systems and rolling stock, the topography, crossing train paths, track allocation, and so on. This is a direct parallel to the situation in the UK described by Watson [14, p. 112], where "achieve all timetable production timescales" is listed as the number one priority among timetable planners at Network Rail and "error free" only as number six, in a ranked list of eight priorities. The British timetable planners that Watson [20, p. 312] interviewed primarily requested *"help with elimination or reduction of the repetitive data manipulation tasks that*

delay them from tackling the 'interesting' conflict resolution and resource minimisation work."

The lack of support and proper tools means that there simply is not enough time or energy left to assess whether the rules and guidelines are the best or most appropriate ones to use. There is not enough time to consider what would make the timetable better, or to ensure that the errors from previous years are not repeated. If the implementation of new tools can help to streamline the workflow for timetable planners, the planners could use more of their time considering how to make the timetable better and more robust.

6. Discussion

This paper is focused on how the timetable process and the decisions of timetable planners contribute to delays and nonpunctuality. Timetable properties have been shown to be important and that changes to the timetable are relatively easy, quick, and inexpensive to make, in comparison to changes to infrastructure, rolling stock, and maintenance practices. While we show that timetable planners make errors that lead to delays, we do not suggest that this is the most important contributor to delays overall.

It is also important to distinguish between what timetable planners do, which contributes to nonpunctuality, and what they do to contribute to punctuality. In the interviews, it is made clear that planners are really struggling to create a feasible, error-free timetable and that they need more assistance. Before this is addressed, there is no time left in the process for them to increase robustness. There is clearly a need for tools to help generate and prove feasible timetables more quickly. Once the planners have these tools they can focus on improving the robustness.

The importance of new tools directed at helping generate timetables more quickly and verifying that they are feasible and error-free is also evident from the interviews. As these new tools and processes are being developed and implemented, it is interesting to consider and discuss the role of timetable planners in the future. One viewpoint that is often raised is that the planners will go from drawing timetables to monitoring the systems that draw them. The focus will shift from creating *a* timetable, to creating a *better* timetable. However, the current system already assumes that this is the case. The results of this paper suggest that planners are already entrusted with a large degree of discretion and are solely responsible for creating good, robust timetables. The problem is that they currently do not have the means to meet that responsibility, because the task of creating *a* timetable is too demanding and time-consuming. If provided with better tools and support and if the feedback loops were improved, it would be possible for the planners to rise to the challenge in the future. As is, they do not appear to be anxious or worried that they will lose their jobs to automation: the impression is perhaps more of frustration that the tools do not work as they should.

Conflicting train paths, track allocation, and constraints due to different signalling control systems could all be handled well through software, but the tools currently used do not do this. As new tools and routines are implemented, it is important to ensure that these important functions are included and that enough high quality data is provided for the systems. While the role conflict cannot be entirely avoided by technology, new tools and processes can diminish the consequences by no longer permitting things such as negative margins and very short dwell times. Thus, more egregious errors could, perhaps, be eliminated. The planners would then also have more support when denying some requests from the train operating companies, but this support could be given in other, less technical ways. Adjusting the penalty associated with excluding trains, which is severe and lacking any theoretical basis, would be one such measure that could easily be implemented on a policy level.

It would also be helpful to clarify the roles and responsibilities between the different parties. With a separation between the infrastructure manager and the train operating companies, it is conceivable that both parties would like to assume responsibility for ensuring that the timetable has a high quality. The train operating companies might want to be responsible, because timetables are core part of their business with significant impacts on their customers' experience. The case is also strong for the infrastructure manager to assume responsibility, because it needs to coordinate traffic from many different companies, as well as engineering works. However, the results of this paper indicate that neither party assumes the responsibility. Clarifying the roles by clearly stating which party is responsible for the quality and robustness of the timetable would help make this interaction more constructive.

The results presented in this paper suggest that both researchers and practitioners should focus more on identifying and improving the relevant feedback loops, to achieve a higher level of learning among those involved. Single-loop learning is both a technical and organizational issue. Since the tools are lacking, planners are hard-pressed merely to finish their work. The time is simply not sufficient to perform quality control. Because there is no systematic review of the quality and outcome, there is no way to begin to improve the rules and guidelines, or to create a better timetable. This supports the findings in Watson [14], Hellström [19], and ONTIME [21]. As the tools do not provide enough assistance, the focus is, and must be, on creating a timetable before creating a better timetable. Creating a better timetable is what we imagine the timetable planners of the future will be tasked with doing, when more of the work has been automated and the software tools provide far more assistance. Rather than trying to manually execute all the details, they will choose which heuristics, goal-functions, and constraints to apply in different scenarios to achieve the best overall results.

This study focuses on Sweden. In the literature, we have seen large similarities with the United Kingdom, which uses the same tools and has a similarly deregulated market. The new tools that are currently being implemented in Sweden have recently been implemented, by the same supplier, in both Norway and Denmark. Experts from the infrastructure manager and largest train operator in the Netherlands [82] describe similar issues with errors causing delays and infeasibility there. We believe that the planning process is largely similar across most European countries, although

the level of deregulation and competition between train operating companies for capacity may vary, as will the tools and contexts.

Reflecting on the methodology, we found it very rewarding to interview the people involved in the actual work of planning in a structured way, and we found them to be surprisingly frank and forthcoming. We are also pleased with how much more information can be extracted from the interviews when the transcripts are concentrated, categorised, and sorted. It is a very time-consuming process, but in our experience, it makes the initial investment of conducting and transcribing the interviews even more worthwhile. Being able to query the transcribed material from different angles, rather than being bound by the initial structure provided by the interview guide, is also very valuable and one of the key methodological takeaways for us.

As a final note, the findings in this study have been reported to both managers and experts at the Swedish Transport Administration on several occasions. They have expressed a keen interest in the research, in its results, and in spreading the findings deeper into the organization and the planning process. The problems with the current set of software are well known internally, which is one of the reasons for the large and currently ongoing project of replacing it. More surprise has been expressed on the theme of role conflict, in the perceived lack of internal support for timetable planners to go against the wishes of the train operating companies, for the benefit of the overall timetable. The managers realise that this is a question of leadership, where they can and must improve. The issues around learning and feedback also generated extensive interest and discussion on the role of timetable planners in the future once the software provides better support and on how to more systematically implement methods and routines for continuous learning and improvement.

7. Conclusions

The process of timetable planning in Sweden has been mostly stable since the beginning of the new millennium, with only minor changes in the processes, norms, and tools used. Over the next few years this will change, following a very significant investment in developing new tools and routines. The large shift that this is brought about presents a rare opportunity to study a major transition in timetable planning. Any large investment like this should be evaluated seriously, to assess the effects and draw lessons for the future. This paper helps establish the baseline for such an evaluation. Finally, a growing research interest into robust timetables, for more punctual train traffic, also justifies studying the timetable planners and process.

Studying the Network Statements of European railways, we found that the processes are largely similar; the difference is mainly in how to prioritize when two or more train operating companies have requested the same train slot. Whereas Sweden uses social cost estimates based on the theory of welfare economics, the British use a more qualitative assessment based on eleven predefined criteria; Germany selects the train that would pay the highest track charges, while the Dutch successively raise the track charges for the conflicting trains until only one party is willing to run the train.

Based on interviews with timetable planners in Sweden, we found that errors in timetables commonly lead to infeasible timetables, which necessitate intervention by traffic control, and to delays occurring, increasing, and spreading. The errors we identified are (a) crossing train paths at stations, (b) wrong track allocation of trains at stations, especially for long trains, (c) insufficient dwell and meet times at stations, and (d) insufficient headways leading to delays spreading. The situation is very reminiscent of the one described by Watson [14] in the UK, preceding our study by almost a decade: the timetable planners really struggle to create a timetable to begin with, and they do not manage to produce one without errors.

Reading through the transcripts of the interviews, we have identified eleven reasons for these errors, and running through these reasons, we have identified and discussed three themes: (1) missing tools and support, (2) role conflict, and (3) single-loop learning. The first theme, that proper tools and support are missing, is mostly a technical issue. The second theme is that of a role conflict for planners, which is mostly an organizational issue. On the one hand, they must strive to meet the demands of the train operating companies and, on the other hand, they must be unbiased and create a timetable that has a very high quality overall. The third theme is that planners, both individually and as a collective, appear to be stuck in single-loop learning [30], which is both a technical and organizational issue.

Appendix

Interview Guide

Note. The questions are translated from Swedish. Questions in italic were optional, intended to follow up.

Introductory Questions

Can you please tell me briefly about yourself and your background?

Can you please tell me briefly about how you work with the upcoming annual timetable?

Questions on Feedback

How long have you worked here?

Can you please tell me a little about what you have learned since you started working here?

Can you please talk briefly about how the work has changes since then?

Approximately how is the work divided between those of you who create the annual timetable?

Can you please give some examples of how you exchange knowledge and experiences, between colleagues?

Can you please explain briefly how those of you who create the annual timetable evaluate it, after the fact?

Can you please give some examples of how you tie back to earlier timetables, coming into the annual timetable for 2017?

On what level of detail is this feedback done?

Can you please briefly state what the punctuality looks like on the lines, trains and stations you plan for?

Questions on Guidelines

Can you please tell me about the guidelines and support you have available when you create an annual timetable?

Can you please give an example of when you turned to the guidelines?

Can you please give an example of when you used "Olsson's minute"? [A running time supplement of one minute directly following a scheduled stop, named after a timetable planner working several decades ago]

Can you please give examples of where you allocate "node supplements"?

How often do you have to allocate supplements for trains using secondary tracks?

Can you please talk a little about what tools you use when creating a timetable?

Can you please give an example of how close you stick to the timetables applied for by the train operating companies?

How specific are the train operating companies' applications with regards to dwell times?

How binding are the train operating companies' applications with regards to dwell times?

How specific are the train operating companies' applications with regards to run times?

How binding are the train operating companies' applications with regards to run times?

How specific are the train operating companies' applications with regards to arrival times?

How binding are the train operating companies' applications with regards to arrival times?

Questions on Rules of Thumb, Heuristics

What do you typically use as a dwell time for passenger trains?

Can you please give examples of factors that contribute to your allocating a longer or shorter dwell time?

Can you please give examples of when you allocate a longer dwell time?

Can you please give examples of when you allocate a shorter dwell time?

What difference does the train type make?

What difference does the volume of passengers make?

What difference do the punctuality statistics make?

Can you please give an example of when margins are needed in a timetable?

If you believe they are required, where do you allocate the margins?

Can you please give examples of how you decide on the size of margins?

Can you please talk a little about which factors come into play here?

What difference does the location make?

What difference does the train type make?

What difference does the volume of passengers make?

What difference do the punctuality statistics make?

Questions on the Trade-Offs in Timetable Planning

Can you please talk a little about the trade-offs you make in your work, creating the annual timetable?

Can you please give some examples of difficult trade-offs you have made recently?

Trade-offs between short and reliable journey times?

The trade-off between margins at stations or on the line?

The trade-off between concentrated and distributed margins?

Can you please give an example of where you willingly created a delay for a train?

Disclosure

This is an extended version of a paper presented at the 20th EURO Working Group on Transportation Meeting (EWGT2017) in Budapest, on 4–6 September 2017 [83].

Conflicts of Interest

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

Acknowledgments

The authors would like to thank the Swedish Transport Administration for providing funding for the research, as well as providing access to the interviewees and to the governing documentation. Finally, the authors also thank the project reference group for their continuous assistance, guidance, and feedback.

References

[1] Transport Analysis, *Rail Traffic 2015*, Transport Analysis, Stockholm, Sweden, 2016.

[2] M. Grimm, *Järnvägens kapacitetsutnyttjande 2016*, Swedish Transport Administration, Borlänge, Sweden, 2017.

[3] Y. Xia, J. N. Van Ommeren, P. Rietveld, and W. Verhagen, "Railway infrastructure disturbances and train operator performance: The role of weather," *Transportation Research Part D: Transport and Environment*, vol. 18, no. 1, pp. 97–102, 2013.

[4] W. Brazil, A. White, M. Nogal, B. Caulfield, A. O'Connor, and C. Morton, "Weather and rail delays: Analysis of metropolitan rail in Dublin," *Journal of Transport Geography*, vol. 59, pp. 69–76, 2017.

[5] Y. Qin, X. Ma, and S. Jiang, "A Stochastic Process Approach for Modeling Arrival Delay in Train Operations," in *Proceedings of the Transportation Research Board 96th Annual Meeting*, 2017.

[6] M. F. Gorman, "Statistical estimation of railroad congestion delay," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 3, pp. 446–456, 2009.

[7] N. O. E. Olsson and H. Haugland, "Influencing factors on train punctuality - Results from some Norwegian studies," *Transport Policy*, vol. 11, no. 4, pp. 387–397, 2004.

[8] N. Olsson, A. H. Halse, P. M. Hegglund et al., *Punktlighet i jernbanen - hvert sekund teller*, SINTEF Akademisk Vorlag, Oslo, Norway, 2015.

[9] M. Veiseth, N. Olsson, and I. A. F. Saetermo, "Infrastructures influence on rail punctuality," *WIT Transactions on The Built Environment*, vol. 96, pp. 481–490, 2007.

[10] J. Parbo, O. A. Nielsen, and C. G. Prato, "Passenger Perspectives in Railway Timetabling: A Literature Review," *Transport Reviews*, vol. 36, no. 4, pp. 500–526, 2016.

[11] P. B. L. Wiggenraad, *Alighting and boarding times of passengers at Dutch railway stations*, TRAIL Research School, Delft, Netherlands, 2001.

[12] L. Nie and I. A. Hansen, "System analysis of train operations and track occupancy at railway stations," *European Journal of Transport and Infrastructure Research*, vol. 1, pp. 31–54, 2005.

[13] C. W. Palmqvist, N. O. E. Olsson, and L. Hiselius, "Delays for passenger trains on a regional railway line in Southern Sweden," *International Journal of Transport Development and Integration*, vol. 1, no. 3, pp. 421–431, 2017.

[14] R. Watson, *Train Planning in A Fragmented Railway - A British Perspective, [Doctoral, thesis]*, Loughborough University, Loughborough, England, November 2008.

[15] A. Kauppi, J. Wikström, B. Sandblad, and A. W. Andersson, "Future train traffic control: Control by re-planning," *Cognition, Technology & Work*, vol. 8, no. 1, pp. 50–56, 2006.

[16] S. Tschirner, B. Sandblad, and A. W. Andersson, "Solutions to the problem of inconsistent plans in railway traffic operation," *Journal of Rail Transport Planning and Management*, vol. 4, no. 4, pp. 87–97, 2014.

[17] B. Sandblad, A. W. Andersson, and S. Tschirner, "Information Systems for Cooperation in Operational Train Traffic Control," *Procedia Manufacturing*, vol. 3, pp. 2882–2888, 2015.

[18] E. M. Roth, N. Malsch, and J. Multer, "Understanding How Train Dispatchers Manage and Control Trains: Results of a Cognitive Task Analysis," DOT-VNTSC-FRA-98-3, National Technical Information Service, Springfield, VA, USA, 2001.

[19] P. Hellström, *Problems in the Integration of Timetabling And Train Traffic Control*, Uppsala University, Uppsala, Sweden, 2013.

[20] R. Watson, "Prospects for computer aided railway scheduling: Perspectives from users and parallels from mass transit," *Transportation Planning and Technology*, vol. 23, no. 4, pp. 303–321, 2000.

[21] ONTIME, *Benchmark analysis, test and integration of timetable tools*, Delft University of Technology, Delft, Netherlands, 2014.

[22] Swedish Transport Administration, *Improved Capacity*, 2017, Available at https://goo.gl/mVkSW6.

[23] Swedish Transport Administration, *Network Statement 2017*, Swedish Transport Administration, Borlänge, Sweden.

[24] I. A. Hansen and J. Pachl, *Railway Timetabling & Operations*, Eurailpress, Hamburg, Germany, 2nd edition, 2014.

[25] F. Avelino, M. te Brömmelstroet, and G. Hulster, "The Politics of Timetable Planning: Comparing the Dutch to the Swiss," in *Proceedings of the Colloquium Vervoersplanologisch Speurwerk 2006*, pp. 23-24, Amsterdam, Netherlands, November 2006.

[26] D. Wood and S. Robertson, "Planning tomorrow's railway - role of technology in infrastructure and timetable options evaluation," in *Capacity Management Systems, Computers in Railways VIII*, J. Allan, R. J. Hill, C. A. Brebbia et al., Eds., 2002.

[27] M. S. Feldman and B. T. Pentland, "Reconceptualizing organizational routines as a source of flexibility and change," *Administrative Science Quarterly*, vol. 48, no. 1, pp. 94–118, 2003.

[28] D. Kahneman, P. Slovic, and A. Tversky, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York, UK, 1982.

[29] M. Bazerman, *Judgement in Managerial Decision Making*, Wiley, New York, NY, USA, 4th edition, 1998.

[30] C. Argyris and D. A. Schön, *Organizational learning II: theory, method, and practice*, Addison-Wesley, Reading, Mass, USA, 1996.

[31] M. Loock and G. Hinnen, "Heuristics in organizations: A review and a research agenda," *Journal of Business Research*, vol. 68, no. 9, pp. 2027–2036, 2015.

[32] C. B. Bingham and K. M. Eisenhardt, "Rational heuristics: The 'simple rules' that strategists learn from process experience," *Strategic Management Journal*, vol. 32, no. 13, pp. 1437–1464, 2011.

[33] G. Kirkebøen, "Decision Behaviour – Improving Expert Judgement," in *Making Essential Choices with Scant Information*, T. M. Williams, K. Samset, and K. Sunnevåg, Eds., Chippenham and Eastbourne: CPI Antony Rowe, 2009.

[34] V. R. Vuchic, *Urban Transit Operations, Planning And Economic*, John Wiley & Sons, New Jersey, NJ, USA, 2005.

[35] V. A. Profillidis, *Railway Management and Engineering*, Ashgate, Surrey, UK, 4th edition, 2014.

[36] N. Harris, H. Haugland, N. Olsson, and M. Veiseth, *An Introduction to Railway Operations Planning*, A & N Harris, London, UK, 2016.

[37] A. Ceder, "Public Transport Scheduling," in *Handbook of Transport Systems and Traffic Control*, J. K. Button and D. A. Hensher, Eds., pp. 539–558, Elsevier Science ltd., Oxford, UK, 1st edition, 2001.

[38] A. Caprara, L. Kroon, M. Monaci et al., "Chapter 3 Passenger Railway Optimization," in *Handbooks in Operations Research and Management Science*, C. Barnhart and G. Laporte, Eds., vol. 14, pp. 129–187, 2007.

[39] M. Veiseth, P. M. Hegglund, I. Wien, N. O. E. Olsson, and Ø. Stokland, "Development of a punctuality improvement method," *TQM Journal*, vol. 23, no. 3, pp. 268–283, 2011.

[40] M. Samà, A. Ariano, F. Corman, and D. Pacciarelli, "A variable neighbourhood search for fast train scheduling and routing during disturbed railway situations," *Computers & Operations Research*, vol. 78, pp. 480–499, 2017.

[41] S. Tschirner, *The GMOC Model: Supporting Development of Systems for Human Control [Ph.D. dissertation]*, Uppsala University, Uppsala, Sweden, 2015.

[42] A. Gunnarsson, "Swedish Railway Policy in the EU Environment – Railway organization and financing," in *Proceedings of the Workshop in Stockholm*, Ministry of Enterprise, Energy and Communications Sweden, September 2013.

[43] Network Rail, *The Network Code*, Network Rail, London, UK, 2017.

[44] D. B. Netz, *Network Statement 2018*, Frankfurt am Main: DB Netz, 2017.

[45] ProRail, *Network Statement 2017*, ProRail, Utrecht, Netherlands, 2016.

[46] Swedish Transport Administration, *Analysmetod och samhällsekonomiska kalkylvärden för transportsektorn: ASEK 6.0*, Trafikverket, Borlänge, Sweden.

[47] P. Ström, "Interviewed by the corresponding author on September 15, 2017".

[48] R. M. P. Goverde and I. A. Hansen, "Performance indicators for railway timetables," in *Proceedings of the 2013 IEEE International Conference on Intelligent Rail Transportation, IEEE ICIRT 2013*, pp. 301–306, September 2013.

[49] C. W. Palmqvist, N. O. E. Olsson, and L. Hiselius, "An Empirical Study of Timetable Strategies and Their Effects on Punctuality," in *Proceedings of the the 7th International Conference on Railway Operations Modelling and Analysis (Rail Lille 2017)*, 2017.

[50] Swedish Transport Administration, *Noder i järnvägssystemet*, Trafikverket, Borlänge, Sweden, 2015.

[51] Swedish Transport Administration, *Riktlinjer täthet mellan tåg*, Trafikverket, Malmö, Sweden, 2016.

[52] Swedish Rail Administration, *Riktlinjer för tidtabellskonstruktion för tåg på statens spåranläggningar*, Swedish Rail Administration, Borlänge, Sweden, 2000.

[53] M. J. C. M. Vromans, *Reliability of Railway Syste [Ph.D ERIM]*, Research in Management 62, 2005.

[54] R. Rudolph, "Allowances and Margins in Railway Scheduling," in *Proceedings of the World Congress on Railway Research*, pp. 230–238, 2003.

[55] L. Haldeman, *Automatische Analyse von IST-Fahrplanen [M.S. thesis], Quoted in Vromans (2005)*, ETH Zürich, Zürich, Switzerland, 2003.

[56] UIC, *UIC Code 451-1 Timetable Recovery margins to guarantee timekeeping - Recovery margins*, International Union of Railways., Paris, France, 2000.

[57] J. R. Hammerton, "Delivering solutions for a changing business world," *Transactions on the Built Environment*, vol. 18, 1996.

[58] H. Sipilä and A. Lindfeldt, Interviewed by the corresponding author on April 9, 2015.

[59] M. Backman and J. Mattisson, Interviewed by the corresponding author on May 11, 2015.

[60] M. Aronsson, M. Forsgren, and S. Gestrelius, "The Road to Incremental Allocation & Incremental Planning," SICS Technical Report 2012:09, ISSN 1100-3154, Swedish Institute of Computer Science, Kista, Sweden, 2012.

[61] P. Kreuger, M. Aronsson, J. Ekman, and T. Franzén, "Rail Traffic Requirements Engineering," SICS Technical Report 2005:12, ISSN 1100-3154, Swedish Institute of Computer Science, Kista, Sweden, 2005.

[62] M. Forsgren, M. Aronsson, S. Gestrelius, and H. Dahlberg, "Opportunities and challenges with new railway planning approach in Sweden," SICS Technical Report 2012:11, ISSN 1100-3154, Swedish Institute of Computer Science, Kista, Sweden, 2012.

[63] S. Gestrelius, M. Bohlin, and M. Aronsson, On the uniqueness of operation days and delivery commitment generation for train timetables, presented at RailTokyo 2015.

[64] M. Aronsson, M. Forsgren, and S. Gestrelius, "Uncovered capacity in Incremental Allocation," SICS Technical Report T2017:01, ISSN 1100-3154, Swedish Institute of Computer Science, Kista, Sweden, 2017.

[65] V. Svedberg, M. Aronsson, and M. Joborn, "Railway Timetabling Based on Cost-Benefit Analysis," *Transportation Research Procedia*, vol. 22, pp. 345–354, 2017.

[66] E. Andersson, A. Peterson, and J. Törnquist Krasemann, "Robustness in Swedish Railway Traffic Timetables," in *Proceedings of the RailRome 2011: Book of Abstracts 4th International Seminar on Railway Operations Modelling and Analysis*, 2011.

[67] A. Peterson, "Towards a robust traffic timetable for the swedish southern mainline," in *Proceedings of the Computers in Railways XIII: Computer System Design and Operation in the Railway and Other Transit Systems*, pp. 473–484, 2012.

[68] F. Khoshniyat and A. Peterson, "Robustness Improvements in a Train Timetable with Travel Time Dependent Minimum Headways," in *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis - RailTokyo*, Tokyo, Japan, March, 2015.

[69] S. Gestrelius, M. Aronsson, M. Forsgren, and H. Dahlberg, "On the delivery robustness of train timetables with respect to production replanning possibilities," in *Proceedings of the 2nd International Conference on Road and Rail Infrastructure*, 2012.

[70] F. Khoshniyat, *Optimization-Based Methods for Revising Train Timetables with Focus on Robustness*, Linköping University Electronic Press, 2016.

[71] E. V. Andersson, A. Peterson, and J. Törnquist Krasemann, "Quantifying railway timetable robustness in critical points," *Journal of Rail Transport Planning and Management*, vol. 3, no. 3, pp. 95–110, 2013.

[72] J. Warg and M. Bohlin, "The use of railway simulation as an input to economic assessment of timetables," *Journal of Rail Transport Planning and Management*, vol. 6, no. 3, pp. 255–270, 2016.

[73] E. V. Andersson, A. Peterson, and J. Thörnquist Krasemann, "Reduced railway traffic delays using a MILP approach to increase robustness in Critical Points," *Journal of Rail Transport Planning & Management*, no. 5, pp. 110–127, 2015.

[74] E. Solinen, G. Nicholson, and A. Peterson, "A Microscopic Evaluation of Robustness in Critical Points," in *Proceedings of the 7th International Conference on Railway Operations Modelling and Analysis (RailLille 2017)*, pp. 83–103, 2017.

[75] M. Forsgren, M. Aronsson, P. Kreuger, and H. Dahlberg, *The Maraca: a tool for minimizing resource conflicts in a non-periodic railway timetable. Presented at RailRome*, 2011.

[76] M. Forsgren, M. Aronsson, S. Gestrelius, and H. Dahlberg, "Using timetabling optimization prototype tools in new ways to support decision making," *WIT Transactions on the Built Environment*, vol. 127, 2013.

[77] S. Gestrelius, M. Aronsson, M. Joborn, and M. Bohlin, "Towards a comprehensive model for track allocation and roll-time scheduling at marshalling yards," *Journal of Rail Transport Planning Management*, vol. 7, no. 3, pp. 157–170, 2017.

[78] J. T. Krasemann, "Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances," *Transportation Research Part C: Emerging Technologies*, vol. 20, no. 1, pp. 62–78, 2012.

[79] J. T. Krasemann, "Computational decision-support for railway traffic management and associated configuration challenges: An experimental study," *Journal of Rail Transport Planning and Management*, vol. 5, no. 3, pp. 95–109, 2015.

[80] S. Kvale, *Den kvaliativa forskningsintervjun (The Qualitative Research Interview, in Swedish)*, Studentlitteratur, Lund, Sweden, 1997.

[81] M. Hammersley and P. Atkinson, *Ethnography: Principles in Practice*, Routledge, 3rd edition, 2007.

[82] H. Olink and G. Scheepmaker, *RailSys use by RailwayLab. Presented at RailSys Swedish and Scandinavian User Group 2017, at the Royal Institute of Technology in Stockholm, Sweden*, 2017.

[83] C. W. Palmqvist, N. O. E. Olsson, L. Winslott Hiselius, and L. Winslott, "Punctuality Problems from the Perspective of Timetable Planners in Sweden," in *Proceedings of the 20th EURO Working Group on Transportation Meeting, EWGT 2017*, Budapest, Hungary, September 2017.

Hindawi
Journal of Advanced Transportation
Volume 2018, Article ID 5812939, 19 pages
https://doi.org/10.1155/2018/5812939

Research Article

A Simulation Platform for Combined Rail/Road Transport in Multiyards Intermodal Terminals

Xuchao Chen,[1] **Shiwei He**,[1] **Tingting Li**,[2] **and Yubin Li**[1]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[2]*School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518055, China*

Correspondence should be addressed to Shiwei He; shwhe@bjtu.edu.cn

Received 28 September 2017; Revised 4 January 2018; Accepted 24 January 2018; Published 28 February 2018

Academic Editor: Francesco Corman

With the rapid development of multiyards railway intermodal terminal (MYRIT) construction in China, performance evaluation has become an important issue for terminal design and management departments. Due to the complexity of the multiyards terminal and the associated rail network, the train moving process and related terminal operations have become more complicated compared with the traditional intermodal container terminal. However, in general simulation platforms, the train moving process is simplified and train route scheduling rules are not considered in existing simulation models. In order to provide an accurate and comprehensive quantitative evaluation tool for MYRIT, a simulation platform based on the Timed Petri Net model has been developed, which can offer decision support for terminal design and management departments. In this platform, a yards and facilities layout module has been created to give simulation users access to designing the railway network on this platform. And a train route dispatching simulation method has been integrated to provide an accurate simulation of the train moving process. Based on a real case of Qianchang railway intermodal terminal that is located in Fujian Province, China, the platform is thoroughly validated against historical data. And the test scenarios show that train routes arrangement and handling equipment configuration both have a significant influence on overall terminal performance, which need to be carefully considered during terminal design and management.

1. Introduction

In the past several years, the Chinese logistics industry has undergone a rapid development. Along with the trends toward the growing demands quantitatively and qualitatively on the freight transportation system [1], *a* large-scale intermodal terminal construction plan has been implemented by China Railway Corporation. According to the official document, 33 1st-class, 175 2nd-class, and 300 3rd-class modern railway intermodal terminals will be built from 2015 to 2017. Different from the intermodal container terminals that have been widely constructed in Europe or America in the past decades, the intermodal terminal in China usually contains multiple cargo yards, which can be called a multiyards railway intermodal terminal (MYRIT). MYRIT is a kind of comprehensive intermodal terminal which comprises distinguished types of cargo yards in order to offer multimodal transport services to different types of cargoes. A typical MYRIT is shown in Figure 1, which contains a bulk cargo (e.g., coal and ore) yard, a container yard, a special cargo yard, and others.

In MYRIT, different cargo yards provide services to different kinds of cargoes and contain independent facilities (railway tracks, storage area, handling equipment, etc.), resulting in a very complex process of terminal design which involves a huge number of decisions [2]. For a better terminal design, one of the effective methods is simulation, which can provide performance evaluation of terminal design schemes based on a simulation model. In fact, intermodal terminals can be regarded as a kind of discrete event dynamic system (DEDS), of which the states only get changed at discrete points in time as a result of stochastic events [3].

Due to the complexity of the MYRIT and the associated railway network, the train moving process simulation becomes an important issue for overall terminal evaluation. With various cargo yards, the internal railway network involves more tracks, switches, and signals, and the

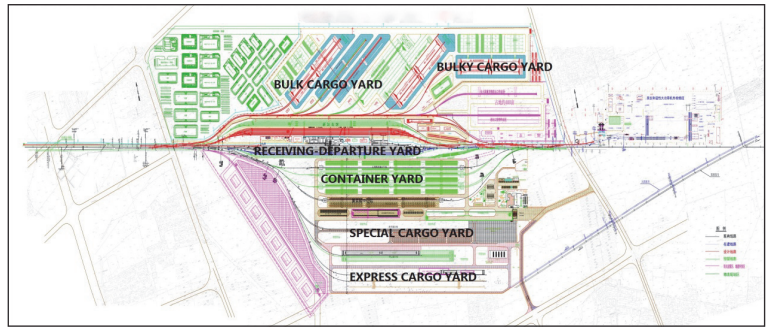


Figure 1: Typical layout of multiyards railway intermodal terminal.

interaction of train routes becomes more complicated, which has a significant influence on the train moving process. However, in existing simulation platforms, there has been a lack of simulation methods which can provide a detailed simulation of the train moving process. The train moving process has been simplified in most general simulation models, which will lead to an inaccuracy in the terminal evaluation.

In order to provide an accurate and comprehensive simulation of MYRIT, a simulation platform is developed and presented in this paper which can be used as an analysis and predictive tool for terminal design and management departments. In this platform, a yards and facilities layout module has been invented which can be used for inputting and editing the railway network of a multiyards terminal, and a train route dispatching simulation method (TRDSM) has been created to provide an accurate simulation of the train moving process considering basic dispatching rules. The architecture and the simulation model of the platform are elaborated in this paper, and it should be noted that this platform is dedicated to rail/road intermodal terminals, and terminals which involve waterway transportation like container ports are not applicable to this system.

The remaining part of this paper is organized as follows. In the following section, a literature review of intermodal terminal optimization and management based on simulation techniques is provided. Then, in Section 3, an overview of operations of MYRIT and the framework of this simulation platform are presented. In Section 4, the simulation models based on Timed Petri Net (TPN) and key simulation methods integrated in the platform are expounded. In Section 5, the function introduction of the simulation platform that we have developed is given. In Section 6, validation and some test scenarios of the simulation platform are detailed. Finally, a summary of this study follows and possible extensions on intermodal terminal simulation are identified.

2. Literature Review

Simulation techniques have been used successfully in terminal planning, design, and optimization problems, and there are many literatures in this field. Rizzoli et al. [4] built a simulation model of the flow of intermodal terminal units (ITUs) among and within inland intermodal terminals using MODSIMIII as a development tool. During the simulation, various statistics are gathered to assess the performance of the terminal equipment. Another simulation tool dedicated to model a single terminal was introduced by Benna and Gronalt [5]. Terminal layout, arrival patterns of trains and trucks, and container settings were specified as part of the input data. A general overview of the macro and micro model was provided by Ballis and Golias [6]. They tested the micro model for 17 different terminal layouts with varying numbers of tracks and cranes as well as lifting technologies. Marinov and Viegas [7] provided a yard simulation modeling methodology for analyzing and evaluating flat-shunted yard operations using SIMUL8. A comparative evaluation tool for rail/road freight transport terminals has been developed by Ballis and Golias [8], which consists of three models (an expert system, a simulation model, and a cost calculation model). Fugihara et al. [9] provided a way of simulation technology which can generate several benefits in the distribution center projects.

This paper focuses on the simulation of rail/road intermodal terminals; however, some literatures with regard to seaport terminal simulation can also be used as references. Hartmann [10] provided an approach for generating scenarios of seaport terminals which can support solving optimization problems in container terminal logistics. Vis [11] compared the use of manned straddle carriers with that of automated stacking cranes. The total travel time required to handle all container moves was applied as a performance measure to determine the yard layout for the seaport terminal. A computer simulation model with on-screen animation graphics which can simulate the operations of a container terminal equipped with straddle carriers was introduced by Ballis and Abacoumkin [12]. Based on this model, different configurations (changes in yard layout, equipment number and productivity, truck arrival pattern, and service discipline) of the simulated system can be evaluated. Shabayek and Yeung [13] developed a simulation model to simulate Kwai Chung container terminals in Hong Kong, using Witness software. The layout of the port and the incoming and outgoing routes for container vessels were considered in the model, which can provide a prediction of terminal operations with a high order of accuracy.

Petri Net is widely used for the description of the structure and dynamics of DES [14]. For more details of the Petri Net

theory and modeling, refer to Peterson [15] and David and Alla [16]. A great number of successful Petri Net models have been designed for terminal simulation and optimization throughout the world. Some of the most relevant ones are expanded in the following paragraphs. Dotoli et al. [17] addressed the issue of modeling and managing Intermodal Transportation Systems (ITS) at the operational level. The system is highly complex, with various types of conveyances alongside scheduling aspects. A Timed Petri Net framework was built to model the ITS, which contains the tollbooth, highways, truck, railway, port, and ship subsystems. A similar exploration has been performed by Maione and Ottomanelli [18]. A container terminal simulation model was proposed within the theoretical framework of Petri Nets, which allows taking into account the different aspects of the considered system. Lee et al. [19] presented the development and application of simulation models for air cargo terminal operations using Timed Color Petri Net (TCPN). The TCPN is able to simulate the operations of various types of material handling equipment and is validated based on actual cargo retrieval schedules records. A high-level Petri Net model which contains timed predicate/transition net has been provided by Hsu et al. [20]. This model aims at solving the three essential operational problems (berth allocation problem, quay crane assignment problem, and quay crane scheduling problem) of container terminals simultaneously, which can result in good overall system performance. Cavone et al. [21] proposed a procedure for planning and managing resources in intermodal terminals, which integrates Timed Petri Nets and Data Envelopment Analysis. Silva et al. [22] described the modeling of a container terminal using Petri Net with predicates, which allowed the evaluation of different configurations and combinations of transport equipment, providing a very complete port system simulation. A general modeling framework based on Timed Petri Net was constructed by Dotoli et al. [23], which allowed simulating and evaluating the performance of key elements within the intermodal transportation chain. In the Petri Net model, places represent resources and capacities or conditions, transitions model inputs, flows, and activities into the terminal, and tokens represent intermodal transport units.

Previous findings show that the simulation method has been very effective and was widely used in the field of terminal optimization and management. These manuscripts provide valuable supports for our work; the simulation framework, control methods of the terminal entities, and Petri Net models of terminal activities are used as references in our simulation platform. However, there has been a lack of simulation methods considering multiple cargo yards, and most papers focus on container terminal simulation. With only one type of cargo, the structure of the container terminal and the associated railway network is relatively simple compared with MYRIT. The simulation model of the train moving process is simplified to a single discrete event in great majority of existing studies, and detailed train route scheduling rules are not considered. However, due to the complexity of the rail network in MYRIT, the interaction of train movements can have a significant influence on the train moving process and terminal performance. Therefore, a detailed simulation method of the train moving process considering train route dispatching rules has been considered as an important element in this platform.

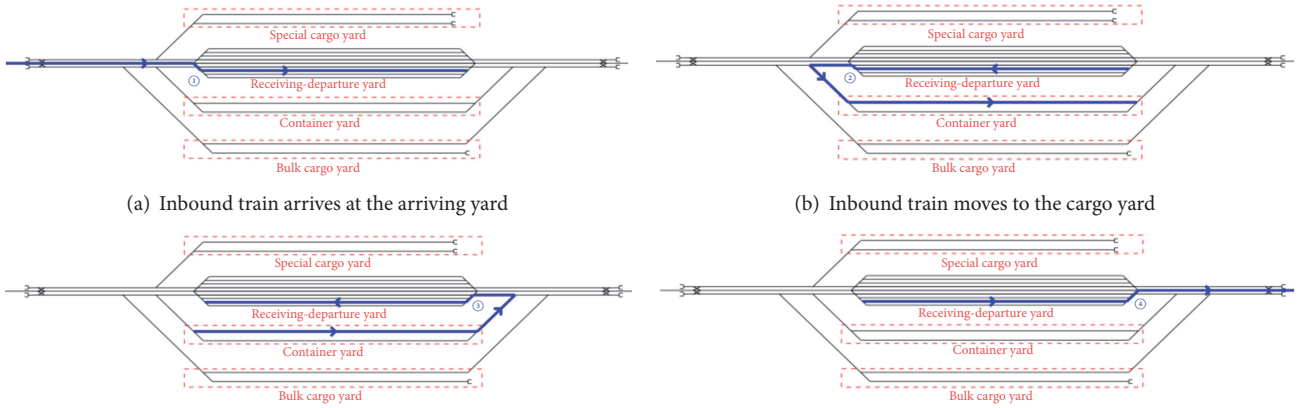
3. Framework Design

MYRIT is a kind of multimodal freight hub where different types of cargoes are delivered and picked up by trains or trucks. Cargoes arrive at terminal by trains/trucks and are unloaded in corresponding yards by cargo-handling appliances (gantry cranes, forklifts, reach stackers, etc.). Certain goods are picked up by trains/trucks directly, while some goods need to be processed and stored within the terminal for a period of time before they are picked up. Various facilities and equipment are needed to finish these tasks, such as railway tracks, platforms, handling machineries, truck parking lots, and repertories, which constitute a large-scale intricate logistics system. In this section, the main operations of MYRIT which are considered in the TPN model are introduced, and the framework of the simulation platform is described as well.

3.1. Overview of MYRIT Operations. The main operations in MYRIT can be divided into three components, namely, the train operations, the truck operations, and the cargo-handling operations.

3.1.1. Train Operations. There are four prime steps of train operations. When an inbound train arrives at the terminal, it enters the receiving-departure yard (or arriving yard) first and then undergoes an arrival inspection (cargo information check, safety inspection, etc.) conducted by a surveyor. After that, according to the type of goods carried by the train, the occupancy states of tracks in the corresponding cargo yard are checked. If there are idle side tracks within the yard, the inbound train will move into the target yard under the command of the train dispatching office. The cargo loading/unloading operations are implemented inside the cargo yard by means of certain cargo-handling appliances. The type and mechanical model of appliances depend on the physical nature of the cargoes. In case the loading/unloading machine is busy, the train will stay on the side track and wait to be served. When the cargo loading/unloading task is finished, the outbound train will move back to the receiving-departure yard (or departure yard), and a departure inspection will be carried out. At last, the outbound train leaves the terminal at the time required by the timetable. It should be noted that, sometimes, the outbound train can leave the terminal from the cargo yard directly without entering the departure yard. The typical trajectories of train operations are shown in Figure 2.

3.1.2. Truck Operations. Trucks arrive at MYRIT to deliver cargoes to outbound trains or to pick up cargoes from inbound trains. When a truck arrives at the entrance gate of the terminal, it joins a first-in first-out (FIFO) queue and waits for arrival inspection which is implemented by the gate system. The arrival inspection includes collecting



(a) Inbound train arrives at the arriving yard

(b) Inbound train moves to the cargo yard

(c) Outbound train moves to the departure yard

(d) Outbound train departs from the terminal

Figure 2: Trajectories of train operations.

information of trucks and goods through a high-resolution camera, weighing arrival trucks, and providing guidance information for truck drivers. Several factors can influence the work efficiency of arrival inspection, such as the number of gates, the number of entrance channels, the location of gates, and the degree of automation of gate information management system.

After the arrival inspection, the arrival truck moves to the corresponding cargo yard according to the guidance information from the gate system. When the truck enters the cargo yard, in the majority of cases, it will join a FIFO queue in the loading/unloading area. If there are available cargo-handling machineries, the truck loading/unloading procedure begins. When the truck has been loaded or unloaded, it will move to the exit gate of the terminal. Similar to the entry process, the departure truck needs a departure inspection at the exit gate.

In addition to the trucks mentioned above, there is another type of trucks which are called the shuttle trucks. Shuttle trucks work inside the terminal and are used for transporting goods between different cargo yards. The number of shuttle trucks is much lower than the number of external trucks which will leave the terminal when the loading and unloading tasks are completed. Shuttle trucks can be regarded as a kind of equipment resources of MYRIT. The sketch map of the truck operations within MYRIT is shown in Figure 3.

3.1.3. Cargo-Handling Operations. There are multiple types of handling machineries in MYRIT, such as gantry cranes, front lifters, reach stackers, and forklifts. Each kind of machinery has different mechanical properties and serves different types of cargoes in diverse cargo yards. The cargo-handling operations can be divided into two types according to the delivery direction of the goods: the train-truck operation which represents the handling operation of goods delivered by trains and picked up by trucks and the truck-train operation which represents the handling operation of goods delivered by trucks and picked up by trains. In this section, the train-truck cargo-handling operation is introduced as a representative.

When the cargo which is delivered by inbound trains is about to be unloaded within the cargo yard, one of the following three circumstances is given:

(i) If the corresponding truck which is used for picking up the cargo has arrived at the cargo yard, the cargo will be unloaded directly from the inbound train to the truck. In this situation, only one loading/unloading operation is required.

(ii) If the corresponding truck cannot catch up with the deadline (the time when the inbound train has been unloaded and must leave), the cargo will be unloaded to the storage area inside the cargo yard and stored there. After that, when the corresponding truck arrives at the cargo yard, the cargo will be loaded from the storage area to the truck. In this case, two loading/unloading operations are required.

(iii) If the cargo needs special storage condition (e.g., refrigerated container) or the cargo needs to be processed inside MYRIT (e.g., express parcels need to be sorted in the sorting workshop of MYRIT before they are picked up by shippers), it will be unloaded to the shuttle truck and delivered to the corresponding area (as shown in Figure 3).

The process of truck-train operation is basically contrary to the process of the train-truck operation and also can be divided into three cases. Due to the limitation of the space and avoidance of repetition, the details of this procedure are not discussed here.

3.2. Framework of Simulation Platform. Based on the operations overview of MYRIT, the framework of this simulation platform is shown in Figure 4. The framework has four layers, namely, the Petri Net layer, the simulator layer, the layout layer, and the user layer, which are elaborated as follows:

(i) Petri Net layer: the Petri Net layer is the basis of the simulation platform which composes three TPN models, that is, the train operation model, the truck operation model, and the cargo-handling model.

(ii) Simulator layer: the simulator layer contains the core simulation modules and key methods integrated in this platform, including the train generation method,

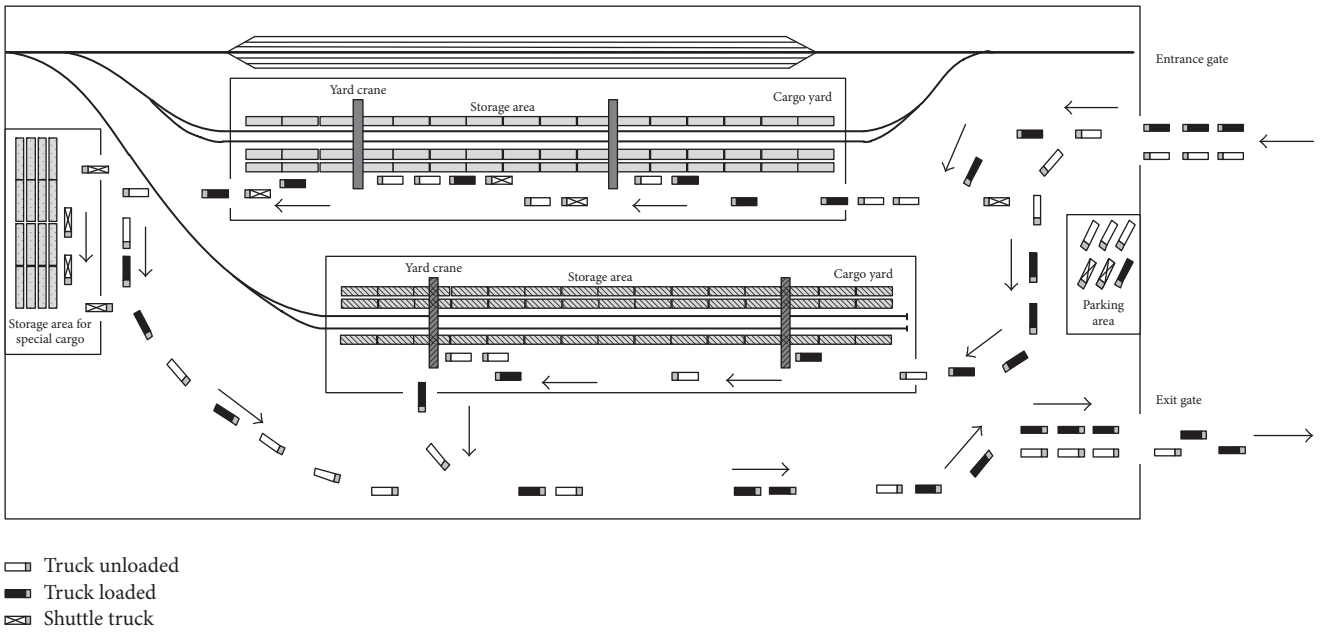


Figure 3: Scheme of truck operations in multiyards railway intermodal terminal.

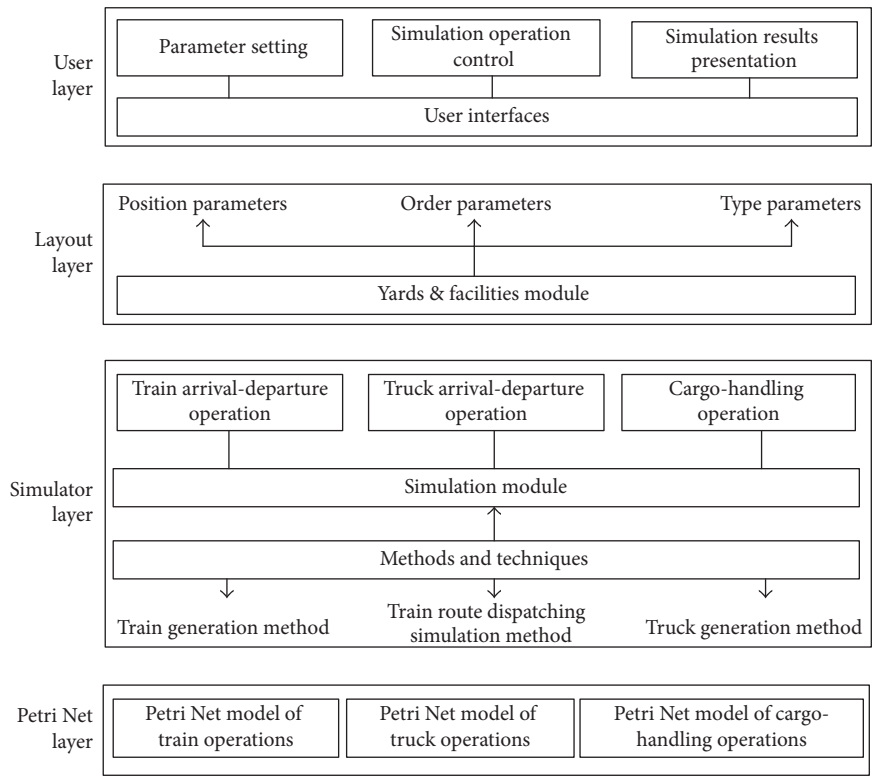


Figure 4: Framework of simulation platform.

the truck generation method, and the train route dispatching simulation method.

(iii) Layout layer: the layout layer contains the yards and facilities layout module which allows users to design or modify the yards and facilities (tracks, switches, signal machines, etc.) locations by adjusting certain parameters. Three types of parameters are designed as follows:

(a) Position parameters, which determine the relative position relation (horizontal and vertical)

between certain cargo yard and the receiving-departure yard.

(b) Order parameter, which determines the relative position relation among different cargo yards.

(c) Type parameter, which determines the type of certain cargo yard. According to the layout of rail tracks, cargo yards can be classified into 3 types: through-type cargo yard, stub-end-type cargo yard, and mixed-type cargo yard.

(iv) User layer: the user layer contains the operation control module, the simulation project database, the simulation results database, and related interfaces. Simulation users are able to access this layer to set up a simulation project, control the simulation progress, and obtain the simulation analysis results via user-friendly interfaces.

4. Simulation Model

Petri Net is a widely used tool using graphic elements as a representation to describe the structure and dynamics of DEDS, such as computer systems and manufacturing systems. TPN is a bipartite diagraph described by the five-tuple as shown in the following formula:

$$\mathrm{TPN} = (P, T, \mathrm{Pre}, \mathrm{Post}, F), \quad (1)$$

where P represents the set of places with $|P| = m$, T is the set of transitions with $|T| = n$, Pre is the preincidence matrix with Pre: $P \times T \rightarrow N^{m \times n}$, Post is the postincidence matrix with Post: $P \times T \rightarrow N^{m \times n}$, and the function F: $T \rightarrow R^{+}$ specifies the timing associated with each transition.

A simulation model has been established based on TPN in this study. There are three types of transitions used in this model: immediate transition, stochastic transition, and deterministic timed transition. Corresponding to the operations of MYRIT, the TPN model is segmented into three submodels which are the train arrival and departure model, the truck arrival and departure model, and the cargo-handling model.

4.1. Train Arrival and Departure Model

4.1.1. Train Generation Method. This platform provides two types of train generation methods. The first one is to generate inbound trains according to a fixed timetable input by simulation users. Each train arrival time is generated based on a historical record of train arrival events. This method applies to the terminals which have been put into use and is particularly useful to perform trace-driven simulation.

The second method is to generate inbound trains arrivals according to a stochastic mathematical distribution which is specified by simulation users. This generation method applies to the terminals that are still in the stage of design or construction, and it can be used to test alternative arrival patterns.

4.1.2. Train Route Dispatching Simulation Method. As introduced above, TRDSM is created to simulate the train dispatching operation which can provide an accurate simulation of the train moving process. According to the relationship between two different train routes, train routes pairs can be divided into two types: conflicting train routes pair (CTRP) and parallel train routes pair (PTRP). Train routes in CTRP contain some of the same equipment. As shown in Figure 5, train route $R1$ (the golden line) shares a section of track (marked by the red circle) with train route $R2$ (the blue line), which means if train route $R1$ is occupied by a train, then train route $R2$ becomes unavailable as well because it is impossible for two trains to move on the same track simultaneously.

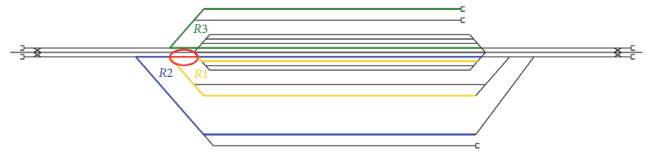


Figure 5: Diagram of train route samples.

Contrary to the CTRP, there is no space conflict in PTRP. Train route $R1$ is independent of train route $R3$ (the green line), which means if one train occupies $R1$, another train can occupy $R3$ at the same time.

Based on the analysis of different types of train routes pairs, the framework of the TRDSM is presented in Figure 6, and the main steps are as follows.

Step 1 (train route database presetting). There are various train routes in MYRIT. The train routes of different types of trains and different train move events are diverse. All train routes should be set by simulation users in advance according to the station operation regulations, and all the preconfigured train routes are stored in the database.

Step 2 (generating train move request). Train move request is generated from the trains which have finished the preceding operation and are ready to move to the next operating location or the trains which have been added to the waiting queue for available train route. When a train move request is generated, the corresponding train route will be invoked from the train route database.

Step 3 (examining the availability of train route). The availability of train route is determined by the availabilities of the key points (switches and signal machines) within it. To be specific, only when all key points of the train route are not occupied can this train route be available. Otherwise, this train route is unavailable and the corresponding train will join the FIFO queue.

Step 4 (generating train move event). If the train route is available, it will be occupied by the train and a train move event will be generated. The occupancy states of the key points in this train route will be updated based on the real-time location of the moving train.

Step 5 (examining the waiting queue). If all train move requests have been executed, the method stops. Otherwise, another train move request is generated according to the FIFO rules, and the method returns to Step 2.

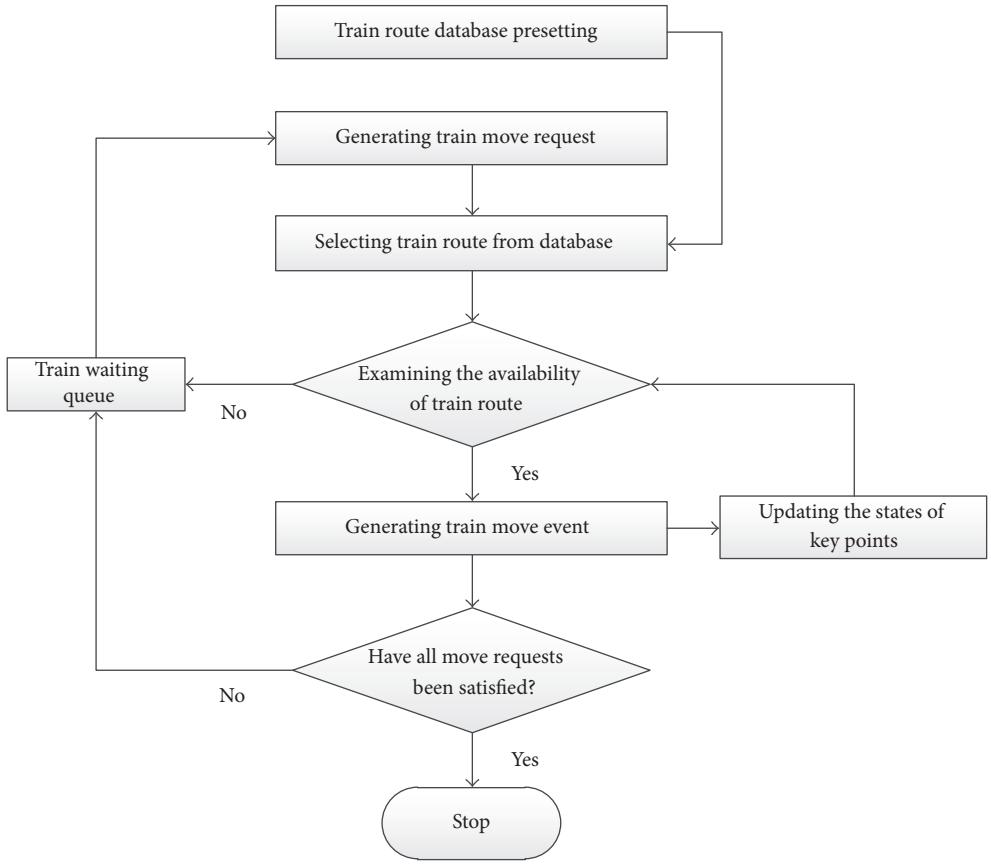


Figure 6: Framework of train route dispatching simulation method.

4.1.3. Description of Train Arrival and Departure Model. The TPN model of train arrival and departure process is presented in Figure 7. Train generation is represented by a stochastic transition T1. For an inbound train (P1), when it is about to enter the arriving yard (or the receiving-departure yard) (T2), the availabilities of the destination track and the corresponding train route must be checked; places P3 and P4 are added to represent the available destination track and the available train route, respectively. Similarly, for each train moving event (T4, T7, T8, and T9), the availabilities of tracks and train routes must be checked. The availability inspection of train route is realized based on TRDSM, which is represented by an immediate transition T10. To check the availabilities of train routes, the real-time location information of the train which is expressed by several places (P17, P19, P20, and P22) needs to be acquired. Places P5 and P6 represent the empty inbound train and the loaded inbound train, respectively. Correspondingly, P7 indicates the train which has been loaded and P8 indicates the train which has been unloaded.

For outbound trains, two places (P9 and P10) are added to represent trains with different departure modes. Place P9 indicates the trains which can leave the terminal directly without moving to the departure yard (or the receiving-departure yard), while place P10 indicates the trains which need to finish the departure inspection at the departure yard before leaving the terminal. Transition T6 is added to distinguish different train departure modes.

4.2. Truck Arrival and Departure Model

4.2.1. Truck Generation Method. A truck generation method is designed to generate stochastic arrivals of trucks according to certain mathematical distribution. According to the study by Rizzoli et al. [4], the truck arrival pattern can be approximately described by the negative exponential distribution.

4.2.2. Description of Truck Arrival and Departure Model. When a truck is generated by the truck generation method (T11), it joins a FIFO queue at the entrance gate. An arrival inspection (T12) will begin if there is a free entrance channel which is represented by the place P38.

Considering that the time cost of the truck moving process is not fixed, the truck moving process inside the terminal (from the terminal gate to a certain cargo yard and vice versa) is represented by two stochastic transitions (T14 and T17). And the time cost of truck moving is produced based on stochastic distribution. Place P37 indicates the available parking spaces within the cargo yard. And T18 is a deterministic timed transition which represents the departure inspection of outbound trucks. The TPN model of truck arrival and departure process is also presented in Figure 7.

4.3. Cargo-Handling Model. Corresponding to the cargo-handling operations, the cargo-handling model can also be divided into two submodels, the train-truck cargo-handling

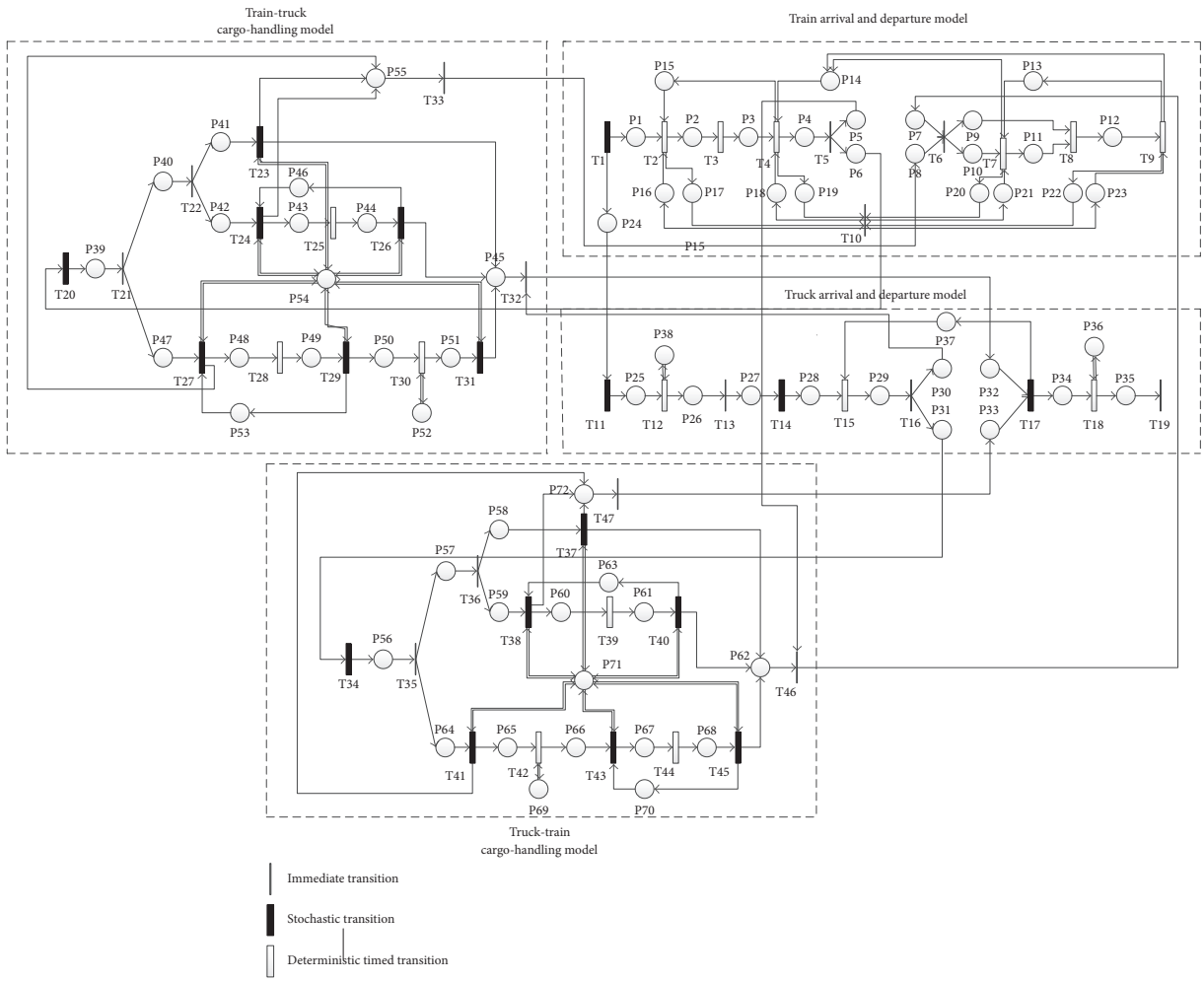


Figure 7: TPN model of multiyards railway intermodal terminal.

model and the truck-train cargo-handling model which are shown in Figure 7. In this section, the train-truck model is introduced as a representative.

The cargo which needs processing operation in MYRIT (or needs to be stored in a special area) is represented by place P47. This kind of cargo needs to be picked up by the shuttle trucks and carried to the corresponding area (T28). Three deterministic timed transitions (T27, T29, and T31) are added to represent the cargo unloading event from trains to shuttle trucks, the cargo unloading event from shuttle trucks to the machining region (or storage area for special goods), and the cargo loading event from machining region (or storage area for special goods) to consignee's trucks, respectively. The processing operation is represented by transition T30. Place P52 represents the resources for processing operation and P53 represents the available shuttle trucks.

Normal cargo which does not need to be machined or stored in the special area is represented by place P40. Place P41 represents the cargo as described in case (i) in Section 3.1.3 which can be picked up by trucks directly. Place P42 represents the cargo which needs to be stored within the terminal temporarily. The cargo unloading event and the storage process are represented by transitions T24 and T25, respectively. Place P48 indicates the available storage space.

Four transitions (T32, T33, T46, and T47) are added to connect different submodels. T32/T46 indicates the event of generating loaded outbound truck/train when the loading operation has been finished, and T47/T33 represents the event of generating empty outbound truck/train when the unloading operation is completed.

5. Multiyards Railway Intermodal Terminal Simulation Platform

Based on the TPN model, a multiyards railway intermodal terminal simulation platform is developed, using C# as a development tool. The MYRIT simulation platform can provide realistic reproduction of the main logistic activities and entities flows (e.g., trains, trucks, and cargoes) which occur inside the terminal. It can be used to evaluate the terminal design scheme and management strategies based on the quantitative analysis of simulation results and to provide decision support for terminal planning and design. In this section, a brief introduction of this platform is presented.

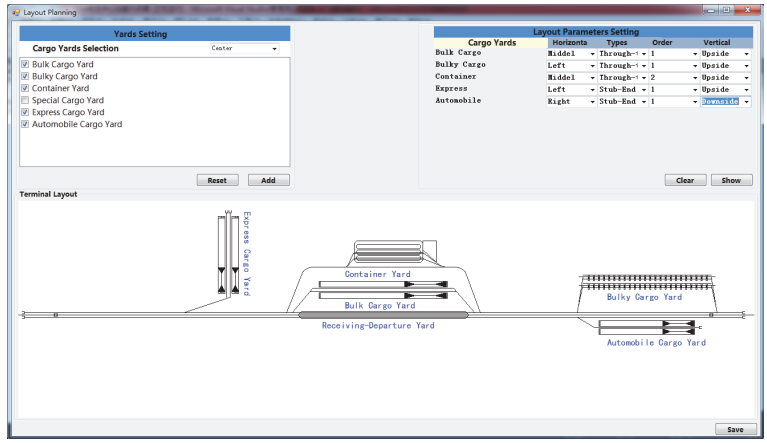


Figure 8: Interface of terminal layout planning.

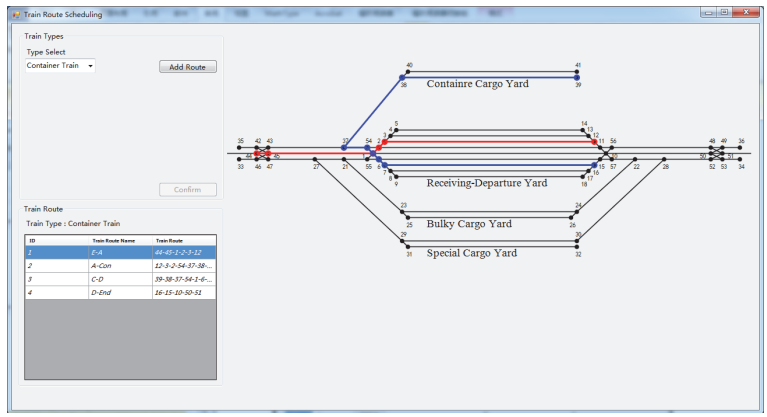


Figure 9: Interface of train route presetting.

As mentioned above, terminal layout has a significant impact on the train operations and should be determined from the very beginning of setting up a simulation project. In this platform, a graphical interface (as shown in Figure 8) for layout planning of MYRIT is provided, where simulation users can design or modify the terminal layout by setting the parameters as introduced in Section 3.2.

Once the terminal layout has been identified, the structure of the internal railway network of MYRIT can be determined. Simulation users should preset train routes according to the station operation regulations via an interface of train route presetting as shown in Figure 9. The internal railway network of MYRIT is described by the lines which represent the rail tracks and the points with unique ID numbers which represent the center points of the turnouts. Each train route can be represented by a sequential list of ID numbers as shown in Figure 9 (e.g., the red train route is represented by the numeral string "44-45-1-2-3-12"). The intersections among various train routes may cause train congestion in the rail yard and can lead to efficiency decline of train operations.

A statistical analysis of the performance of storage area can be provided by the platform. As shown in Figure 10, a sketch map of the operation in cargo yard during the simulation is reported. A train is being unloaded within the cargo yard and a truck arrives to pick up cargo. The real-time volume curve of the cargoes which are stored in the storage area is displayed. Based on the detailed records of the cargo volumes, the utilization ratio of storage area in a certain cargo yard can be calculated, as shown on the right side of Figure 10.

In Figure 11, several indexes of the cargo volumes are reported. The pie chart shows the proportion of the quantity of each type of goods to the total volume of goods during the simulation period. And from the histogram, the volume of each type of goods delivered by trains and trucks and the volume of each type of goods picked up by trains and trucks can be obtained. A group of line charts show the relationships between the volumes of inbound cargo and outbound cargo.

Based on the TRDSM, this platform can provide elaborate reproduction of the train moving process within the terminal, and major time information of the train moving process can be accessed by users via the interface as shown in Figure 12. A detailed statistical analysis of train performance can be obtained on the platform based on the simulation results. Several statistical charts about train operation time are reported in Figure 13. Among them, the first chart shows the number of trains and the time consumption of trains staying at the cargo yard.

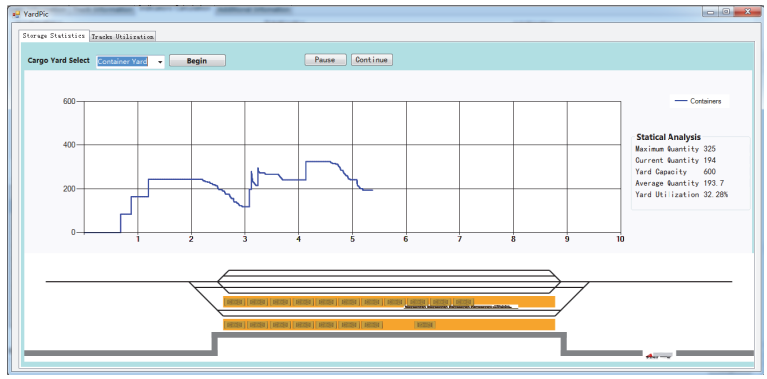


Figure 10: Snapshot of the platform interface during simulation.

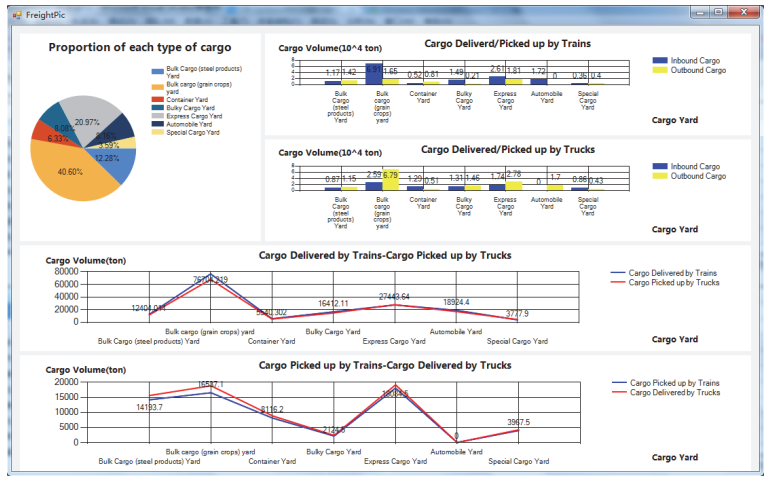


Figure 11: Statistical analysis of cargo volumes based on the simulation results.

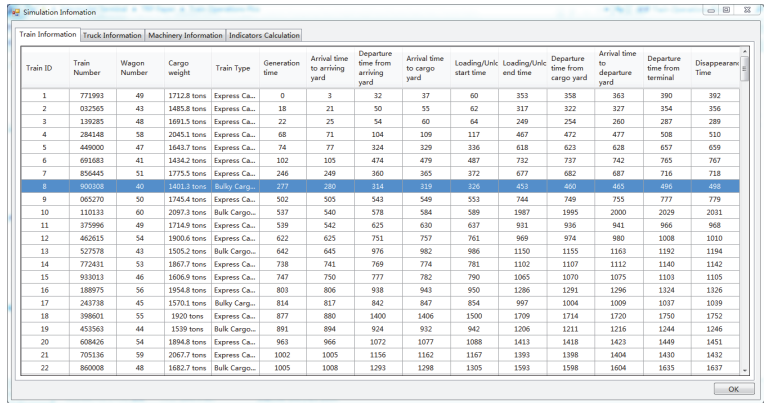
The utilization ratio of handling equipment can also be analyzed. An example of the historical record of the crane utilization ratio in the container yard during the simulation period is shown in Figure 14. Utilization ratio is one of the key indicators of handling equipment, which reflects the busy degree of the equipment resources. In general, low utilization ratio may indicate that there is a problem of equipment redundancy, and high utilization ratio (in a reasonable range) means the handling equipment is fully used. Therefore, from the perspective of investment benefit, high utilization ratio is more likely to be beneficial for the investors. However, exorbitant high utilization ratio also means possibility of having lengthy truck queue in the cargo yard. Therefore, the queue length of trucks needs to be analyzed as well. The sample result of truck queue length in cargo yard is presented in Figure 15.

6. Validation and Test Scenarios of Simulation Platform

Validation is extremely important in the validation phase to ensure that the result of the simulation model is able to reproduce the reality under different conditions [24]. In this section, a simulation case of Qianchang railway intermodal terminal has been performed to verify the capacity of the platform of reproducing the real terminal behavior. And two test scenarios are presented to investigate the impact of variations of train route arrangement and handling equipment configuration on terminal performance.

6.1. Simulation Case Setup. The simulation case is set up based on the Qianchang railway intermodal terminal which is located in Fujian Province, China. Qianchang railway intermodal terminal provides transport services of various types of cargoes, including commercial vehicles, electromechanical equipment, steel products, and cold chain goods. The intermodal terminal covers an area of more than 2 square kilometers which contains one receiving-departure yard and seven independent cargo yards. These cargo yards consist of a bulk cargo yard for steel products, a bulk cargo yard for grain crops, a container yard, a bulky cargo yard, an express cargo yard, an automobile cargo yard for commercial vehicles, and a special cargo yard for cold chain cargoes. The layout of Qianchang terminal edited by the yards and facility module of the simulation platform is shown in Figure 16.

The initial data used in the validation is provided by the Qianchang railway intermodal terminal based on a monthly statistical report. The dataset spans a 30-day period and



Train ID	Train Number	Wagon Number	Cargo weight	Train Type	Generation time	Arrival time to arriving yard	Departure time from arriving yard	Arrival time to cargo yard	Loading/Unlc start time	Loading/Unlc end time	Departure time from cargo yard	Arrival time to departure yard	Departure time from terminal	Disappearan Time
1	771993	49	1712.8 tons	Express Ca...	0	3	32	37	60	353	358	363	390	392
2	032565	43	1485.8 tons	Express Ca...	18	21	50	55	62	317	322	327	354	356
3	139285	48	1691.5 tons	Express Ca...	22	25	54	60	64	249	254	260	287	289
4	284148	58	2045.1 tons	Express Ca...	68	71	104	109	117	467	472	477	508	510
5	449000	47	1643.7 tons	Express Ca...	74	77	324	329	336	618	623	628	657	659
6	691683	41	1434.2 tons	Express Ca...	102	105	474	479	487	732	737	742	765	767
7	856445	51	1775.5 tons	Express Ca...	246	249	360	365	372	677	682	687	716	718
8	900308	40	1401.3 tons	Bulky Carg...	277	280	314	319	326	453	460	465	496	498
9	065270	50	1745.4 tons	Express Ca...	502	505	543	549	553	744	749	755	777	779
10	110133	60	2097.3 tons	Bulk Cargo...	537	540	578	584	589	1987	1995	2000	2029	2031
11	375996	49	1714.9 tons	Express Ca...	539	542	625	630	637	931	936	941	966	968
12	462615	54	1900.6 tons	Express Ca...	622	625	751	757	761	969	974	980	1008	1010
13	527578	43	1505.2 tons	Bulk Cargo...	642	645	976	982	986	1150	1155	1163	1192	1194
14	772431	53	1867.7 tons	Express Ca...	738	741	769	774	781	1102	1107	1112	1140	1142
15	933013	46	1606.9 tons	Express Ca...	747	750	777	782	790	1065	1070	1075	1103	1105
16	188975	56	1954.8 tons	Express Ca...	803	806	938	943	950	1286	1291	1296	1324	1326
17	243738	45	1570.1 tons	Bulky Carg...	814	817	842	847	854	997	1004	1009	1037	1039
18	398601	55	1920 tons	Express Ca...	877	880	1400	1406	1500	1709	1714	1720	1750	1752
19	453563	44	1539 tons	Bulk Cargo...	891	894	924	932	942	1206	1211	1216	1244	1246
20	608426	54	1894.8 tons	Express Ca...	963	966	1072	1077	1088	1413	1418	1423	1449	1451
21	705136	59	2067.7 tons	Express Ca...	1002	1005	1156	1162	1167	1393	1398	1404	1430	1432
22	860008	48	1682.7 tons	Bulk Cargo...	1005	1008	1293	1298	1305	1593	1598	1604	1635	1637

Figure 12: Time information of the train moving process.

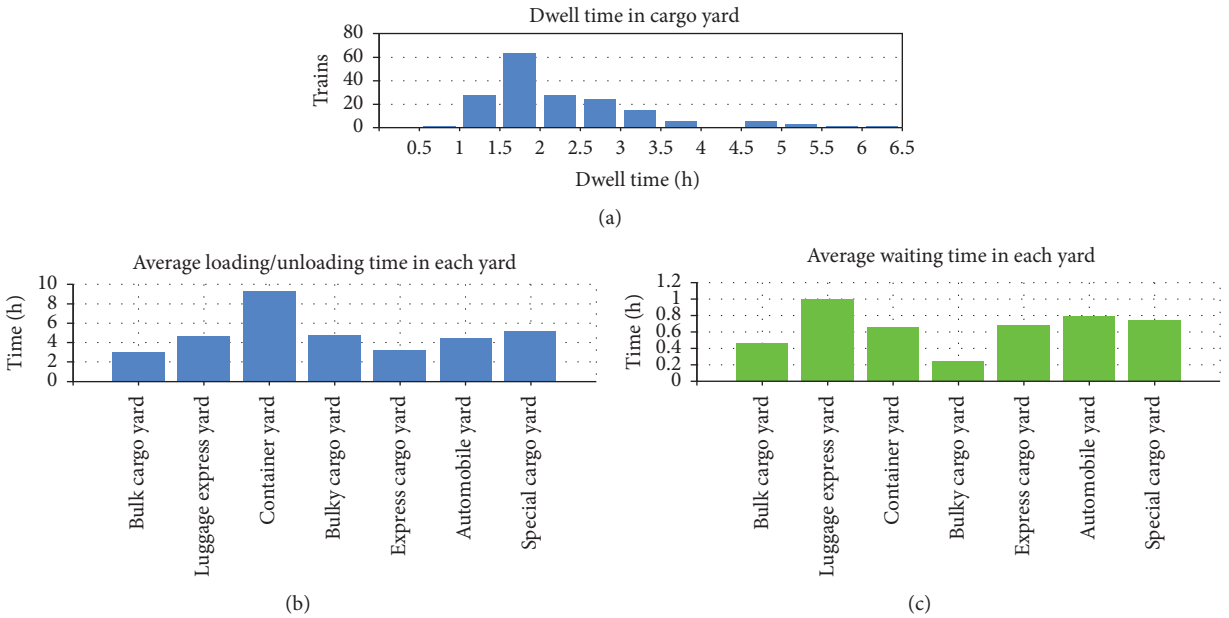


Figure 13: Statistical analysis of train performance.

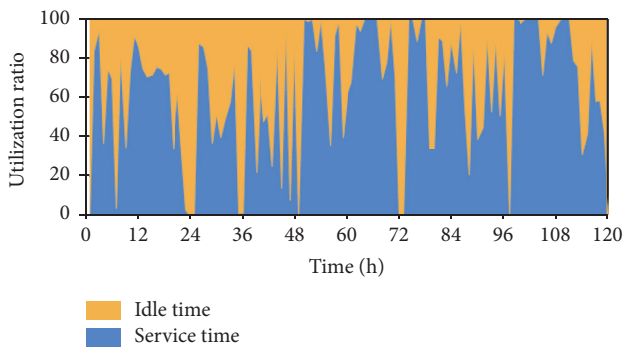


Figure 14: Crane utilization ratio in container yard.

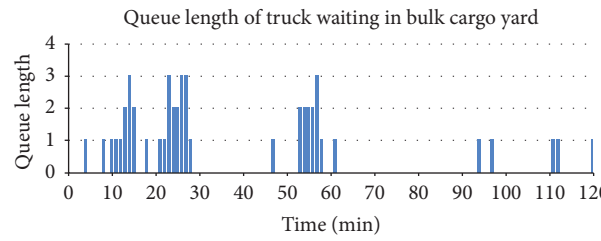


Figure 15: Queue length of trucks in bulk cargo yard.

records detailed information on several key performance indicators which include the cargo volumes, average train loading/unloading time, average truck terminal dwell time, and utilization rate of handling machineries.

6.2. Results and Analysis. Considering the variability in the simulation procedure, the results used for validation are summarized from a set of 500 simulations, which provide reliable simulation statistics of the activities within the intermodal terminal. The efficiency of the constructed simulation framework ensures that a single repetition costs about 4.7 seconds on a 2.5 GHz i5-3210M dual-core processor with

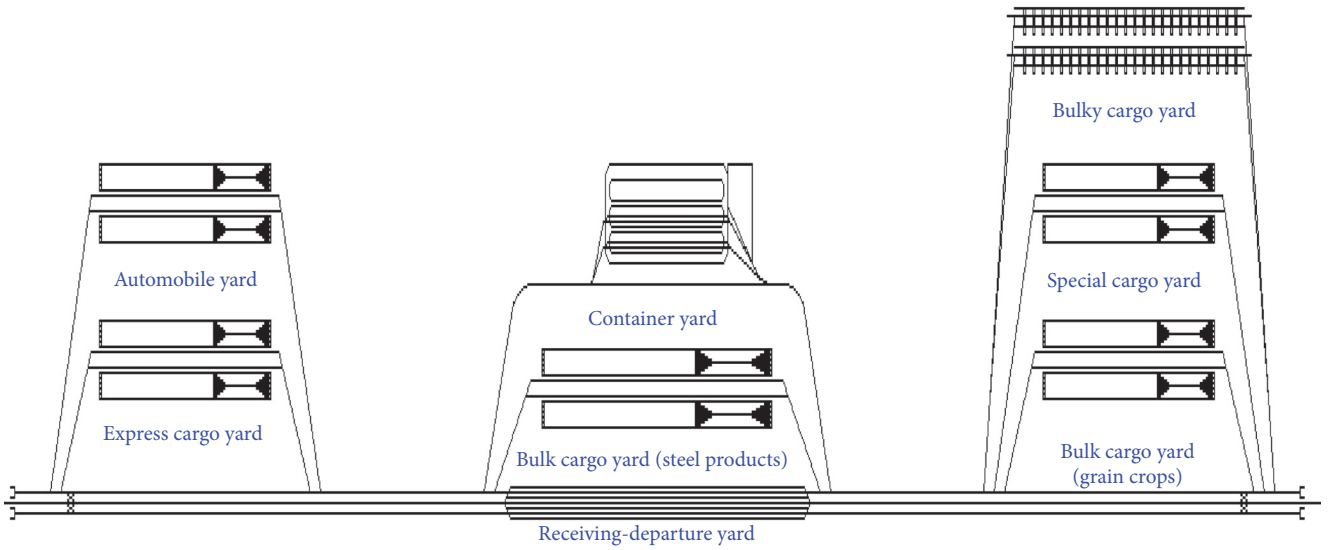


Figure 16: Yards layout of Qianchang railway intermodal terminal.

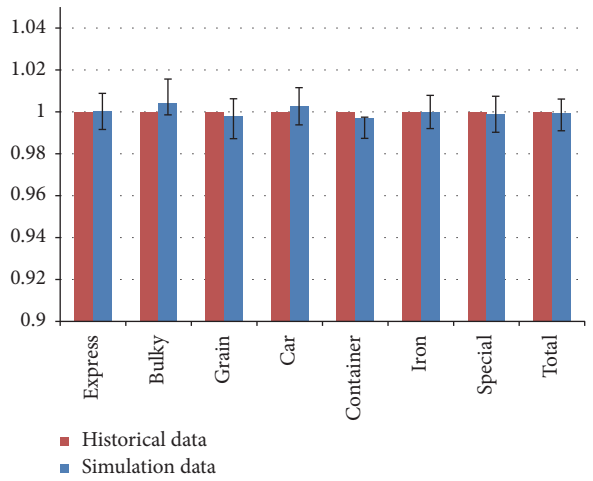


Figure 17: Normalized cargo volumes from historical and simulation datasets.

4 GB of RAM. To provide intuitive comparative analysis results in the validation phase, most simulated values are normalized with respect to the values in the historical dataset, which means the value of 1 corresponds to the actual value in the real life system.

In Figure 17, we present a bar plot of the normalized cargo volumes (the total volume and the respective volumes for each type of cargo) of the historical data and the simulation data which are in the 95% confidence interval. The black bars on histograms indicate the minimum and maximum of the simulation data. Although the trains and trucks are generated according to the historical record, the specific capacity of each railway wagon and truck and all types of delays (the delays that are related to the lack of railway tracks, handling equipment, and warehouses) introduce variabilities in the quantity of cargo volume. The differences between the simulated and actual data are very limited. The biggest difference of simulated data in 95% confidence interval is below 2%, which suggests good agreement with the historical dataset.

A similar study is performed in the case of utilization rates of handling machineries, as presented in Figure 18, where the bars also indicate the limits of the 95% confidence interval. The utilization rates of machineries in seven cargo yards are measured and displayed in the histograms, which vary widely from yard to yard. The utilization of handling equipment in the express cargo yard almost reached 50%, while the utilization rates in grain crops yard and special cargo yard are below 20%. Returning to the analysis of the validation, the similarity of handling machineries utilization is again noticeable. As shown in Figure 18, the differences between the simulated utilization rates and real utilization rates are below 2%, and 95% of the simulation results lie within 5% of the expected value, which means the simulation platform can reproduce the activities of handling machineries accurately.

Figures 19 and 20 display the normalized train loading/unloading time accumulated by each train from the historical and simulation dataset in the same order. In specific, each loading/unloading time is normalized with respect to the actual mean loading/unloading time which is 266.0 minutes. As shown in Figure 19, most data points are scattered in the interval from 0.3 to 2.0 and have no apparent patterns. However, a noticeable feature is that some data points are concentrated on the horizontal line with the value of 0.4 (as shown by the orange dotted line in Figure 19) and are relatively far from the rest of the points. These data points correspond to the container trains which can be loaded or unloaded with higher stevedoring efficiency with the support of gantry cranes. In fact, the historical average loading/unloading time of container trains is 100.3 minutes which is only 37.6% of the overall mean loading/unloading time.

These characteristics are all reflected in the simulated results as shown in Figure 20. To verify the similarities

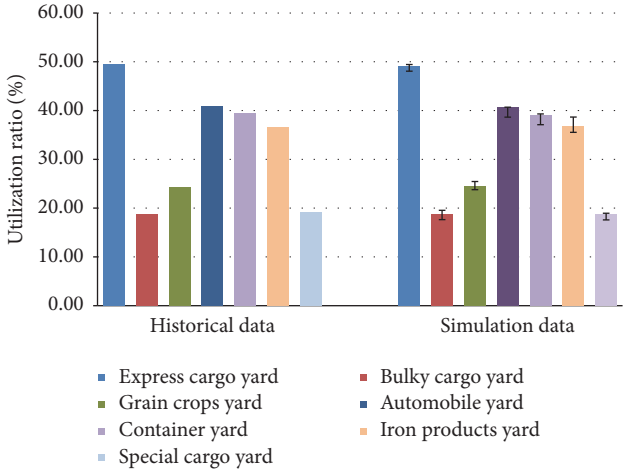


Figure 18: Historical and simulated utilization ratio of handling machineries.

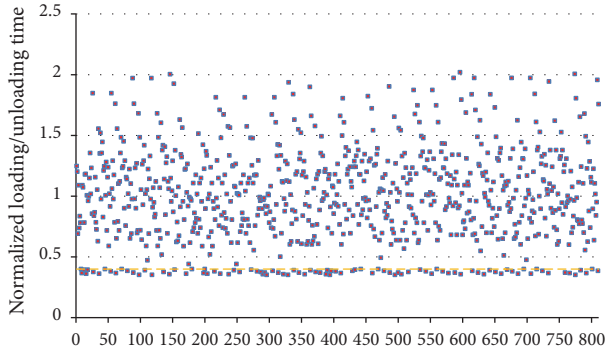


Figure 19: Historical train loading/unloading time distribution.

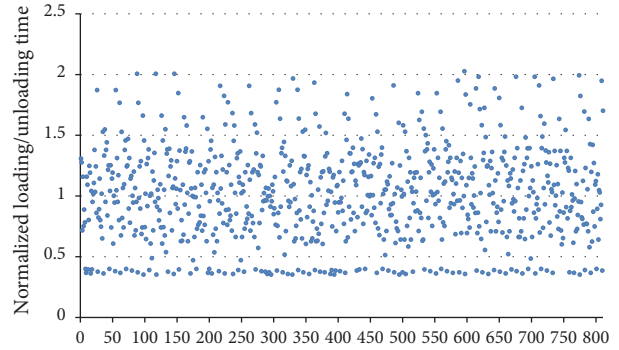


Figure 20: Simulated train loading/unloading time distribution.

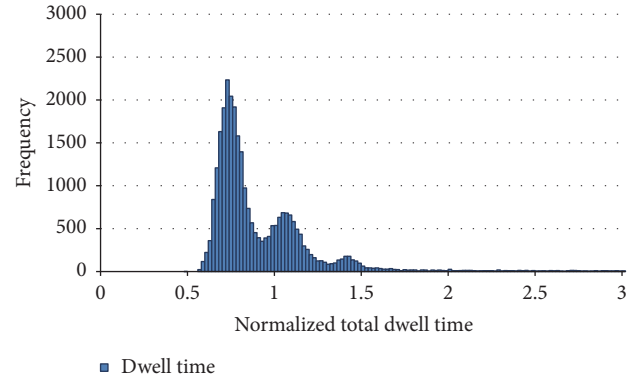


Figure 21: Historical distribution of total terminal dwell time of trucks.

between the historical data and simulation data in the future, we employ the two-sample Kolmogorov-Smirnov test to assess the quantitative differences between the distributions. The progressive significance coefficient (p value) is 0.553 which is significantly higher than the typical mark of 0.05. This verification indicates that the simulated train loading/unloading time distribution is statistically not different from the historical record and hence describes an accurate representation.

To verify the simulation accuracy of truck operation procedures, the total dwell time at the intermodal terminal is calculated as one of the key performance indicators. Considering the large quantity of trucks which are more than 30,000, it is difficult for the scatter plot to display the inherent law of the data intuitively. Therefore, histograms are used to describe the distribution of the total truck dwell time, which are shown in Figures 21 and 22. Similarly, the dwell time of each truck has been normalized with respect to the historical mean dwell time, and the simulated distribution is drawn based on the simulation results which are in the 95% confidence interval.

As shown in the histograms, the simulated dwell time distribution is very similar to the actual time distribution overall. Both distributions have multipeaks, and great majority of data points are clustered in the section from 50% to 150% of the mean dwell time. It is easy to explain this relatively strong variation in the dwell time distribution. In Qianchang intermodal terminal, there are multiple types of trucks. Due to the differences in cargo characteristics, the dwell time of different types of truck tends to have obvious differences. In fact, the historical mean dwell time of container trucks is 23% less than the average dwell time of special cargo trucks. In Figure 22, the red continuous vertical line indicates the value of the historical mean dwell time, and the green dotted line illustrates the average simulated dwell time within the 95% confidence interval, and the difference is negligible. The two distributions have been again compared in terms of the two-sample Kolmogorov-Smirnov test, with p value of 0.484, which is significantly larger than 0.05. The test result indicates that the simulation platform reproduces the activities of the real life system accurately.

In order to further verify the effectiveness of the simulation data, the mean error $D(Q)$ which is proposed in the literature [25] (defined in (2)) is introduced to quantify the mean absolute percentage error. In (2), Q is defined as a quantity of one indicator, Q_h is introduced as notation for the historical value of this indicator, and Q_s is defined as the value of the respective indicator obtained in the ith simulation. Furthermore, the percentage error between the simulated

Table 1: Simulation error coefficient results.

Cargo type	Error coefficients	Indicators			
		Cargo volume	Utilization rate	Loading/unloading time	Dwell time
Express cargo	$D(Q)$	0.012944	0.011589	0.018251	0.008871
	$E(Q)$	0.50%	0.94%	1.81%	1.75%
Bulky cargo	$D(Q)$	0.016761	0.035432	0.015348	0.015244
	$E(Q)$	0.65%	0.98%	1.53%	1.52%
Grain crops	$D(Q)$	0.020680	0.022395	0.021014	0.012596
	$E(Q)$	0.82%	0.70%	2.10%	1.26%
Automobile	$D(Q)$	0.011556	0.015392	0.000179	0.048468
	$E(Q)$	0.80%	0.59%	0.02%	3.83%
Container	$D(Q)$	0.011646	0.012187	0.000863	0.003889
	$E(Q)$	0.22%	1.01%	0.09%	1.53%
Iron products	$D(Q)$	0.012918	0.020601	0.014844	0.003483
	$E(Q)$	0.92%	1.62%	1.48%	1.88%
Special cargo	$D(Q)$	0.011461	0.019520	0.010384	0.034711
	$E(Q)$	0.69%	1.70%	1.04%	3.47%

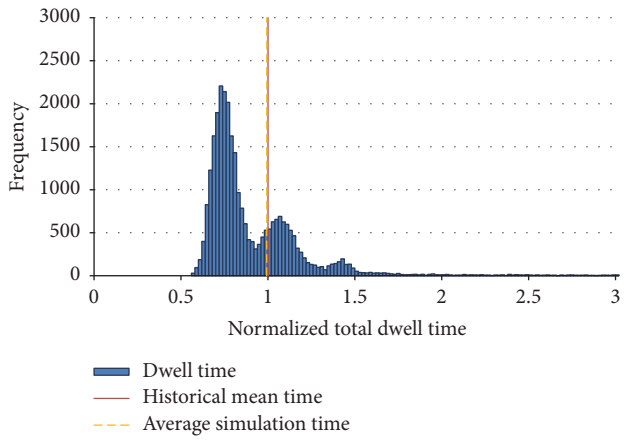


Figure 22: Simulated distribution of total terminal dwell time of trucks.

data and the historical value is described by $E(Q)$, which is defined in (3).

$$D(Q) = \frac{1}{n}\sum_{i=1}^{n}\frac{|Q_s(i) - Q_h|}{Q_h}, \quad (2)$$

$$E(Q) = \left|1 - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{Q_s(i)}{Q_s}\right)\right| \cdot 100. \quad (3)$$

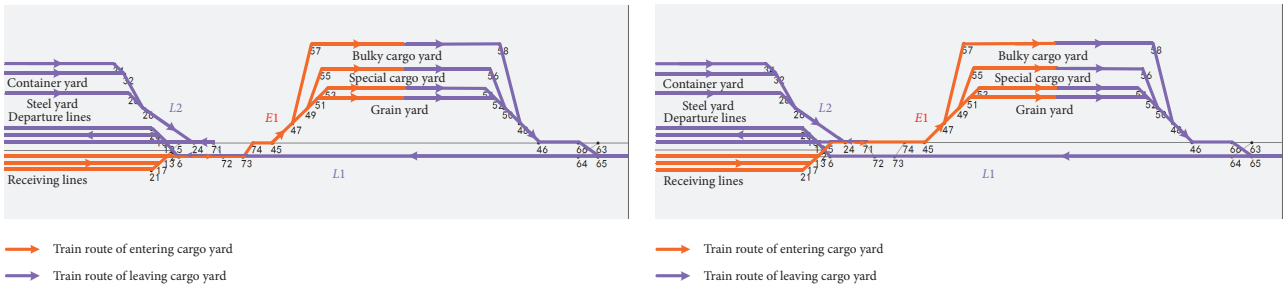
A detailed error coefficient analysis of the four key performance indicators is summarized in Table 1. The mean error $D(Q)$ and the percentage error $E(Q)$ of each type of cargo are calculated based on the historical record and the simulation data which are in the 95% confidence interval. The mean errors of the cargo volumes are obviously small, which is consistent with the results in Figure 17. The error coefficients of truck dwell times are relatively higher, which can be explained as follows. Since there is no fixed timetable for truck activities (which is different from trains), the trucks are designed to mainly follow the principle of FIFO rule in the simulation model, which is mentioned in Section 4.2. However, in practice, the queuing rule is more flexible and is easy to be artificially changed, which obviously affects the dwell time of trucks. As for trains, the arrival and departure time of trains are determined by the timetable, which reduces the variabilities in the simulation process to a certain extent. This deviation can be narrowed by introducing more flexible queuing rules into the simulation model in the future.

In general, the similarities between the simulation data and the actual record are noticeable. The biggest percentage error is smaller than 4%, and the percentage errors of most performance indicators are less than 2%. Following the comprehensive analysis of both qualitative and quantitative features, the simulated results have been proved to have good agreements with the historical values, which is able to verify the validity of this simulation platform.

6.3. Test Scenarios. Having established the convincing performance of the model via several validation studies, we now aim to use the platform as a decision-support and predictive tool. Some test scenarios of relevance for the activities in MYRIT have been formulated.

In Section 6.3.1, we describe the impact of train route arrangement on train operation efficiency. Two train route schemes are analyzed based on simulation results. Then, in Section 6.3.2, we pursue to investigate the effects of handling equipment configuration variation on loading/unloading operations. Some suggestions on equipment management are given according to the simulation analysis.

6.3.1. Train Route Arrangement. Train moving procedure is one of the most important terminal activities within MYRIT. This process has an instant impact on the train operations and can influence all relevant workflows of the terminal. Inside the MYRIT, each train must move in accordance with the prescribed train route. Due to the complexity of the railway



(a) Scheme A (b) Scheme B

Figure 23: Two local train routes' arrangement schemes of Qianchang railway terminal.

Table 2: Performance indicators of the train moving process based on simulation tests.

Train type	Scheme A		Scheme B	
	MDTRY (min)	MDTCY (min)	MDTRY (min)	MDTCY (min)
Bulky cargo	31.8	323.8	33.7	318.7
Grain	31.7	217.8	33.9	212.6
Container	31.9	428.2	34	422.5
Steel	32	120.8	33.1	130.7
Special cargo	117.4	392.4	121.2	422.5

line topology in MYRIT, train route arrangement can have a significant influence on train moving efficiency and is an important issue in the field of terminal management.

The goals of train route arrangement include reducing routes conflicts and improving train moving efficiency. Based on the simulation platform, the performance of certain train route arrangement can be analyzed, and different arrangement schemes can be compared and selected based on simulation results. In this section, we construct a simulation test which contains two partial train route arrangement schemes of Qianchang railway intermodal terminal. The detailed train route schemes are shown in Figures 23(a) and 23(b).

The two train route schemes shown in Figure 23 show part of the overall train route arrangement scheme of Qianchang terminal, which mainly contain the train routes of entering cargo yards and leaving cargo yards. According to the direction, the train routes which are shown in these two schemes can be divided into three types: the train routes of entering bulky, special, and grain cargo yards which are named $E1$, the train routes of leaving bulky, special, and grain cargo yards which are named $L1$, and the train routes of leaving container and steel cargo yards which are named $L2$. The detailed arrangements of $E1$ routes in Scheme A and Scheme B are different, while other train routes in the two schemes are the same. To investigate the influences of these two train route schemes on the efficiency of the train moving procedure, simulation experiments are implemented. With the exception of the train routes variation, all other parameters are designed in the same manner as in the validation study.

Table 2 summarizes the results of the investigation. Two indicators are used to analyze the train operation efficiency, which are the mean dwell time in receiving yard (MDTRY) and the mean dwell time in cargo yard (MDTCY). Each indicator is calculated based on the values from 500 simulation repetitions which are in the 95% confidence interval. Since the inspection time in receiving yard and the loading/unloading rates in cargo yards are the same in the two simulation experiments, the indicators of dwell time can reflect the performances of train moving efficiency. Figures 24 and 25 present the visualizations of the findings in Table 2, which show the value differences on indicators between Scheme A and Scheme B.

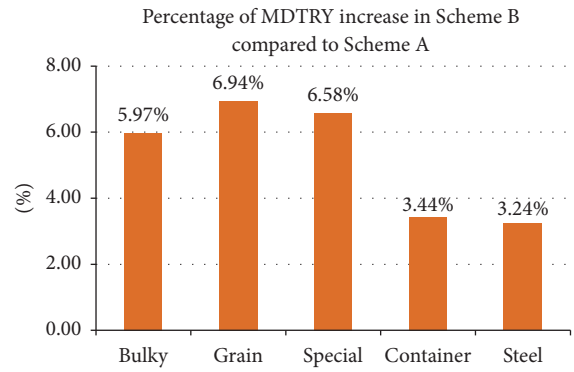


Figure 24: Differences of MDTRY between Scheme A and Scheme B.

We notice an obvious increase on MDTRYs when using Scheme B as shown in Figure 24. Compared with Scheme A, the MDTRYs of bulky, grain, and specially cargo trains have been extended by about 6%, which indicates that the routes $E1$ in Scheme B have more conflicts with other train routes. The train routes of entering cargo yard for container and steel cargo trains are the same in both Scheme A and Scheme B. However, the MDTRYs of container and steel cargo trains in Scheme B are still about 3% larger than the MDTRYs in

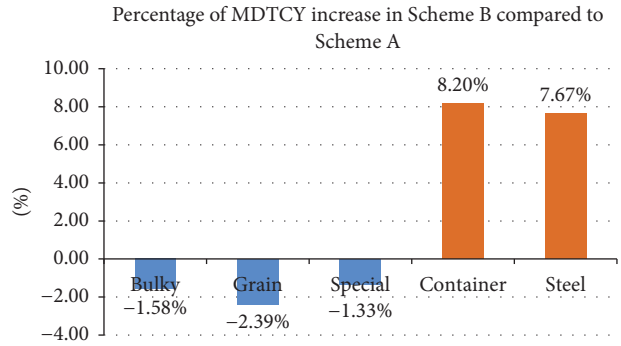


Figure 25: Differences of MDTCY between Scheme A and Scheme B.

Scheme A. In terms of dwell time in cargo yard, the MDTCYs of container and steel cargo trains in Scheme B have increased sharply compared with the MDTCYs in Scheme A, while the MDTCYs of bulky, grain, and special cargo trains have been reduced. This phenomenon indicates that, in Scheme B, the performance of $L1$ train routes has been improved compared to Scheme A; however, the conflicts between $L2$ and other train routes have become more serious.

The reason of the simulation results needs to be analyzed. There are train routes conflicts in both Scheme A and Scheme B. In Scheme A, routes $E1$ mainly have conflicts with routes $L1$, while in Scheme B, routes $E1$ mainly have conflicts with routes $L2$. However, the number of trains which need to occupy $L1$ and the number of trains that need to occupy $L2$ are different. In fact, the total number of container and steel cargo trains is 52% larger than the total number of bulky, grain, and special cargo trains, which makes the conflicts between $E1$ and $L2$ much more serious than the conflicts between $E1$ and $L1$. Hence, the MDTRY of bulky, grain, and special cargo trains and the MDTCY of container and steel cargo trains are all increased when using Scheme B. In addition, since the MDTCY of container and steel cargo trains increased, these types of trains need to wait more for available side tracks in the receiving yard, which leads to an increase in the MDTRY of container and steel cargo trains. Although the MDTCY of bulky, grain, and special cargo trains decreased when using Scheme B, the MDTCY of all trains in Scheme B is still 3.9% larger than that of Scheme A.

Therefore, in general, Scheme A has advantages over Scheme B in train operation efficiency. And in reality, Scheme A is adopted by the dispatch department of Qianchang terminal. The simulation tests show that the number of trains is one of the key factors in train route dispatching. It can be inferred that when the quantity of trains changes greatly, train routes schemes need to be adjusted accordingly.

6.3.2. Handling Equipment Configuration. The second set of investigations is constructed in order to study the impact of varying handling equipment configuration on performance indicators of loading/unloading operation. Generally, increasing the quantity of handling equipment can reduce the loading/unloading time and improve the operation efficiency. However, the quantitative relationship between the performance indicators of loading/unloading operation and the equipment quantity needs to be analyzed, which can provide decision-support information for the terminal managers.

Considering the realistic quantity limitations of the machineries, we are interested in what happens when the quantity of handling equipment is varied from 50% to 150% of the current value in Qianchang terminal. The increment is set to 25%, leading to a total of 5 studies, each designed to collect statistics from 500 simulations. The performance indicators include the mean laytime of train (MLT_Train), the mean laytime of truck (MLT_Truck), the mean waiting time for handling operation of train (MWT_Train), and the mean waiting time for handling operation of truck (MWT_Truck), which are calculated from the 95% confidence interval of the simulation results and are normalized with respect to the historical values. The results are shown in Figure 26, where we present the lower, upper, and mean values of the relevant indicators obtained from the simulation tests.

It can be observed that increasing the handling equipment quantity can lead to a decrease in laytime and waiting time of trains and trucks. However, the sensitivities of the four indicators are different.

As the machinery quantity increases, the decreases of MLT_Train and MLT_Truck are both almost linear, which is easily understood. However, the variation of MLT_Train is much bigger than the variation of MLT_Truck. This phenomenon is caused by the difference between the equipment usage pattern of trains and that of trucks. In general, most types of trains can be unloaded/loaded by multiple handling machineries at the same time. Therefore, with the increase of handling equipment, the loading/unloading rates of trains rise simultaneously. Nevertheless, except for bulk cargo trucks and express cargo trucks, many types of trucks (e.g., bulky cargo trucks, container trucks, and automobile trucks) can only be served by one handling machine due to the characteristics of goods. Thus, with the increase of handling equipment quantity, the growth of the overall loading/unloading rate of all types of trucks is more moderate compared with trains. And little variation of MLT_Truck can be observed with the increase of handling equipment.

We notice a strong link between the MWT_Train/MWT_Truck and machinery quantity dynamics. A 50% decrease of handling equipment can cause almost 130% increase of the waiting time of trains and almost 100% increase of the waiting time of trucks. Very large variation of waiting time indicates that reduction of available machineries can significantly reduce the efficiency of loading/unloading operation. Hence, maintaining an efficient stevedoring servicing is vital to the terminal system.

As the handling equipment quantity increases, both the means and variations in MWT_Train and MWT_Truck decrease obviously. However, the rates of decline gradually slow down. It can be inferred that, even with more handling machineries, the waiting time of trains and trucks in the terminal does not completely disappear. In fact, with another 50% increase of handling equipment (200% of the current value), only 8% decrease of MWT_Train and 5% decrease of MWT_Truck can be obtained. We can conclude that,

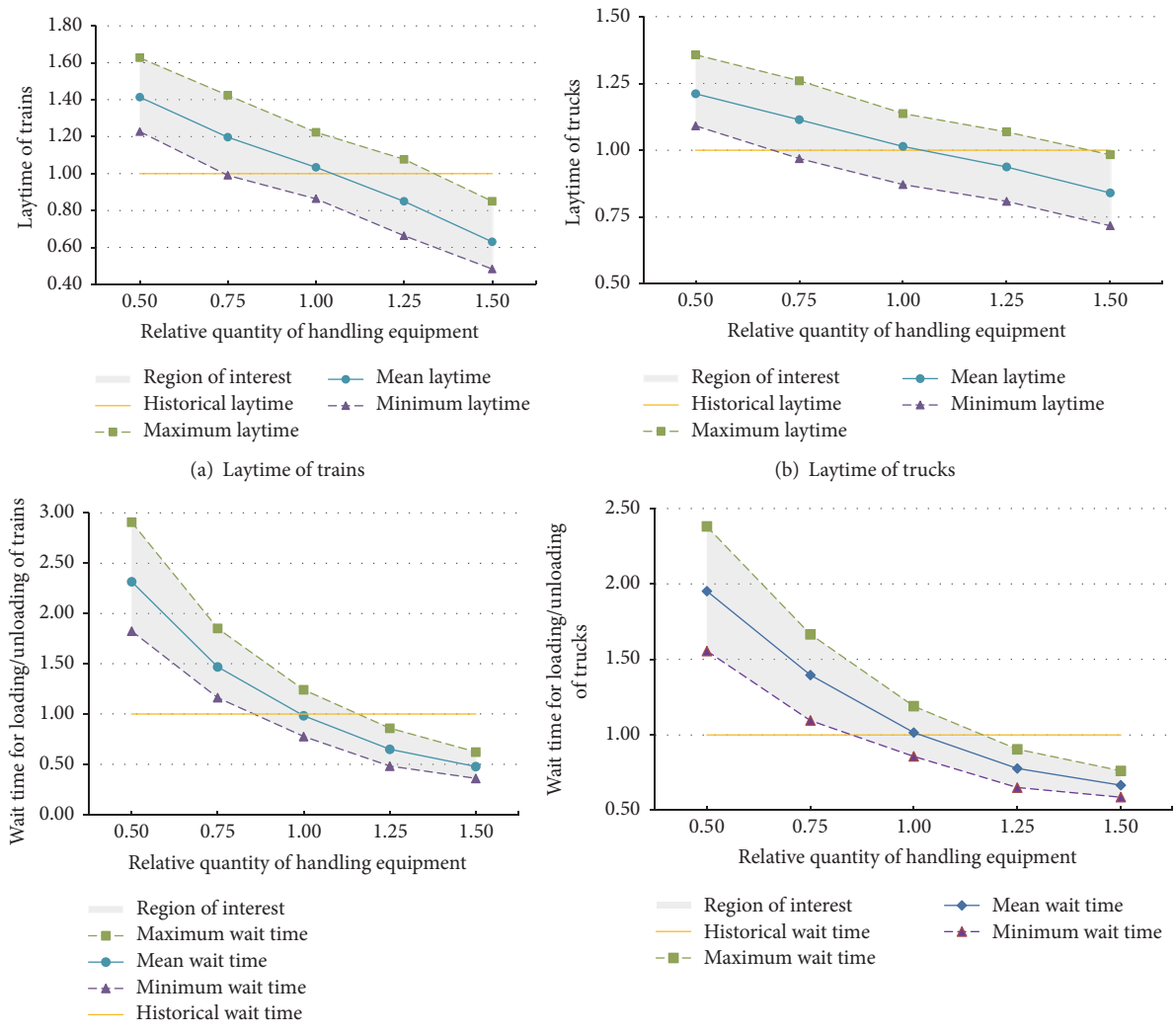


Figure 26: Impact of handling equipment configuration variation on loading/unloading operation efficiency.

in Qianchang intermodal terminal, increasing the handling equipment to more than 150% of the current value would only lead to negligible improvements of loading/unloading efficiency. With this conclusion, future investment and management of handling equipment can be considered in a reasonable and efficient manner.

7. Conclusions and Future Work

A simulation platform for providing a quantitative evaluation of MYRIT has been presented. The simulation model includes three sub-TPN models which correspond to the three major operations of MYRIT. In this platform, a yards and facilities layout module has been created where users can flexibly design or modify the terminal layout, and a TRDSM has been designed to provide an accurate simulation of the train moving process. Based on a comprehensive simulation analysis of Qianchang railway intermodal terminal, the simulation platform has been proved to be able to reproduce the activities of the real life system accurately. In addition, useful suggestions for terminal design and management can be obtained based on simulation tests of train route arrangement and handling equipment configuration.

Compared with existing simulation models, this platform has advantage in providing a detailed simulation of train moving process considering train routes dispatching rules. The results in the case study indicate that integrating the control method of train route dispatching can significantly improve the simulation accuracy of MYRIT. In general, this simulation platform can be used as an evaluation and analysis tool for terminal planning and design departments. Moreover, with the support of the yards and facilities module and the TRDSM, it can also be used as a management decision-support tool for dispatching managers.

Some limitations of the current simulation platform and the simplification procedure within the TPN model need to be analyzed, which can be considered as suggested improvements to the platform in future research.

Firstly, in the TPN model, the truck moving process has been simplified as one discrete event, and the connections between adjacent trucks and truck congestion circumstances are not considered, which will lead to inaccurate simulation results when the truck flow rate of the terminal is huge. This could be improved by integrating a car-following model into the simulation platform. Secondly, the basic queuing rule adopted by the platform is FIFO policy. Nevertheless, in some circumstances, the EDD (earliest due date) or WSPT (weighted shortest processing time first) rules can be used to reduce the delay time, and sometimes queuing rules are artificially determined. In most situations, this is a suitable assumption, since most of the terminal activities are required to follow the principle of FIFO rule. However, a study of mixed queue policy might provide possible improvement. Finally, an optimization mechanism is not taken into account in the framework, which can be improved. For example, numerous optimization methods have been proposed in the field of train route scheduling [26–28]. Some of the methods can be integrated into the platform to perform a simulation-based train route optimization for terminal managers. And in terms of determining equipment configuration or terminal layout, developing an integral approach considering multiple input scenarios can be a valuable future research direction, which can assist the terminal designers in a good terminal design.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is jointly supported financially by the National Natural Science Foundation of China (no. 61374202), National Key R&D Program of China (2016YFE0201700), the Research Project of China Railway Corporation (2017X004-D, 2017X004-E), and the Fundamental Research Funds for the Central Universities (2017YJS098).

References

[1] H. S. Mahmassani, K. Zhang, J. Dong, C.-C. Lu, V. C. Arcot, and E. Miller-Hooks, "Dynamic network simulation-assignment platform for multiproduct intermodal freight transportation analysis," *Transportation Research Record*, no. 2032, pp. 9–16, 2007.

[2] B. C. Kulick and J. T. Sawyer, "Use of simulation modeling for intermodal capacity assessment," in *Proceedings of the 2000 Winter Simulation Proceedings*, pp. 1164–1167, December 2000.

[3] S. Hou, "Distribution center logistics optimization based on simulation," *Research Journal of Applied Sciences, Engineering & Technology*, vol. 5, no. 21, pp. 5107–5111, 2013.

[4] A. E. Rizzoli, N. Fornara, and L. M. Gambardella, "A simulation tool for combined rail/road transport in intermodal terminals," *Mathematics and Computers in Simulation*, vol. 59, no. 1–3, pp. 57–71, 2002.

[5] T. Benna and M. Gronalt, "Generic Simulation for rail-road container Terminals," in *Proceedings of the Winter Simulation Conference*, pp. 2656–2660, Miami, Fla, USA, December 2008.

[6] A. Ballis and J. Golias, "Towards the improvement of a combined transport chain performance," *European Journal of Operational Research*, vol. 152, no. 2, pp. 420–436, 2004.

[7] M. Marinov and J. Viegas, "A simulation modelling methodology for evaluating flat-shunted yard operations," *Simulation Modelling Practice and Theory*, vol. 17, no. 6, pp. 1106–1129, 2009.

[8] A. Ballis and J. Golias, "Comparative evaluation of existing and innovative rail-road freight transport terminals," *Transportation Research Part A: Policy and Practice*, vol. 36, no. 7, pp. 593–611, 2002.

[9] M. K. Fugihara, A. Audenhove, and N. T. Karassawa, "Randomless as a critical point: Simulation fitting better planning of distribution centers," in *Proceedings of the Conference on Winter Simulation*, pp. 2371-2371, 2007.

[10] S. Hartmann, "Generating scenarios for simulation and optimization of container terminal logistics," *OR Spectrum*, vol. 26, no. 2, pp. 171–192, 2004.

[11] I. F. A. Vis, "Survey of research in the design and control of automated guided vehicle systems," *European Journal of Operational Research*, vol. 170, no. 3, pp. 677–709, 2006.

[12] A. Ballis and C. Abacoumkin, "A container terminal simulation model with animation capabilities," *Journal of Advanced Transportation*, vol. 30, no. 1, pp. 37–57, 1996.

[13] A. A. Shabayek and W. W. Yeung, "A simulation model for the Kwai Chung container terminals in Hong Kong," *European Journal of Operational Research*, vol. 140, no. 1, pp. 1–11, 2002.

[14] J. Montoya Francisco and R. K. Boel, "Introduction to Discrete Event Systems," *IEEE Transactions on Automatic Control*, vol. 46, no. 2, pp. 353-354, 2002.

[15] J. L. Peterson, "Petri net theory and the modeling of systems," *Computer Journal*, vol. 25, no. 1, 1981.

[16] R. David and H. Alla, *Discrete, Continuous, and Hybrid Petri Nets*, Springer-Verlag, Berlin Heidelberg, Germany, 2005.

[17] M. Dotoli, M. P. Fanti, A. M. Mangini, G. Stecco, and W. Ukovich, "The impact of ICT on intermodal transportation systems: a modelling approach by Petri nets," *Control Engineering Practice*, vol. 18, no. 8, pp. 893–903, 2010.

[18] G. Maione and M. Ottomanelli, "A Petri net model for simulation of container terminal operations," *Advanced OR and AI Methods in Transportation*, pp. 373–378, 2005.

[19] C. Lee, H. C. Huang, B. Liu, and Z. Xu, "Development of timed Colour Petri net simulation models for air cargo terminal operations," *Computers & Industrial Engineering*, vol. 51, no. 1, pp. 102–110, 2006.

[20] H.-P. Hsu, C.-N. Wang, C.-C. Chou, Y. Lee, and Y.-F. Wen, "Modeling and solving the three seaside operational problems using an object-oriented and timed predicate/transition net," *Applied Sciences (Switzerland)*, vol. 7, no. 3, article no. 218, 2017.

[21] G. Cavone, M. Dotoli, N. Epicoco, and C. Seatzu, "Intermodal terminal planning by Petri Nets and Data Envelopment Analysis," *Control Engineering Practice*, vol. 69, pp. 9–22, 2017.

[22] C. A. Silva, C. Guedes Soares, and J. P. Signoret, "Intermodal terminal cargo handling simulation using Petri nets with predicates," *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, vol. 229, no. 4, pp. 323–339, 2015.

[23] M. Dotoli, N. Epicoco, M. Falagario, and G. Cavone, "A Timed Petri Nets Model for Performance Evaluation of Intermodal Freight Transport Terminals," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 842–857, 2016.

[24] M. Bielli, A. Boulmakoul, and M. Rida, "Object oriented model for container terminal distributed simulation," *European Journal of Operational Research*, vol. 175, no. 3, pp. 1731–1751, 2006.

[25] R. Cimpeanu, M. T. Devine, and C. O'Brien, "A simulation model for the management and expansion of extended port terminal operations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 98, pp. 105–131, 2017.

[26] P. Pellegrini, G. Marlière, J. Rodriguez, and G. Marlière, "Optimal train routing and scheduling for managing traffic perturbations in complex junctions," *Transportation Research Part B: Methodological*, vol. 59, pp. 58–80, 2014.

[27] M. Samà, P. Pellegrini, A. D'Ariano, J. Rodriguez, and D. Pacciarelli, "Ant colony optimization for the real-time train routing selection problem," *Transportation Research Part B: Methodological*, vol. 85, pp. 89–108, 2016.

[28] M. Samà, A. D'Ariano, F. Corman, and D. Pacciarelli, "A variable neighbourhood search for fast train scheduling and routing during disturbed railway traffic situations," *Computers & Operations Research*, vol. 78, pp. 480–499, 2017.