Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 3513517, 21 pages
https://doi.org/10.1155/2023/3513517

*Research Article*

# Analytical Approximation-Based Approach for Passenger Flow Control Strategy in Oversaturated Urban Rail Transit Systems

**Qian Zhu, Xiaoning Zhu, Pan Shang, and Lingyun Meng**

*School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Pan Shang; shangpan@bjtu.edu.cn

Received 21 January 2023; Revised 3 June 2023; Accepted 5 June 2023; Published 23 June 2023

Academic Editor: Domokos Esztergár-Kiss



Focusing on a heavily congested urban rail corridor, this study investigates the passenger flow control strategy optimization problem from a mesoscopic perspective to reduce platform congestion and enhance service quality. Based on a quadratic functional approximation for passenger arrival rates, an analytical formula for calculating passenger waiting time is derived based on the classic deterministic queueing theory. We formulate the problem as a continuous nonlinear programming model to minimize the total passenger waiting time within transportation capacity constraints. A Lagrangian relaxation approach effectively transforms the original complex problem into an unconstrained minimization program. The analytical solution relating to optimal flow control strategy is derived by directly solving the unconstrained program. To further provide an integrated optimization framework from both the supply and demand sides, we extend the abovementioned passenger flow control optimization model into an integrated mixed-integer nonlinear programming model to jointly optimize the passenger-flow control strategy and train frequency setting. Numerical examples are presented to demonstrate the applicability and effectiveness of the proposed models. The computational results show that the produced high-quality passenger flow control strategy significantly reduces total passenger delay.

## 1. Introduction

Peak-hour traffic congestion in megacities is a very common but critical problem when travel demand temporally and spatially exceeds transportation capacity. Urban rail transit currently faces significant pressure due to its higher punctuality, larger capacity, eco-friendliness, and popularity among travelers, which have led to increased demand and heightened expectations. The overloaded passenger flow not only aggravates potential accident risks at platforms but also increases the likelihood of train delays. In urban rail transit operation and management, passenger travel demands cannot be satisfied, even with the minimum departure headway during the peak period. The severe imbalance between travel demand and transportation resources urgently necessitates accurate passenger management methods from the demand side.

Recently, passenger flow control has received significant attention in the practical operations and management of rail transit. Collaborative passenger flow control refers to a coordinated and synchronized passenger management method within urban rail systems. Instead of following a first-in-first-service principle, it requires passengers with different origins and destinations to board trains in a manner that enhances operational efficiency or improves service levels. With collaborative passenger flow control strategies, passengers arriving at origin stations are required to wait in the station hall and then enter the platform through the entrance facilities, following the flow control rates. However, current passenger flow control methods in practical operations, such as the Beijing subway system, only focus on one or two stations and fail to consider the competition between passengers at different stations. Therefore, developing an approach to quickly compute collaborative passenger flow strategy within transportation capacity is essential.

This study explores the optimization of collaborative passenger flow control from a systemic perspective, aiming to enhance the operational efficiency and safety of an

oversaturated urban rail corridor. The optimal inflow rates of each origin-destination (OD) pair are systematically computed, considering the transportation capacity given the predetermined train timetable. Accordingly, under oversaturated conditions, the limited transportation resources can be efficiently reallocated by considering the temporal and spatial distribution characteristics of passenger flow. We summarized the following contributions of this work:

(1) By formulating passenger arrival rates using a quadratic functional approximation method, a continuous nonlinear programming model with the objective of minimizing total passenger waiting time is developed to optimize passenger flow control strategy.

(2) A Lagrangian relaxation solution approach is designed to solve the constructed model. With this solution procedure, an analytical solution can be derived to quickly generate the flow control strategies among all OD pairs.

(3) To further design the passenger flow control strategy and train frequency synchronously, a nonlinear mixed-integer programming model is proposed to consider the complicated interactions between supply and demand sides systematically.

(4) Numerical examples involving a three-node network and Beijing Metro Line 1 demonstrate that the formulated model can produce passenger flow control strategies with less passenger delay compared with the scenario without flow control.

The paper is organized as follows. We review the relevant studies in Section 2. In Section 3, the passenger flow control problem for an oversaturated rail corridor is described. In Section 4, the collaborative passenger flow control optimization model is developed, and the analytical solution of the model is investigated. In Section 5, a comprehensive mixed-integer programming model is rigorously constructed to design the optimal passenger flow control and frequency setting simultaneously. In Section 6, a simple case and large-scale instance on Beijing Metro Line 1 are conducted to evaluate the developed models. Finally, in Section 7, the conclusions and future research directions are presented.

## 2. Literature Review

With the rapid increase in travel demand in recent decades, the oversaturation situation in heavily congested rail corridors has led to massive passenger delays and increased accident risks, becoming a significant concern [1–10]. To address this issue, researchers have focused on optimizing operational efficiency and service quality in rail transit systems, with passenger waiting time as a key evaluation index [9, 11–16]. Along this line, some researchers have proposed various train operation and management strategies from the supply side, such as frequency design, timetabling adjustment, and scheduling optimization. For instance, LeBlanc [17] built a mode-split assignment model to determine the frequencies of transit lines in a network and further described the standard Hooke–Jeeves algorithm. Based on dynamic trip records collected from automatic fare collection systems, Niu and Zhou [13] designed a timetabling optimization approach for oversaturated railway lines. A nonlinear binary programming model was developed and solved using a genetic algorithm (GA). Barrena et al. [1] formulated two mathematical optimization models to optimize timetables and developed an adaptive neighborhood search heuristic approach to solve the cases. Yin et al. [14] focused on the metro train rescheduling problem, taking into consideration uncertain and time-dependent travel pattern. They developed a stochastic programming model to reduce passenger waiting time and train operation costs and adopted an approximated dynamic programming algorithm to solve this problem.

Although service-oriented train scheduling can significantly improve operational efficiency and reduce passenger waiting time, the severe passenger congestion on platforms still cannot be avoided due to the great attraction of rail transit systems. Several studies have demonstrated that reasonable demand management measures can efficiently reduce congestion, such as fare pricing schemes [18–21], reservations [22–25], and passenger flow control [26–29].

Passenger flow control, which considers boarding activities, has received considerable attention from researchers. It has been proven to be an effective method for reducing passenger waiting time [12, 18, 26, 27, 29–31]. For instance, Xu et al. [29] developed multiobjective mathematical programming formulations to adjust the inbound and transfer passenger arrival rates among multiple stations, in which the route choice behavior was considered through a logit-based stochastic equilibrium assignment model. Shi et al. [18] considered passenger transfer behaviors in a metro network and constructed an integer linear programming model to optimize the passenger flow control strategy. Wang et al. [32] established a mixed-integer programming model to realize coordinated and dynamic flow control considering transfer passengers. Zhou et al. [31] focused on coordinated passenger flow control from the line, station, and ticket gate dimensions by establishing a linear programming model. Meng et al. [26] studied a robust passenger flow control strategy that takes into account stochastic and dynamic passenger demand. They designed a heuristic algorithm to solve the formulated integer linear programming model, which aimed to minimize the total passenger waiting time over a metro line.

In addition, other researchers combined passenger flow control problems with train timetable optimization in a single rail line without considering transfer behaviors. For example, Shi et al. [27] developed an effective method to design train timetables and passenger flow control strategies jointly given dynamic travel demand. They presented an integrated integer linear programming model and implemented a hybrid algorithm that combined local search and CPLEX to reduce the total waiting time. Liu et al. [33] formulated an integrated nonlinear programming model to collaboratively optimize passenger flow control strategy, train schedules, and connections. A Lagrangian relaxation-based algorithm was designed to solve this problem. Yuan

et al. [34] introduced a novel mixed-integer linear programming model to jointly optimize train scheduling and passenger flow control strategies. They used ILOG CPLEX to solve a real-world instance of the Beijing Metro Line 5.

Table 1 presents a comparison of modeling approaches in related publications, with key features including decision item, model, algorithm, analytical solution, and study granularity. It is important to emphasize that the presence of an analytical solution is considered a critical criterion. This is because an analytical solution can quickly offer precise and insightful passenger flow control strategies for certain real-world scenarios while requiring fewer computational resources. Furthermore, it provides a deeper understanding of how the passenger flow strategy changes in response to relevant parameters, as demonstrated in Section 4.3.

In the existing literature, numerous studies have examined flow control strategies for congested rail corridors using a microscopic perspective. These studies often model boarding and alighting activities and discretize the time horizon into finite intervals to estimate passenger waiting times. However, the real-time control of large-scale rail lines requires a more practical and efficient analytical method for generating solutions. Furthermore, recent research has not thoroughly investigated the integration of passenger flow control and train frequency optimization. To address these gaps, our study explores a collaborative passenger flow control problem from a mesoscopic perspective in oversaturated rail lines and introduces an integrated optimization framework that jointly designs passenger flow control strategies and train frequency settings to enhance system-wide service levels.

## 3. Problem Statement

This study considers a passenger flow control strategy for a directional urban rail corridor with a set of consecutively numbered links, as shown in Figure 1. The links are numbered as $l \in \{1,2,\ldots,L\}$, where all trains depart from the start terminal station and arrive at the return terminal station, following an all-stopping pattern during peak periods. We assume that all trains maintain the same velocity and dwell time, leading to their space-time trajectories being parallel to each other. Let $K$ denote the set of OD pairs, where each OD pair $k \in K$ can be connected by one path along the rail corridor. Furthermore, let $\delta_k^l$ denote the link-OD incidence parameter, which equals 1 if link $l$ is on the path of OD pair $k$ and 0 otherwise. To consider the time-dependent continuous passenger demand over the morning or evening peak period $[0,T]$, we denote the number of passengers for the $k$th OD pair arriving at the origin station at time $t$ as $\lambda_k(t)$. These data can be conveniently collected from an automatic fare collection system typically employed in an urban rail system.

Many passengers are left behind on platforms if they are unable to board the arriving train. This accumulation of passengers can potentially cause severe congestion or operational safety risks due to the limited capacity of the platforms. Furthermore, this noncollaborative boarding pattern may cause severe delays for passengers and hence poses a huge challenge to the transport efficiency and service level of the rail transit system. Considering the analyses mentioned above, collaborative passenger flow control can significantly improve rail transit system performance and has received considerable attention from many researchers and practitioners.

In undersaturated conditions, passengers queuing at platforms can always get on the next incoming train. Therefore, controlling the inflow rate of passengers into platforms is redundant. Under oversaturated conditions, we divide the station area into two parts: the waiting area in the hall and the boarding area at the platforms, as depicted in Figure 2(a). For simplicity, we will illustrate the passenger flow control strategy and corresponding queuing evolution process using a single OD pair; consequently, we can omit the OD pair subscription $k$. We will use $\lambda(t)$ to represent the passenger arrival rate at time $t$.

We consider each entrance facility as a traffic bottleneck; therefore, the passenger queuing evolution process for each OD pair can be analyzed based on Newell's deterministic fluid approximation model [35]. Accordingly, the time-dependent passenger arrival rate $\lambda(t)$ is assumed to follow a polynomial trend and is shown in Figure 2(c). The cumulative arrival count by time $t$ is denoted as $A(t)$, $A(t) = \int_0^T \lambda(t)$, which increases over time, as shown in Figure 2(b). Passengers who have just arrived must first queue in the station hall due to limited platform capacity. They then gradually enter the platform to board trains, in accordance with the control service rate $\mu(t)$, through entrance facilities. Similarly, the cumulative departure count is denoted by $D(t)$, $D(t) = \int_0^T \mu(t)$. The time-dependent number of passengers $Q(t)$ queuing in the waiting area is the vertical length between $A(t)$ and $D(t)$, and it reaches a maximum at time $t_2$ when $\lambda(t) = \mu(t)$. Following Newell's deterministic fluid approximation model [35], Appendix A presents the derivation process for the time-dependent queue length and waiting time. For modeling convenience, we ignore the walking time between the waiting and boarding areas. This implies that passengers can arrive at the platform immediately after passing through the entrance facilities.

To rigorously formulate the collaborative passenger flow control optimization model, this study makes the following assumptions.

*Assumption 1.* Trains maintain the same speed and dwell time throughout the study horizon. Overtaking and crossing events are strictly prohibited, meaning their space-time trajectories remain identical and parallel.

*Assumption 2.* The boarding sequence of waiting passengers with the same OD pair follows the exact first-in-first-out principle. Otherwise, estimating the total waiting times of passengers would be extremely complicated due to the difficulty in recognizing the boarding order within the metro network.

*Assumption 3.* In-service trains can always satisfy all passenger demands over the considered time horizon, which implies that no passengers are left behind at the end of the operation time horizon.

Table 1: Comparisons of passenger flow control related modeling methods in different publications.

| Publication | Decision item | Model | Solution approach | Analytical solution | Microscopic or mesoscopic |
|---|---|---|---|---|---|
| Xu et al. [29] | Passenger flow control | NLP | Simulation-based algorithm | No | Microscopic |
| Shi et al. [18] | Passenger flow control | ILP | CPLEX solver | No | Microscopic |
| Wang et al. [32] | Passenger flow control | MIP | GUROBI solver | No | Microscopic |
| Zhou et al. [31] | Passenger flow control | LP | Particle swarm optimization algorithm | No | Microscopic |
| Meng et al. [26] | Passenger flow control | ILP | Heuristic algorithm | No | Microscopic |
| Shi et al. [27] | Train timetable, and passenger flow control | ILP | Hybrid algorithm | No | Microscopic |
| Liu et al. [33] | Train schedule, train circulation, and passenger flow control | MINLP | Lagrangian relaxation | No | Microscopic |
| Yuan et al. [34] | Train schedule, and passenger flow control | MINLP | CPLEX solver | No | Microscopic |
| This study | Train frequency, and passenger flow control | NLP | Lagrangian relaxation | Yes | Mesoscopic |

*Note*. NLP: nonlinear programming; ILP: integer linear programming; MIP: mixed-integer programming; LP: linear programming; MINLP: mixed-integer nonlinear programming.

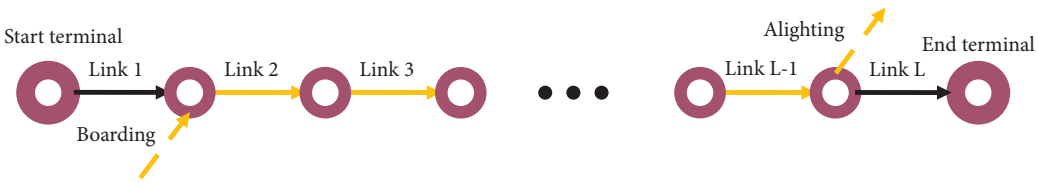



Figure 1: Urban rail transit line.

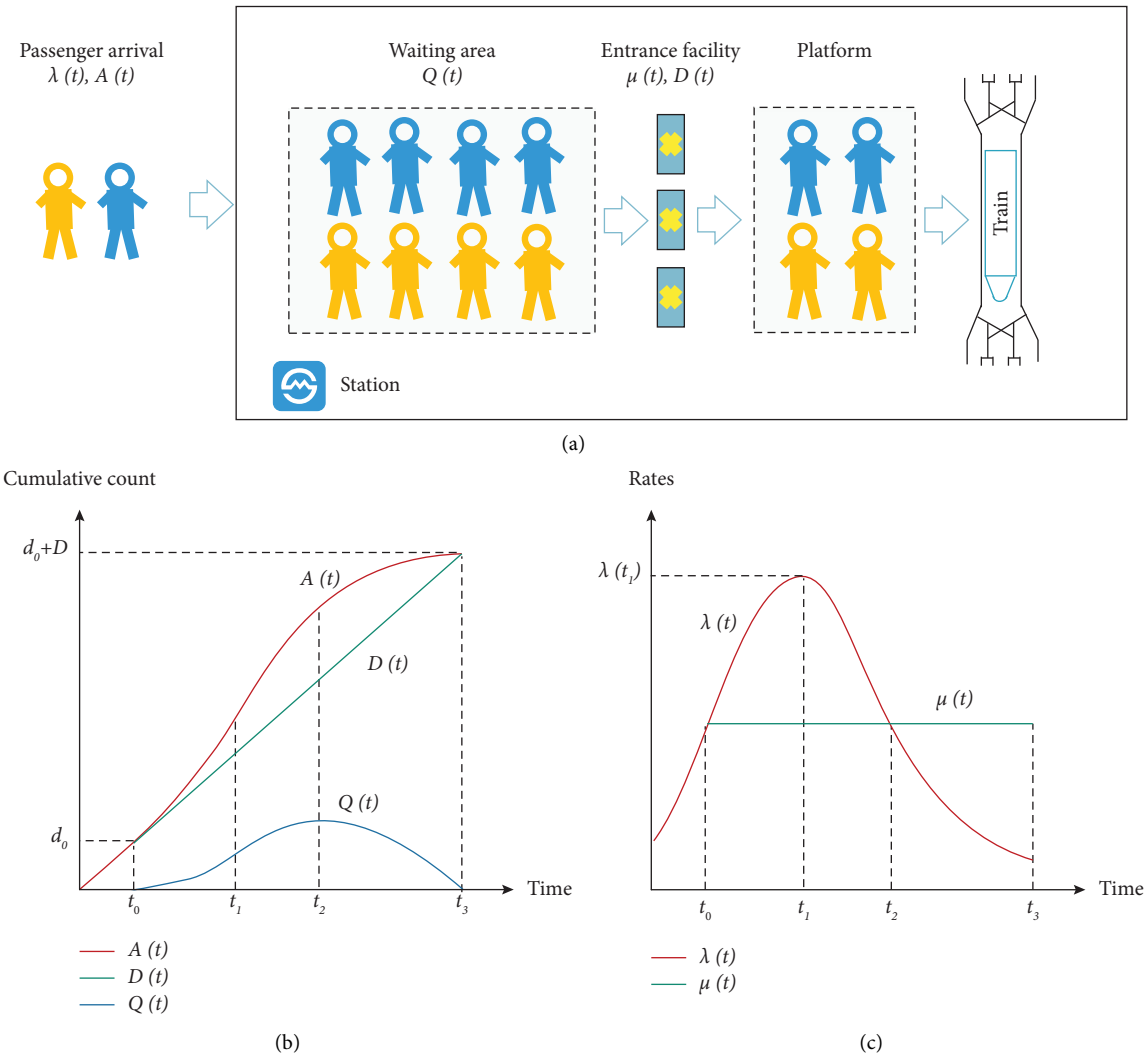



Figure 2: Illustration of passenger flow control process. (a) Passenger flow control process at station. (b) Cumulative arrival and departure counts. (c) Time-dependent arrival and departure rates.

*Assumption 4.* To distinguish passengers arriving at stations according to their OD trip information, the entrance facilities are preset for each destination, as illustrated by Shi et al. [18].

Given (1) the urban rail transit line, (2) the timetable of train services, and (3) time-dependent passenger demand, this study seeks to develop an analytical collaborative passenger flow control strategy under oversaturated conditions to reduce the total passenger delay from a system optimization perspective. As a result, the core decision variables of the investigated problem are service rates for all OD pairs at their origin stations, which are strictly restricted to the train service capacity. Notably, the proposed model can be applied in both directions when we consider the passenger travel demand as input data. However, because passenger congestion always occurs in one direction during a fixed period (i.e., morning and evening peaks), we model this problem on a one-way metro line.

This study disregards the integer nature of time-dependent passenger arrival counts and treats their arrival rate as a continuous fluid. Figure 3 presents the cumulative number of arriving and departure passengers by time $t$. The actual cumulative number of departure passengers $D'(t)$ is a step function with integer steps at each train departure time, which is represented by the yellow step line in Figure 3. Because this study assumes that trains are dispatched at discrete time intervals with a minimum constant headway, the service rate $\mu(t)$ would be a constant parameter under this constant headway. This equals the ratio of vehicle capacity $c$ to the headway $h$. As shown in Figure 3, we can obtain a straight line $D(t)$ by connecting the peak points of the steps of $D'(t)$. The total passenger waiting time $W$ is defined as the area between the arrival and departure curves. Notably, $W$ calculated using the approximate cumulative departure curve $D(t)$ is less than that obtained directly using the step departure curve $D'(t)$. However, the optimization direction of the flow control optimization using the approximation method is consistent with the original problem owing to the constant headway. This approximate method is convenient for using an analytical method to determine the optimal passenger flow control strategy in urban rail corridors.

To clearly characterize the problem of interest, Table 2 lists the notations used in the optimization models.

## 4. Passenger Flow Control Strategy with a Queueing Theoretic Performance Model

In this section, we systematically develop a passenger flow control strategy optimization model to minimize total passenger delay, subject to train service capacity. Based on Newell's deterministic queueing theory [35], an analytical solution of optimal flow control strategy is derived by introducing a Lagrangian relaxation solution approach.

*4.1. Collaborative Passenger Flow Control Strategy Optimization Model.* The passenger flow control strategy executed at stations relates to the system-level spatial redistribution of capacity along an urban rail corridor. To maximize system efficiency, the proposed collaborative passenger flow control (CPFC) optimization model aims at reducing the total passenger delay from a system optimization perspective, subject to the transportation capacity of each link on the rail corridor. The objective function, as formally stated in equation (1), aims to maximize the utility of the passenger flow control strategy. This study uses the total waiting time as the performance evaluation criterion for the proposed passenger flow control approach. Therefore, maximizing the utility of flow control is equivalent to maximizing the negative sum of the passenger waiting times over all OD pairs. The total delay $W_k$ for each OD pair $k$ in equation (1) is reflected in the flow control strategy $\mu_k$.

$$\max_{\mu_k} Z(\boldsymbol{\mu}) = -\sum_{k=1}^{K} W_k(\mu_k), \tag{1}$$

subject to the following constraints.

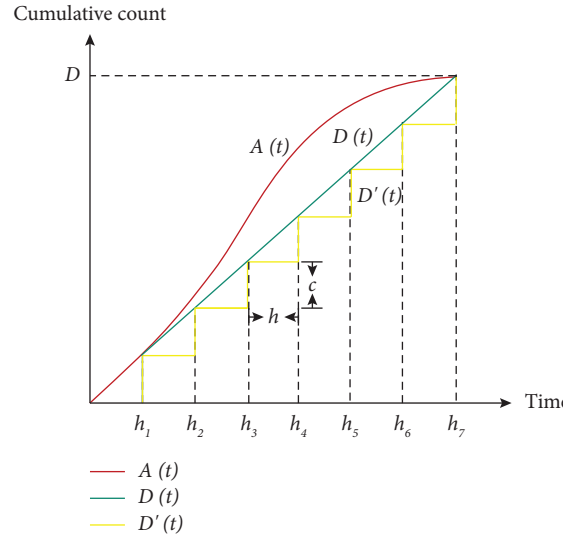



Figure 3: Relationship between approximate continuous method and original discrete problem.

Table 2: Notations.

| Symbols | Definition |
|---|---|
| $L$ | Set of links |
| $K$ | Set of OD pairs |
| $l$ | Index of links, $l \in L$ |
| $k$ | Index of OD pairs, $k \in K$ |
| $t_{1,k}$ | Time index with maximum arrival rate of $k$th OD pair |
| $\lambda_k(t)$ | Arrival rate of $k$th OD pair at time $t$ |
| $\mu_k$ | Service rate of $k$th OD pair |
| $\beta_k$ | Parameter in waiting time function, which equals to $-1/2\lambda_k''(t_{1,k})$ |
| $\delta_k^l$ | Link-OD incidence, = 1, indicates link $l$ is on the path of $kth$ OD pair, =0, otherwise |
| $W_k$ | Total delay of $k$th OD pair |
| $\mathrm{Cap}_l$ | Capacity of $l$th link |
| $\rho$ | Reciprocal of unit time waiting cost per passenger |
| $H_{\max}$ | Maximum permitted headway |
| $H_{\min}$ | Minimum permitted headway |
| $n_1$ | The number of train carriages of each train |
| $U_1$ | Energy cost of each train |
| $N$ | Total number of available train carriages |
| $T$ | The length of studied time horizon |
| **Variables** | **Definition** |
| $u_k$ | Departure rate of OD pair $(r, s)$ at time $t$ |
| $f$ | Train frequency during the studied time horizon |

*4.1.1. Capacity Constraint*

$$\sum_{k=1}^{K} \mu_k \delta_k^l \le \mathrm{Cap}_l, \quad \forall l \in L. \tag{2}$$

*4.1.2. Positive Variables*

$$\mu_k \ge 0, \quad \forall k \in K. \tag{3}$$

This study introduces a novel analytical formulation of passenger waiting time based on a quadratic approximation of the passenger arrival rate at the stations. Equation (A.13)

presents the functional form of the total delay $W_k$ for each OD pair during peak hours, with the detailed derivation process outlined in Appendix A. The total delay for each OD pair, $W_k, \forall k \in K$, depends on the assigned service rate $\mu_k$ with the assumption of a quadratic time-dependent arrival rate $\lambda_k(t)$. Consequently, by substituting equation (A.13) into equation (1), we derive the following system-level total delay function:

$$\max_{\mu_k} Z(\boldsymbol{\mu}) = -\sum_{k=1}^{K} \frac{9\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]^2}{4\beta_k}. \tag{4}$$

*4.1.3. Boundary Constraints.* According to practical experience, a delay occurs only when a queue exists. Therefore, equation (5) is proposed to restrict the upper bound for the service rate. Otherwise, if no queue exists, equation (4) indicates that there will be no delay, i.e., $\mu_k = \lambda_k(t_{1,k})$.

$$\mu_k \le \lambda_k\left(t_{1,k}\right), \quad \forall k \in K. \tag{5}$$

To ensure that the peak period of each OD pair is fully included in the entire study horizon $[0, T]$, the following boundary constraints are required:

$$\mu_k \ge \lambda_k(t_0), \quad \forall k \in K, \tag{6}$$

$$\mu_k \ge \lambda_k(t_T), \quad \forall k \in K. \tag{7}$$

Equation (6) suggests that the congestion period for each OD pair has not yet started at the start time $t_0$ while equation (7) indicates that it vanishes at the ending time $t_T$. The corresponding time interval $[t_0, t_T]$ relates to the horizon $[0, T]$. Moreover, the capacity constraint presented in equation (2) should also be incorporated into the optimization model.

**Lemma 5.** *CPFC model is concave and has a unique solution that maximizes the passenger flow control utility, as expressed in equation (1).*

*Proof.* First, we must prove that the objective function in equation (4) of the CPFC model is concave, which is equal to proving that $\sum_{k=1}^{K}(9[\lambda_k(t_{1,k}) - \mu_k]^2)/(4\beta_k)$ is convex. For convenience, we let $F = \sum_{k=1}^{K}(9[\lambda_k(t_{1,k}) - \mu_k]^2)/(4\beta_k)$. The Hessian matrix of $F$ is given as follows:

$$\begin{bmatrix} \frac{\partial^2 F}{\partial \mu_1^2} & \frac{\partial^2 F}{\partial \mu_1 \partial \mu_2} & \cdots & \frac{\partial^2 F}{\partial \mu_1 \partial \mu_K} \\ \frac{\partial^2 F}{\partial \mu_2 \partial \mu_1} & \frac{\partial^2 F}{\partial \mu_2^2} & \cdots & \frac{\partial^2 F}{\partial \mu_2 \partial \mu_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial \mu_K \partial \mu_1} & \frac{\partial^2 F}{\partial \mu_K \partial \mu_2} & \cdots & \frac{\partial^2 F}{\partial \mu_K^2} \end{bmatrix} = \begin{bmatrix} \frac{9}{2\beta_1} & 0 & \cdots & 0 \\ 0 & \frac{9}{2\beta_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{9}{2\beta_K} \end{bmatrix}, \tag{8}$$

where $\beta_k$ denotes a positive parameter. Evidently, the Hessian matrix mentioned above is definite, implying that the function $F$ is convex. This in turn means that the objective function of the CPFC model is concave. Furthermore, the linear capacity constraint in equation (2) leads to a convex feasible region. Therefore, the CPFC model is convex and has a unique optimal solution, thus proving Lemma 5. □

*4.2. Analytical Solution Analysis.* A Lagrangian relaxation solution approach is designed to derive an analytical solution for the investigated problem. In this section, only the objective function in equation (4) and capacity constraint (2) are considered. Let $p_l$ be the Lagrange multiplier associated with the capacity constraint at link $l$, and let $\mathbf{p}$ denote the vector of Lagrange multipliers. Subsequently, the Lagrangian form of equation (4) is expressed as follows:

$$L(\boldsymbol{\mu}, \mathbf{p}) = -\sum_{k=1}^{K} \frac{9\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]^2}{4\beta_k} - \sum_{l=1}^{L} p_l \left( \sum_{k=1}^{K} \mu_k \delta_k^l - \mathrm{Cap}_l \right). \tag{9}$$

Setting $(\partial L/\partial \mu_k) = 0$ for each OD pair $k$ gives

$$\begin{aligned} \frac{\partial L}{\partial \mu_k} &= \frac{9\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]}{2\beta_k} - \sum_{l=1}^{L} p_l \delta_k^l \\ &= 0. \end{aligned} \tag{10}$$

Consequently, we can obtain the following analytical solution:

$$\mu_k = \lambda_k\left(t_{1,k}\right) - \frac{2\beta_k}{9}\sum_{l=1}^{L} p_l\delta_k^l, \quad \forall k \in K. \tag{11}$$

Furthermore, the KKT conditions require that

$$p_l\left(\sum_{k=1}^{K}\mu_k\delta_k^l - \mathrm{Cap}_l\right) = 0 \text{ and } p_l \geq 0, \quad \forall l \in L. \tag{12}$$

If $p_l = 0$, the above formula is insignificant.

Otherwise, $p_l > 0$, $\sum_{k=1}^{K}\mu_k\delta_k^l - \mathrm{Cap}_l = 0$, and substituting equation (11) into equation (12) yields

$$\sum_{k=1}^{K}\left(\lambda_k\left(t_{1,k}\right) - \frac{2\beta_k}{9}\sum_{l'=1}^{L} p_{l'}\delta_k^{l'}\right)\delta_k^l - \mathrm{Cap}_l = 0, \quad \forall l \in L. \tag{13}$$

Then,

$$\sum_{k=1}^{K}\frac{2\beta_k}{9}\delta_k^l\sum_{l'=1}^{L} p_{l'}\delta_k^{l'} = \sum_{k=1}^{K}\lambda_k\left(t_{1,k}\right)\delta_k^l - \mathrm{Cap}_l, \quad \forall l \in L. \tag{14}$$

For simplicity, we use $\lambda_k$ as a substitute for $\lambda_k(t_{1,k})$. Let

$$b_l = \sum_{k=1}^{K}\lambda_k\left(t_{1,k}\right)\delta_k^l - \mathrm{Cap}_l. \tag{15}$$

Then, equation (14) can be rewritten as follows:

$$\sum_{k=1}^{K}\frac{2\beta_k}{9}\delta_k^l\sum_{l'=1}^{L} p_{l'}\delta_k^{l'} = b_l, \quad \forall l \in L. \tag{16}$$

Evidently, $L$ formulas with $L$ variables can be obtained for a given urban rail corridor, and the coefficient of each variable $p_l$ depends on the specific network structure. The unique solutions with respect to the Lagrangian multiplier $p_l$ can be derived from the $L$ formulas with $L$ variables. By substituting $p_l$ into equation (11), we obtain a unique set of optimal service rates for each OD pair that minimizes the total passenger waiting time.

Let $a_{i,j}$ denote the coefficient of the $j$th Lagrangian multiplier $p_j$ associated with the $i$th link. Equation (16) can then be rewritten as follows:

$$\begin{cases} a_{1,1}p_1 + a_{1,2}p_2 + \cdots + a_{1,L}p_L = b_1, \\ a_{2,1}p_1 + a_{2,2}p_2 + \cdots + a_{2,L}p_L = b_2, \\ \cdots \\ a_{L,1}p_1 + a_{L,2}p_2 + \cdots + a_{L,L}p_L = b_L. \end{cases} \tag{17}$$

Let $\mathbf{A}$ denote the $L \times L$ coefficient matrix in equation (17), and

$$\mathbf{P} = \begin{Bmatrix} p_1 \\ p_2 \\ \vdots \\ p_L \end{Bmatrix}, \mathbf{B} = \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_L \end{Bmatrix}. \tag{18}$$

We can express equation (17) in a matrix form as follows:

$$\mathbf{AP} = \mathbf{B}. \tag{19}$$

The solution of the Lagrangian multiplier matrix can be obtained by solving the following equation.

$$\mathbf{P} = \mathbf{A}^{-1}\mathbf{B}. \tag{20}$$

*Cramer's rule* is an efficient method that uses determinants to solve systems of equations that have the same number of equations as the variables [36], as represented by equation (17). Let $\mathbf{D}$ be the determinant of the coefficient matrix of the equation (17).

$$\mathbf{D} = \begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,L} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,L} \\ \vdots & \vdots & & \vdots \\ a_{L,1} & a_{L,2} & \cdots & a_{L,L} \end{vmatrix}. \tag{21}$$

The solution using Cramer's Rule is given as follows:

$$p_1 = \frac{\mathbf{D}_1}{\mathbf{D}}, p_2 = \frac{\mathbf{D}_2}{\mathbf{D}}, \cdots, p_L = \frac{\mathbf{D}_L}{\mathbf{D}}, \tag{22}$$

where $\mathbf{D_j}$ denotes the determinant formed by replacing the $j$th column values with the answer-column values. The condition $\mathbf{D} \neq 0$ is naturally satisfied because it is impossible for all link flows in the network to be equal to 0. Let $A_{i,j}$ represent the cofactor of $a_{i,j}$

$$\begin{aligned} \mathbf{D_j} &= \sum_{i=1}^{L} b_i A_{i,j} \\ &= \begin{vmatrix} a_{1,1} & \cdots & a_{1,j-1} & b_1 & a_{1,j+1} & \cdots & a_{1,L} \\ a_{2,1} & \cdots & a_{2,j-1} & b_2 & a_{2,j+1} & \cdots & a_{2,L} \\ \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{L,1} & \cdots & a_{L,j-1} & b_L & a_{L,j+1} & \cdots & a_{L,L} \end{vmatrix}, \quad \forall j = 1, 2, \cdots, L. \end{aligned} \tag{23}$$

*4.3. Three-Node Network Example.* In this section, we use a hypothetical three-node network to examine the sensitivity of the proposed model to various decision factors. These factors include the arrival and service rates of each OD pair. The network consists of three OD pairs and two links, as depicted in Figure 4. Specifically, we have $k = 3, l = 2$.

Applying Cramer's rule, we can obtain the following equations involving two variables, $p_1$ and $p_2$:

$$\begin{aligned} \frac{2}{9}\left[\beta_1\left(p_1 + p_2\right) + \beta_2 p_2\right] &= b_1 \text{ for } l = 1, \\ \frac{2}{9}\left[\beta_1\left(p_1 + p_2\right) + \beta_3 p_2\right] &= b_2 \text{ for } l = 2. \end{aligned} \tag{24}$$

We can conveniently solve the above equations, resulting in


Figure 4: Simplified three-node network.

$$p_1 = \frac{9\left[(\beta_1+\beta_2)b_1 - \beta_1 b_2\right]}{2\left(\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3\right)}, \tag{25}$$

$$p_2 = \frac{9\left[(\beta_1+\beta_2)b_2 - \beta_1 b_1\right]}{2\left(\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3\right)}, \tag{26}$$

where

$$\begin{aligned} b_1 &= (\lambda_1+\lambda_2) - \mathrm{Cap}_1, \\ b_2 &= (\lambda_1+\lambda_3) - \mathrm{Cap}_2. \end{aligned} \tag{27}$$

By substituting the above formulas into equation (11), we can easily obtain an optimized passenger flow control strategy for each OD

$$\begin{aligned} \mu_1 &= \lambda_1 - \frac{2\beta_1}{9}(p_1+p_2) \\ &= \lambda_1 - \frac{\beta_1\beta_2(2\lambda_1+\lambda_2+\lambda_3-\mathrm{Cap}_1-\mathrm{Cap}_2)}{\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3}, \end{aligned} \tag{28}$$

$$\begin{aligned} \mu_2 &= \lambda_2 - \frac{2\beta_2}{9}p_1 \\ &= \lambda_2 - \frac{\beta_2\left[(\beta_1+\beta_2)(\lambda_1+\lambda_2-\mathrm{Cap}_1) - \beta_1(\lambda_1+\lambda_3-\mathrm{Cap}_2)\right]}{\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3}, \end{aligned} \tag{29}$$

$$\begin{aligned} \mu_3 &= \lambda_3 - \frac{2\beta_3}{9}p_2 \\ &= \lambda_3 - \frac{\beta_3\left[(\beta_1+\beta_2)(\lambda_1+\lambda_3-Cap_2) - \beta_1(\lambda_1+\lambda_2-Cap_1)\right]}{\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3}. \end{aligned} \tag{30}$$

For simplicity, we assume that the time-dependent arrival rate functions for these three OD pairs are the same, denoted as $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ and the parameters $\beta_1 = \beta_2 = \beta_3 = \beta$. In addition, we assume that the capacities of each link are identical, represented as $\mathrm{Cap}_1 = \mathrm{Cap}_2 = \mathrm{Cap}$. With these assumptions, equations (25) and (26) can be simplified to

$$\begin{aligned} p_1 &= p_2 \\ &= \frac{3(2\lambda - \mathrm{Cap})}{2\beta}. \end{aligned} \tag{31}$$

Equations (28)–(30) become

$$\mu_1 = \frac{2}{3}\mathrm{Cap} - \frac{1}{3}\lambda, \tag{32}$$

$$\begin{aligned} \mu_2 &= \mu_3 \\ &= \frac{1}{3}\mathrm{Cap} + \frac{1}{3}\lambda. \end{aligned} \tag{33}$$

To ensure a positive service rate $\mu_k$, we must have $\lambda \le 2\mathrm{Cap}$. When $\lambda$ exceeds 2Cap, the above mentioned analytical formula becomes insignificant.

Equations (28)–(30) indicate that the optimal passenger flow control strategy for this simple three-node network depends on the maximum arrival flow rate $\lambda_k(t_{1,k})$, which is simplified as $\lambda$ for convenience. However, if the conditions of $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ and $\beta_1 = \beta_2 = \beta_3 = \beta$ are violated, the passenger flow control strategy will also be affected by the quadratic coefficient $\beta_k$ in the estimated arrival rate function.

We assume that the link capacity Cap is 1000, and the optimal passenger flow control rates for each OD pair with respect to the maximum arrival flow rate $\lambda$ can be easily obtained, as shown in Figure 5. The service rate curves of OD pair 1-2 and 2-3 coincide because they share the same function, as shown in equation (33).

## 5. Integrated Optimization of Passenger Flow Control and Frequency Setting

The preceding section aims at enhancing the service level of the rail corridor from the demand side. To further expedite congestion reduction at platforms during the peak period, we extended the passenger flow control optimization model into a comprehensive mixed-integer programming model, considering both supply and demand factors. The model jointly designs an optimal flow control strategy and

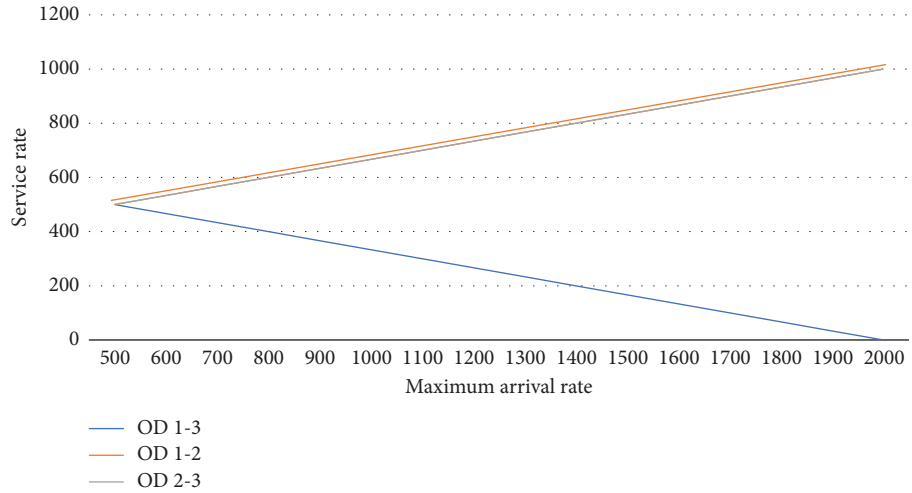



Figure 5: Optimized service rates for each OD pair under different maximum arrival rates.

frequency setting from a macroscopic perspective. A common assumption in such studies is that the departure times of vehicles are evenly distributed throughout the study period.

We present the objective function and constraints applied in the developed model as follows:

Objective function:

$$\min L(\boldsymbol{\mu}, \mathbf{f}) = \sum_{k=1}^{K} W_k(\mu_k) + \rho U_1 f, \tag{34}$$

subject to the following constraints.

Capacity constraint:

$$\sum_{k=1}^{K} \mu_k \delta_k^l T \le Cap \times f, \quad \forall l \in L. \tag{35}$$

Boundary constraint:

$$\mu_k \le \lambda_k\left(t_{1,k}\right), \quad \forall k \in K. \tag{36}$$

Headway constraint:

$$\frac{T}{H_{\max}} \le f \le \frac{T}{H_{\min}}. \tag{37}$$

Train unit constraint:

$$f n_1 \le N. \tag{38}$$

Variable domain:

$$\mu_k \ge 0, \quad \forall k \in K, \tag{39}$$

$$f \in Z^+. \tag{40}$$

In addition to the passenger flow control variable $\mu_k$, an integer decision variable $f$ was introduced to represent the scheduled frequency of the vehicles. The objective function (34) is developed to minimize the total passenger delay and the energy consumption. The first term of the objective function is consistent with that used in equation (4). $U_1$ represents the energy cost of one vehicle and $\rho$ denotes the weight. Constraint (35) ensures that the total passenger volume on each link does not exceed its transportation capacity. Constraint (36) restricts the allocated service rate for a specific OD pair to be less than or equal to its maximum arrival rate, avoiding resource waste. Constraint (37) ensures that the frequency value satisfies the requirements for maximum and minimum headways between two sequential vehicles. Constraint (38) is utilized to formulate the available fleet-size constraints. Notably, we did not consider the connection problem of train carriages at terminal stations on the unidirectional rail corridor. Constraints (39) and (40) define the feasible regions for the two types of decision variables.

Next, we will discuss the specific form of energy cost $U_1$. We assume that the energy cost of a vehicle mainly depends on its mass, which is identical to its capacity. In this study, we adopted the energy-cost function proposed by Chen et al. [37]:

$$U_1 = c_f + c_v(\mathrm{Cap})^{\alpha}, \tag{41}$$

where Cap represents the capacity of one train. $c_f$ denotes the fixed energy cost, which is independent of the vehicle capacity. $c_v$ is a positive coefficient and $\alpha$ is a power index with a value of $\alpha \le 1$. The energy cost function exhibits concavity over the vehicle capacity and shares the properties of economies of scale.

By substituting equations (4) and (41) into equation (34), the objective function can be rewritten as follows:

$$\min L(\boldsymbol{\mu}, \mathbf{f}) = \sum_{k=1}^{K} \frac{9\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]^2}{4\beta_k} + \rho\left(c_f + c_v(\mathrm{Cap})^{\alpha}\right) f. \tag{42}$$

## 6. Numerical Examples

This section presents two sets of numerical experiments, including a three-node network and Beijing Metro Line 1, to demonstrate the performance and effectiveness of the proposed approaches. All models were solved using the general-purpose, high-level Gurobi version 9.1 on a Windows 10 personal computer with an AMD Ryzen 7 5800H 3.20 GHz processor and 16 GB memory.

*6.1. Simple Case.* We examined the simple three-node rail line, as used in Section 4.3, with a link capacity of 1000 passengers. The time horizon studied lasted from 0 to 17. Given the estimated time-dependent arrival rate functions of each OD pair, the optimal passenger service rate for each OD pair can be calculated by solving the CPFC model (i.e., equations (4)–(7) and (2)) in less than 0.01 seconds. The arrival rate functions and the optimal service rates for each OD pair are listed in Table 3.

The cumulative arrival (CA) and departure (CD) curves without passenger flow control were plotted according to the traffic loading simulation procedure detailed in Appendix B. It is worth noting that, in the absence of passenger flow control, the service rate of each OD is calculated differently. First, the arrival counts of this OD pair are divided by the total arrival counts of all OD pairs that start from the same station. Then, this ratio is multiplied by the remaining available capacity of the following link during peak periods. However, if the following vehicle has sufficient capacity to accommodate all waiting passengers of this OD pair, the service rate of an OD pair equals the passenger arrival rate. The proposed traffic loading procedure without passenger flow control is consistent with actual passenger boarding rules.

Figure 6 displays the resulting CA and CD curves for each OD pair, both with and without flow control, as well as the total delay. We observe that under the passenger flow control scenario, the vehicle capacity is prioritized for passengers with short-distance OD pairs. Notably, the waiting time of the OD pair (1, 3) with passenger flow control is significantly greater than that in the scenario without flow control. This is because passengers with long-distance OD pairs would occupy vehicle capacity for an extended period. Consequently, more passengers with short-distance OD pairs would have to wait for the next arrival train, thereby resulting in an increase in waiting time. Furthermore, the proposed flow control method enhances the boarding equity for passengers departing from downstream stations during peak periods. This finding aligns with the conclusions presented by Wu et al. [38] and Shang et al. [39]. Regarding the evaluation of efficiency performance, the total delay under flow control is 18474, marking a reduction of 13 percent compared to the scenario without passenger flow control.

To further evaluate the integrated model of passenger flow control and train frequency setting, we conduct an additional numerical experiment. This experiment uses the three-node network and employs the time-dependent arrival rates listed in Table 2. According to the current operational information of the Beijing metro system, each train has a capacity of 3000 passengers (six units with 500 passengers/unit). We set the minimum and maximum headways to 3 min and 6 min, respectively. Energy cost-related parameters $c_f$, $c_v$, $\alpha$, and $\rho$ were set to 2.049, 0.37, 0.5, and 9.09 min/$, respectively, which were adapted from the work of Chen et al. [40]. The integrated model can be solved in 0.01 seconds, yielding optimal service rates for the three OD pairs as follows: $\mu_1 = 329$ and $\mu_2 = \mu_3 = 553.35$. Consequently, with a limited fleet size of 45 available, the resulting scheduled frequency is five. The objective function value is 38925.06, comprising a total waiting time of 37910.86 and a train energy cost of 1014.2. While this study does not consider other train operation costs (e.g., staff and maintenance costs), it should be noted that such costs usually account for a small percentage of the total cost when compared to passenger travel costs.

Table 3: Arrival rate functions and resulting optimal service rates.

| OD | Arrival rate (passengers/min) | Service rate (passengers/min) |
|---|---|---|
| 1–3 | $\lambda_1(t) = 550 - 5 \times (t-6)^2$ | $\mu_1 = 397$ |
| 1-2 | $\lambda_2(t) = 650 - 6 \times (t-6)^2$ | $\mu_2 = 603$ |
| 2-3 | $\lambda_3(t) = 750 - 7 \times (t-6)^2$ | $\mu_3 = 603$ |

*6.2. Large-Scale Experiments.* We further examine the effectiveness of the proposed passenger flow control strategy on Beijing Metro Line 1, as depicted in Figure 7. This rail corridor comprises 23 stations, including 12 transfer stations, which are represented as purple circles. Table 4 lists the names of the stations along with their corresponding numbers. This study only focuses on the direction from Pingguoyuan station to Sihuidong station. To cover the entire congestion process from start to dissipation, we consider the morning peak hours from 7:00 am to 10:00 am as our time horizon.

We used dynamic travel demand recorded by an automatic fare collection system on a weekday in February 2017 as input demand data. For transfer passengers with subroutes on this line, the first station of the subroute is considered the origin, and the last station is regarded as destination. Transfer passengers are identified within the transit network using the shortest path method. We do not introduce demand data preprocessing in detail because it is beyond the scope of this study. The total number of trips is 158925. The uneven demand profiles with a 1 min collection resolution among the OD pairs and stations during the period of interest are depicted in Figures 8 and 9, respectively. Due to space limitations, only a subset of the OD pairs and stations are marked on the horizontal axis. Notably, passenger volumes are extremely high for some stations, such as Guomao, Dawanglu, Pingguoyuan, and Wukesong.

Given the predetermined all-stopping pattern and even headway of the train, the capacity of the sections between two consecutive stations is set at 500 passengers/min. The

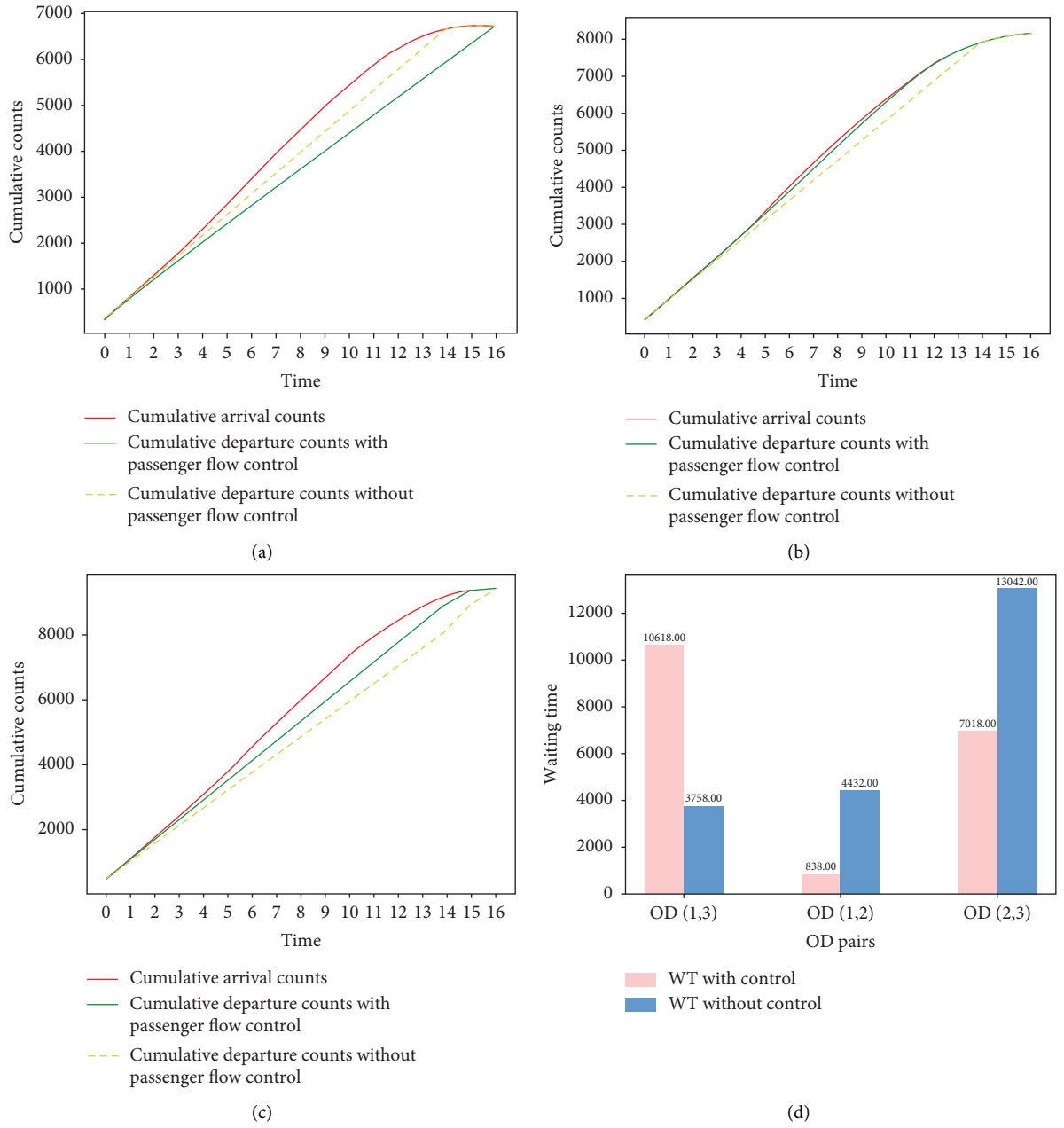



Figure 6: Comparison of cumulative arrival (CA) and departure (CD) curves of each OD pair and total waiting time. (a) CA and CD curves of OD (1, 3). (b) CA and CD curves of OD (1, 2). (c) CA and CD curves of OD (2, 3). (d) Total waiting time of with and without passenger flow control.

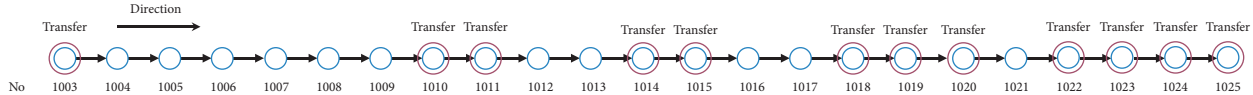



Figure 7: Beijing Metro Line 1 network.

CPFC model can be solved in 0.02 seconds. The optimized collaborative service rate allocation for each OD pair is illustrated in Figure 10. Notably, a greater proportion of transportation resources is allocated to the first and last few stations, which have higher passenger volumes, to reduce passenger waiting time. Given that this research only considers one direction of Line 1, only the upper right corner of the entire demand matrix presents nonnegative service rates.

Figures 11 and 12 present a comparison of passenger waiting time distributions, for each OD pair and station, respectively, with and without the application of the

Table 4: Station number and corresponding name.

| No. | Name | No. | Name | No. | Name |
|---|---|---|---|---|---|
| 1003 | Pingguoyuan | 1011 | Junshibowuguan | 1019 | Dongdan |
| 1004 | Gucheng | 1012 | Muxidi | 1020 | Jianguomen |
| 1005 | Bajiaoyouleyuan | 1013 | Nanlishilu | 1021 | Yonganli |
| 1006 | Babaoshan | 1014 | Fuxingmen | 1022 | Guomao |
| 1007 | Yuquanlu | 1015 | Xidan | 1023 | Dawanglu |
| 1008 | Wukesong | 1016 | Tiananmenxi | 1024 | Sihui |
| 1009 | Wanshoulu | 1017 | Tiananmendong | 1025 | Sihuidong |
| 10010 | Gongzhufen | 1018 | Wangfujing | | |

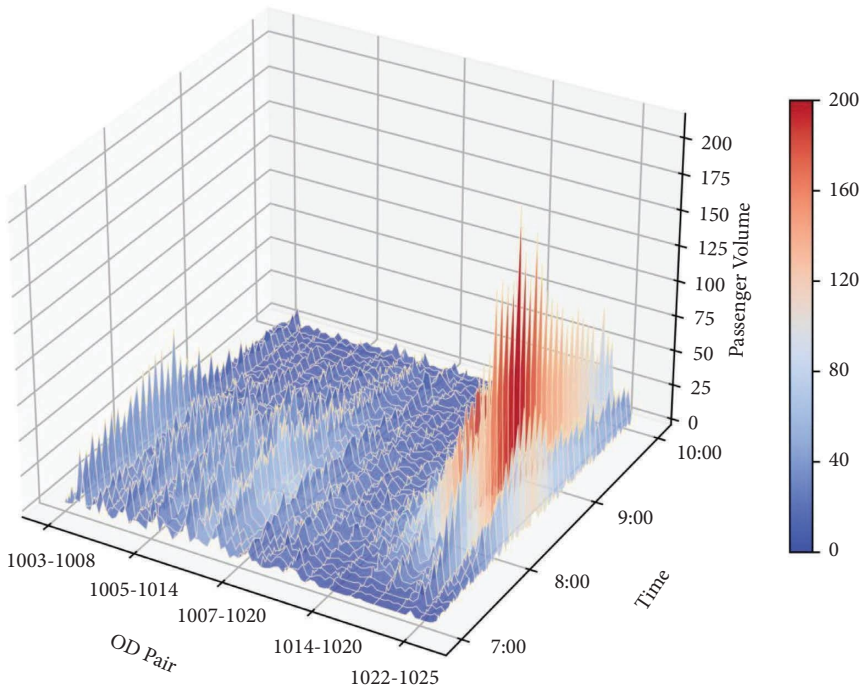



Figure 8: Time-dependent demand of each OD pair on Line 1.

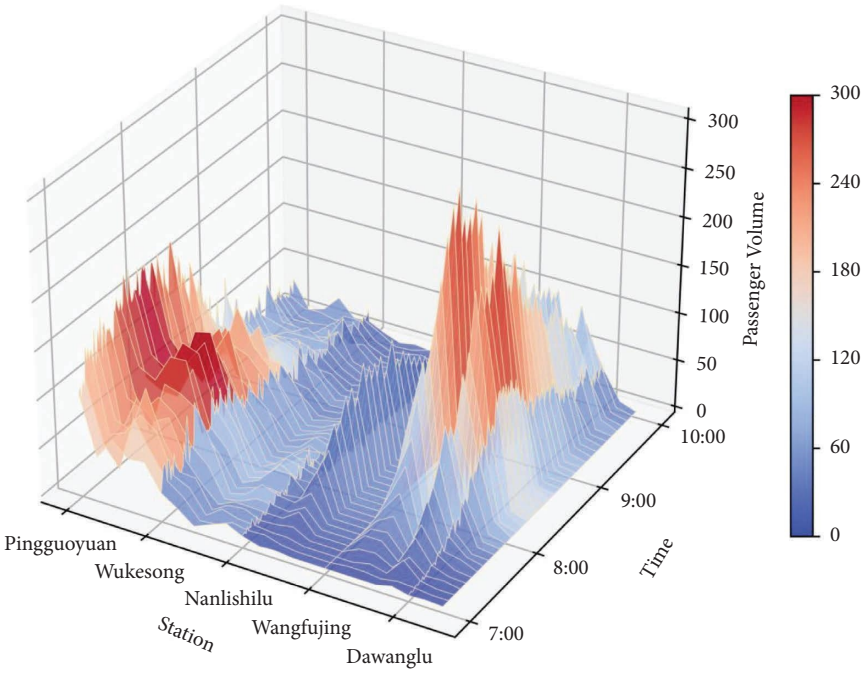



Figure 9: Time-dependent demand of each station on Line 1.

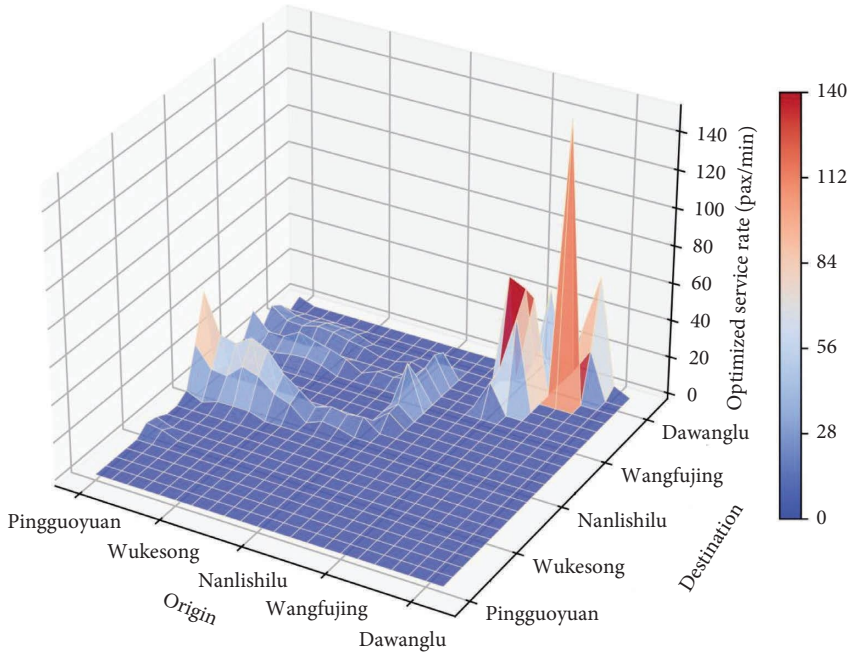
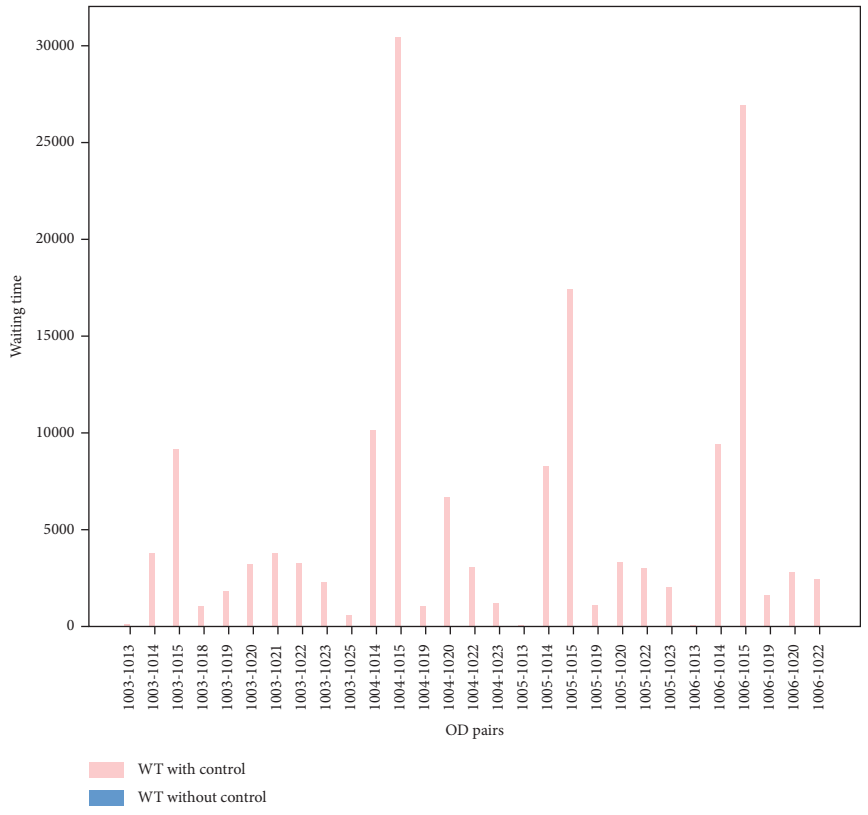
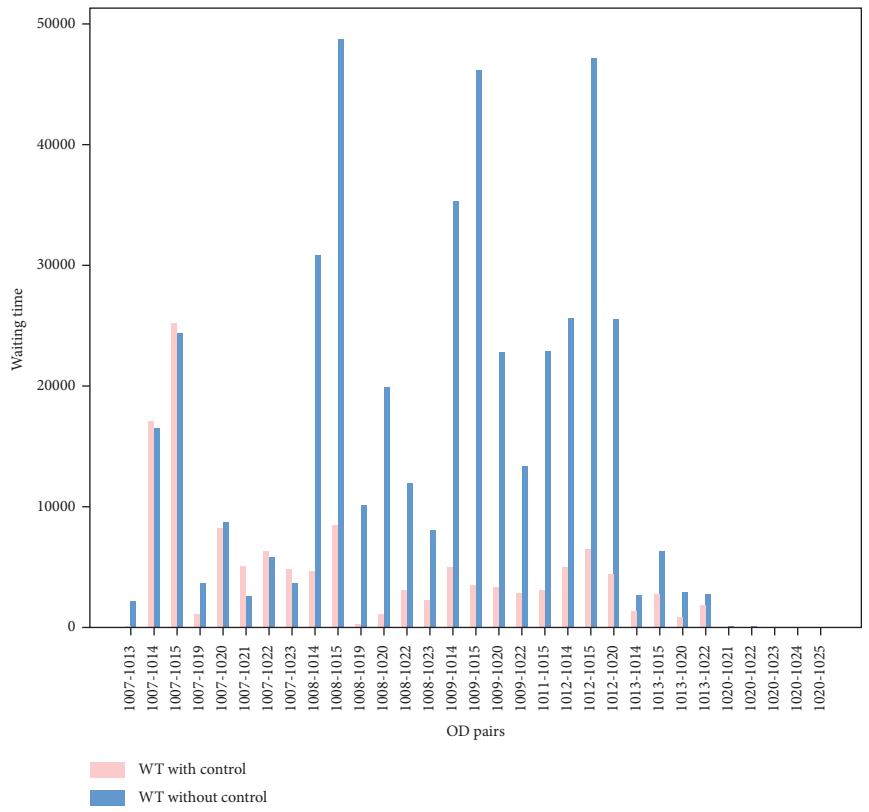



Figure 10: Optimized service rate for each OD pair.




(a)

Figure 11: Continued.




(b)

Figure 11: Comparison of waiting times for each OD with and without passenger flow control. (a) Waiting times of OD pairs originating from between Pingguoyuan station (No. 1003) to 2 Babaoshan station (No. 1006). (b) Waiting times of OD pairs originating from between Yuquanlu station (No. 1007) and Muxidi 1 station (No. 1020).

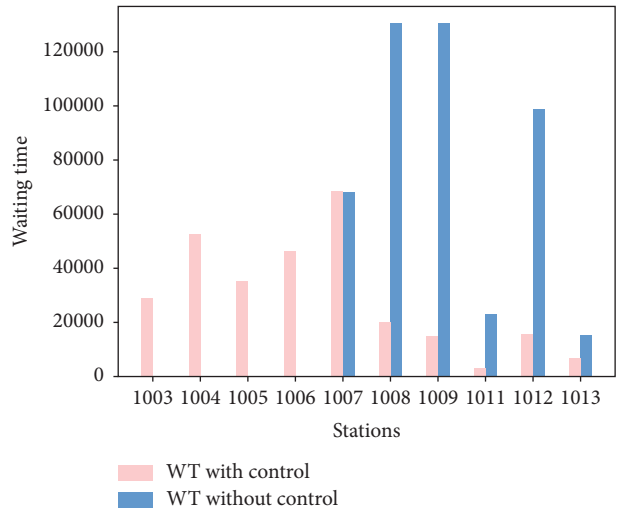



Figure 12: Comparison of waiting times for each station with and without passenger flow control.

passenger flow control strategy. For illustration, OD pairs and stations that did not experience any waiting time under either scenario have not been included. Without passenger flow control, passengers originating from Pingguoyuan station (No. 1003) to Babaoshan station (No. 1006) do not have to wait at the platforms, as they can successfully board the first-arrival train. Consequently, there are no waiting times depicted for OD pairs in Figure 11(a), and for stations numbered from 1003 to 1006 in Figure 12. As most of the transportation capacity is utilized by passengers originating from upstream stations, those departing from the latter stations often have significant difficulty boarding the first-

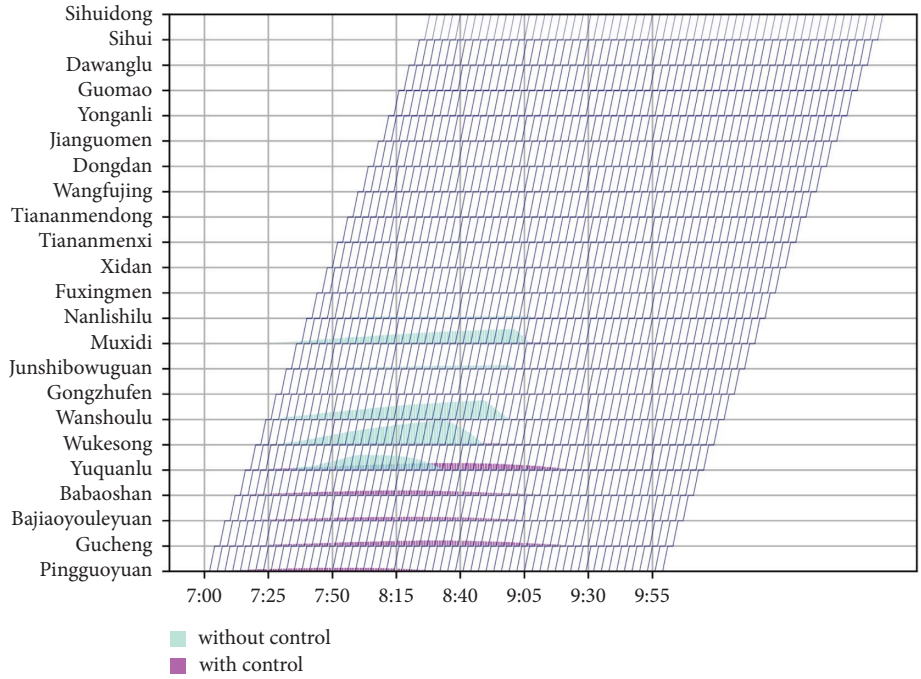



Figure 13: Time-dependent queue length at station hall.

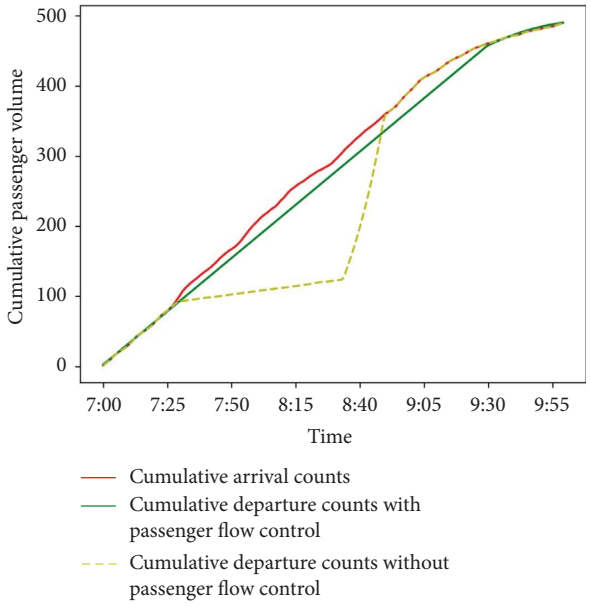



Figure 14: Cumulative arrival and departure curves of OD pair from Wukesong to Dawanglu.

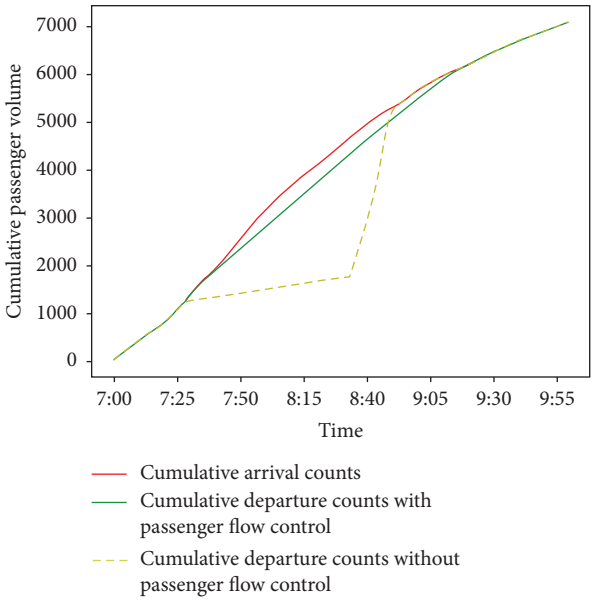



Figure 15: Cumulative arrival and departure curves of Wukesong station.

arrival train, leading to increased waiting times. Under a collaborative passenger flow control strategy, the total waiting time is 347872 with passenger flow control implemented, which represents a 23.2% reduction compared to the scenario without flow control.

Figure 13 indicates that severe passenger accumulation mainly occurrs from Yuquanlu station to Muxidi station without passenger flow control. This is largely due to the high volume of commuters traveling from suburban areas to downtown during the morning peak period. With passenger flow control implemented, traffic congestion is shifted to the first several upstream stations, specifically from the Pingguoyuan station to the Yuquanlu station. To further reveal the benefits of congestion shift, Figures 14 and 15 depict the cumulative passenger arrival and departure counts for the OD pair from Wukesong to Dawanglu, as well as Wukesong station, respectively. In Figure 14, the arrival rate begins to exceed the service rate at approximately 7:30 am, causing

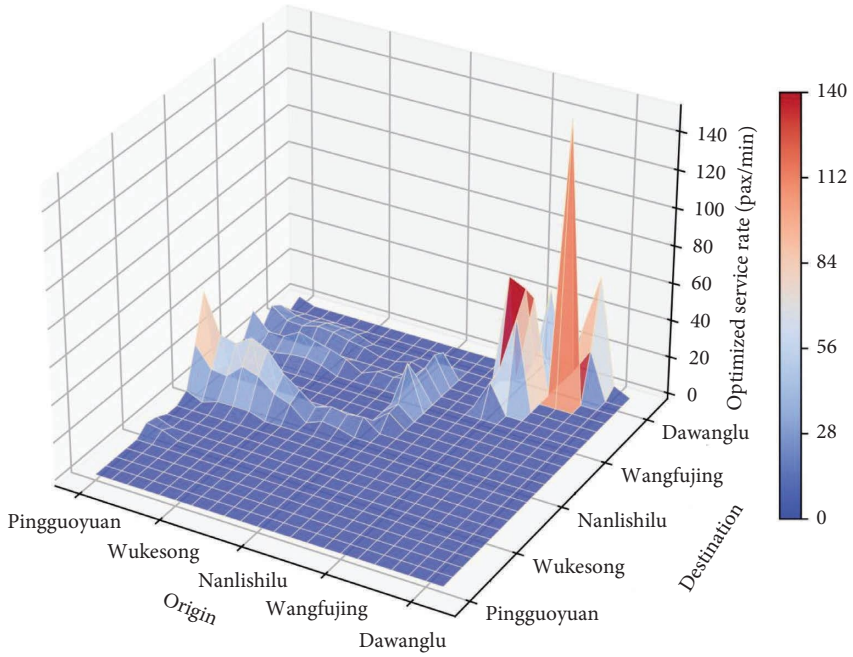



Figure 16: Optimized service rate considering frequency setting.

Table 5: Passenger loading procedure without passenger flow control.

| |
|---|
| For each $t \in T$ |
| $r\_Cap_l(t) = Cap_l$ |
| For each $s \in S$ |
| For each $k \in K, k(o) = s, k(d) > s$ |
| Step 1: calculate number of passengers that can be served |
| $a_k(t) = \min\{b_k(t), (b_k(t)/\sum_{k' \in K, k(o)=s, k(d)>s} b_{k'}(t)) r\_cap_{l(s,s+1)}(t)\}$ |
| Step 2: calculate number of remaining unserved passengers |
| $r_k(t) = b_k(t) - a_k(t)$ |
| Step 3: update total number of passengers to be served next time |
| $b_k(t+1) = r_k(t) + \lambda_k(t+1)$ |
| Step 4: update remaining available capacity for each link |
| For each $l \in L$ |
| $r\_cap_l(t) = r\_cap_l(t) - a_k(t)\delta_k^l$ |

passengers to wait in the station hall thereafter. Without passenger flow control, the cumulative departure curve rises slowly from 7:30 am to 8:30 am because train capacities are primarily occupied by passengers originating from upstream stations. Under passenger flow control, the cumulative departure curve rises steadily because more transportation resources can be assigned to this OD pair. In Figure 15, the curve trends are similar to those observed in Figure 14. Those similarities are due to the near-constant ratio of passenger volume for this OD pair to the total volume departing from the same station. We can conclude that the use of our proposed collaborative passenger flow control strategy allows for a more efficient and equitable distribution of limited transportation resources.

Subsequently, we evaluated the performance of the integrated optimization model for passenger flow control and train frequency settings. Its computation CPU time is 0.02 seconds. We set the minimum and maximum headways at 2 min and 6 min, respectively. The available fleet size stands at 58, with each train having a capacity of 1428 passengers. In addition, the other parameters remain consistent with those used in the experiments on the three-node network, as detailed in Section 6.1. Figure 16 plots the optimal service rates for each OD pair over the study period. The optimal number of scheduled vehicles stands at 58, with the total passenger waiting time amounting to 892322.57 and the total energy cost coming in at 8451.8. Interestingly, the service rate allocation pattern resembles that in Figure 10. However, the total waiting time was significantly higher than in experiments where frequency settings were not taken into consideration. The key distinction between these two experiments lies in the setting of link capacity. When implementing a collaborative passenger flow control strategy, the total passenger waiting time proves to be more sensitive to the distribution pattern of the optimized service rates, with transportation capacity following closely.

## 7. Conclusions

Focusing on a heavily congested urban rail corridor, a collaborative passenger flow control strategy was systematically investigated from a mesoscopic perspective to enhance transportation efficiency and improve passenger service levels. By formulating the problem using a continuous optimization method based on classic deterministic queuing theory, our proposed approach can not only optimize the passenger flow control strategy directly in large-scale urban rail lines experiencing time-dependent congestion but also avoid complex modeling representations for time-dependent and microscopic passenger and train activities. By introducing a Lagrangian relaxation solution approach,

an analytical solution was derived to quickly generate a reasonable passenger flow control strategy for real-world scenarios. By further developing an integrated passenger flow control strategy and train frequency setting model, our proposed approach can fully incorporate the interacting relationship between passenger demand and limited transportation resource. The optimization procedure of the passenger flow control strategy can assist urban rail operators in redistributing the transportation capacity more efficiently and equitably.

The proposed method provides a practical passenger flow control strategy, which can be effectively and quickly implemented for congested urban rail transit systems. Under this collaborative passenger boarding strategy, the system passenger waiting time is significantly reduced compared to scenarios without passenger flow control. Limited transportation capacity can also be fully utilized. Furthermore, the proposed joint control strategy can prevent passenger accumulation on platforms, thereby reducing potential accident risks and train delays. The passenger service quality and train operational efficiency are enhanced to a large extent. The proposed method is applicable to passenger demand management in the daily operation of rail systems and provides a tool for operators to schedule passengers collaboratively. This research sheds new light on understanding, modeling, and systematically optimizing passenger-oriented service concept and train schedules.

Our model only considers deterministic passenger demand and train timetable as input. Passenger flow control models were formulated based on deterministic demand distribution characteristics and predetermined train timetables. Besides, our developed approaches only consider a single rail line, neglecting complicated passenger route choice behaviors within a transportation network. Based on the abovementioned limitations, future research should concentrate on developing a model based on the following considerations. (1) In practical operations, unexpected events often disrupt regular train schedules and passenger demand is highly elastic owing to bad weather or other irregular occurrences. Therefore, exploring flexible flow control strategies applied to elastic demand and timetables is an essential topic for future research. (2) We will focus on generating a passenger flow control strategy that considers complex transfer activity and route choice behavior under a network-based train schedule in the future.

## Appendix

## A. Performance evaluation based on point queue theory

The passenger waiting time calculation method is inspired by Newell's deterministic fluid approximation model [35]. In oversaturated urban rail systems, the travel demand is relatively concentrated during peak hours; we therefore approximate the time-dependent passenger arrival rates using a polynomial function. Accordingly, the passenger waiting time can be derived conveniently.

Taking a specific OD pair as an example, Figure 2 shows how the total delay depends on the service rate $\mu$ with given an approximate deterministic cumulative arrival count $A(t)$. As shown, the cumulative departure curve $D(t)$ follows $A(t)$ very closely (essentially zero queue) until time $t_0$ when the arrival rate $\lambda(t)$ is equal to the constant service rate $\mu$. A queue $Q(t)$ starts to form at time $t_0$, reaches a maximum at time $t_2$, and vanishes at time $t_3$ when the total service count has caught up with the arrivals, that is, $D(t_3) - D(t_0) = A(t_3) - A(t_0) = D$. We can conveniently draw graphs of the actual arrival rate $\lambda(t)$ and service rate $\mu$, as shown in Figure 2. $\lambda(t)$ gradually increases from $t_0$ and reaches a maximum at $t_1$ and decreases until it is equal to $\mu$ again at time $t_2$.

Practically, the passengers of each OD pair arrive at their origin station continuously according to a certain time-dependent pattern. Thus, we assume that the passenger arrival rate $\lambda_k(t)$ of $k$th OD pair follows a polynomial trend with respect to the time. We denote $t_{1,k}$ as the time with the maximum arrival flow rate of $k$th OD pair, as shown in Figure 2. According to the Taylor series expansion, the time-dependent passenger arrival rate $\lambda_k(t)$ can be adequately represented by an $n$th-order polynomial function at times $t_{1,k}$

$$\lambda_k(t) = \lambda_k\left(t_{1,k}\right) + \lambda_k'\left(t_{1,k}\right)\left(t - t_{1,k}\right) + \frac{1}{2}\lambda_k''\left(t_{1,k}\right)\left(t - t_{1,k}\right)^2 + \ldots + \frac{1}{n!}\lambda_k^{(n)}\left(t_{1,k}\right)\left(t - t_{1,k}\right)^n + R_n(t), \tag{A.1}$$

where $\lambda_k^{(n)}(t_{1,k})$ denotes the $n$th-order derivative at $t_{1,k}$, and $R_n(t)$ is the remainder term of the Taylor's formula.

Newell assumed that the passenger arrival rate near time $t_{1,k}$ can be approximated by a quadratic function [35]. As a result, equation (A.1) degenerates into

$$\lambda_k(t) = \lambda_k\left(t_{1,k}\right) + \lambda_k'\left(t_{1,k}\right)\left(t - t_{1,k}\right) + \frac{1}{2}\lambda_k''\left(t_{1,k}\right)\left(t - t_{1,k}\right)^2. \tag{A.2}$$

Evidently, $\lambda_k'(t_{1,k}) = 0$, and letting $\beta_k = -1/2\lambda_k''(t_{1,k})$, equation (A.2) can be rewritten as follows:

$$\lambda_k(t) = \lambda_k\left(t_{1,k}\right) - \beta_k\left(t - t_{1,k}\right)^2. \tag{A.3}$$

For simplicity, we assume that the permitted service rate of $k$th OD pair is constant, that is, $\mu_k$. As shown in Figure 2, both $t_{0,k}$ (represented as $t_0$) and $t_{2,k}$ (represented as $t_2$) denote times when $\mu_k$ is equal to $\lambda_k(t)$, that is, $\mu_k = \lambda_k(t_{0,k}) = \lambda_k(t_{2,k})$

$$\mu_k = \lambda_k\left(t_{1,k}\right) - \beta_k\left(t_{0,k} - t_{1,k}\right)^2 = \lambda_k\left(t_{1,k}\right) - \beta_k\left(t_{2,k} - t_{1,k}\right)^2. \tag{A.4}$$

We can then obtain two real roots of $t_{0,k}$ and $t_{2,k}$ as follows

$$t_{0,k} = t_{1,k} - \left[\frac{\lambda_k\left(t_{1,k}\right) - \mu_k}{\beta_k}\right]^{1/2}, \qquad t_{2,k} = t_{1,k} + \left[\frac{\lambda_k\left(t_{1,k}\right) - \mu_k}{\beta_k}\right]^{1/2}. \tag{A.5}$$

According to the assumption of a quadratic $\lambda_k(t)$, the function $\lambda_k(t) - \mu_k$ has zeros at $t = t_{0,k}$ and $t = t_{2,k}$, and the second derivative with respect to $t$ is $-2\beta_k$. Now, we can write $\lambda_k(t) - \mu_k$ in factored form:

$$\lambda_k(t) - \mu_k = \beta_k\left(t - t_{0,k}\right)\left(t_{2,k} - t\right). \tag{A.6}$$

The queue length at any time $t_{0,k} < t < t_{3,k}$ can be calculated as follows:

$$Q_k(t) = A_k(t) - D_k(t) = \int_{t_{0,k}}^{t} \left[\lambda_k(\tau) - \mu_k\right] \mathrm{d}\tau. \tag{A.7}$$

By substituting equation (A.6) into equation (A.7), we can obtain the time-dependent queue length as

$$Q_k(t) = \beta_k\left(t - t_{0,k}\right)^2\left[\frac{\left(t_{2,k} - t_{0,k}\right)}{2} - \frac{\left(t - t_{0,k}\right)}{3}\right]. \tag{A.8}$$

The queue length has a maximum at time $t_{2,k}$

$$Q_k\left(t_{2,k}\right) = \frac{\beta_k}{6}\left(t_{2,k} - t_{0,k}\right)^3 = \frac{4\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]^{3/2}}{3{\beta_k}^{1/2}}. \tag{A.9}$$

The queue dissipates at time $t_{3,k}$, that is, $Q_k(t_{3,k}) = 0$, and $t_{3,k}$ can be derived as follows:

$$t_{3,k} = t_{0,k} + \frac{3}{2}\left(t_{2,k} - t_{0,k}\right) = t_{0,k} + 3\left(t_{1,k} - t_{0,k}\right). \tag{A.10}$$

Therefore, equation (A.8) can be rewritten as follows:

$$Q_k(t) = \frac{\beta_k}{3}\left(t - t_{0,k}\right)^2\left(t_{3,k} - t\right). \tag{A.11}$$

The total delay over the peak period is equal to the area between $A_k(t)$ and $D_k(t)$ in Figure 2, which can be calculated by integrating equation (A.11) between times $t_{0,k}$ and $t_{3,k}$

$$W_k = \int_{t_{0,k}}^{t_{3,k}} \left[\frac{\beta_k}{3}\left(\tau - t_{0,k}\right)^2\left(t_{3,k} - \tau\right)\right] d\tau = \frac{9\left[\lambda_k\left(t_{1,k}\right) - \mu_k\right]^2}{4\beta_k}. \tag{A.12}$$

Noticeably, the total delay over the rush over is quadratic with respect to the service rate $\mu_k$ given an estimated time-dependent passenger arrival rate $\lambda_k(t)$.

## B. Passenger loading procedure without passenger flow control

Table 5 provides the passenger loading procedure within transportation capacity to capture passenger boarding and alighting activities without passenger flow control. For convenience, the following notations used in the proposed algorithm are first introduced.

$a_k(t)$: number of passengers of OD pair $k$ that can be served at time t

$r_k(t)$: number of remaining unserved passengers of OD pair $k$ at time $t$

$b_k(t)$: total number of passengers of OD pair $k$ waiting to be served at time $t$, which equals the sum of the remaining passengers at time $t-1$ and passengers arriving at time $t$

$r\_cap_l(t)$: remaining available capacity of link $l$ at time $t$

$l(s, s+1)$: link $l$ with start station $s$ and end station $s+1$

## Data Availability

The data used to support the findings of this study are restricted by the Beijing Metro Line 1. Reasonable requests for data will be available from the corresponding author.

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Authors' Contributions

Qian Zhu conceptualized the study, proposed the methodology, visualized the study, and wrote the study. Xiaoning Zhu conceptualized the study, proposed the methodology, and reviewed and edited the manuscript. Pan Shang conceptualized the study, proposed the methodology, and acquired the funding. Lingyun Meng supervised the study, proposed the methodology, and acquired the funding.

## Acknowledgments

The study was funded by the Natural Science Foundation of Beijing Municipality, China (Grant no. L201016), and the National Natural Science Foundation of China (Grant no. 72001020 and U2034208).

## References

[1] E. Barrena, D. Canca, L. C. Coelho, and G. Laporte, "Single-line rail rapid transit timetabling under dynamic passenger demand," *Transportation Research Part B: Methodological*, vol. 70, pp. 134–150, 2014.

[2] V. Cacchiani, F. Furini, and M. P. Kidd, "Approaches to a real-world train timetabling problem in a railway node," *Omega*, vol. 58, pp. 97–110, 2016.

[3] F. Corman, A. D'Ariano, A. D. Marra, D. Pacciarelli, and M. Samà, "Integrating train scheduling and delay management in real-time railway traffic control," *Transportation*

*Research Part E: Logistics and Transportation Review*, vol. 105, pp. 213–239, 2017.

[4] F. Jiang, V. Cacchiani, and P. Toth, "Train timetabling by skip-stop planning in highly congested lines," *Transportation Research Part B: Methodological*, vol. 104, pp. 149–174, 2017.

[5] L. Kang and Q. Meng, "Two-phase decomposition method for the last train departure time choice in subway networks," *Transportation Research Part B: Methodological*, vol. 104, pp. 568–582, 2017.

[6] 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.

[7] L. Sun, J. G. Jin, D. H. Lee, K. W. Axhausen, and A. Erath, "Demand-driven timetable design for metro services," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 284–299, 2014.

[8] P. Vansteenwegen and D. V. Oudheusden, "Developing railway timetables which guarantee a better service," *European Journal of Operational Research*, vol. 173, no. 1, pp. 337–350, 2006.

[9] 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.

[10] L. Yang, J. Qi, S. Li, and Y. Gao, "Collaborative optimization for train scheduling and train stop planning on high-speed railways," *Omega*, vol. 64, pp. 57–76, 2016.

[11] C. Leiva, J. C. Muñoz, R. Giesen, and H. Larrain, "Design of limited-stop services for an urban bus corridor with capacity constraints," *Transportation Research Part B: Methodological*, vol. 44, no. 10, pp. 1186–1201, 2010.

[12] F. Meng, L. Yang, Y. Wei, S. Li, Z. Gao, and J. Shi, "Collaborative passenger flow control on an oversaturated metro line: a path choice approach," *Transportation Business*, vol. 8, no. 1, pp. 376–404, 2020.

[13] 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.

[14] J. Yin, T. Tang, L. Yang, Z. Gao, and B. Ran, "Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: an approximate dynamic programming approach," *Transportation Research Part B: Methodological*, vol. 91, pp. 178–210, 2016.

[15] Y. Yin, D. Li, K. Zhao, and R. Yang, "Optimum equilibrium passenger flow control strategies with delay penalty functions under oversaturated condition on urban rail transit," *Journal of Advanced Transportation*, vol. 2021, Article ID 3932627, 27 pages, 2021.

[16] Y. Zhou, Y. Bai, J. Li, B. Mao, and T. Li, "Integrated optimization on train control and timetable to minimize net energy consumption of metro lines," *Journal of Advanced Transportation*, vol. 2018, Article ID 7905820, 19 pages, 2018.

[17] L. J. LeBlanc, "Transit system network design," *Transportation Research Part B: Methodological*, vol. 22, no. 5, pp. 383–390, 1988.

[18] J. Shi, L. Yang, J. Yang, F. Zhou, and Z. Gao, "Cooperative passenger flow control in an oversaturated metro network with operational risk thresholds," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 301–336, 2019.

[19] G. Currie, "Quick and effective solution to rail overcrowding: free early bird ticket experience in Melbourne, Australia," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2146, no. 1, pp. 35–42, 2010.

[20] L. W. Lan, C.-H. Wen, and H.-Y. Lee, *Effects of Temporally Differential Fares on Taipei Metro Riders' Mode And Time-of-Day Choiches*, La Nuova Italia; RIET; Fabrizio Serra, Tuscany, Italy, 2010.

[21] S. Peer, J. Knockaert, and E. T. Verhoef, "Train commuters' scheduling preferences: evidence from a large-scale peak avoidance experiment," *Transportation Research Part B: Methodological*, vol. 83, pp. 314–333, 2016.

[22] H. Yang and Y. Tang, "Managing rail transit peak-hour congestion with a fare-reward scheme," *Transportation Research Part B: Methodological*, vol. 110, pp. 122–136, 2018.

[23] H. Akahane and M. Kuwahara, "A basic study on trip reservation systems for recreational trips on motorways," *Proceedings of the Third World Congress on Intelligent Transportation Systems*, vol. 660, pp. 1–7, 1996.

[24] J.-T. Wong, "Basic concepts for a system for advance booking for highway use," *Transport Policy*, vol. 4, no. 2, pp. 109–114, 1997.

[25] W. Liu, H. Yang, and Y. Yin, "Efficiency of a highway use reservation system for morning commute," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 293–308, 2015.

[26] F. Meng, L. Yang, J. Shi, Z.-Z. Jiang, and Z. Gao, "Collaborative passenger flow control for oversaturated metro lines: a stochastic optimization method," *Transportmetrica: Transportation Science*, vol. 18, no. 3, pp. 619–658, 2021.

[27] J. Shi, L. Yang, J. Yang, and Z. Gao, "Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: an integer linear optimization approach," *Transportation Research Part B: Methodological*, vol. 110, pp. 26–59, 2018.

[28] Y. Wang, S. Zhu, A. D'Ariano, J. Yin, J. Miao, and L. Meng, "Energy-efficient timetabling and rolling stock circulation planning based on automatic train operation levels for metro lines," *Transportation Research Part C: Emerging Technologies*, vol. 129, Article ID 103209, 2021.

[29] X. Xu, J. Liu, H. Li, and M. Jiang, "Capacity-oriented passenger flow control under uncertain demand: algorithm development and real-world case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 87, pp. 130–148, 2016.

[30] S. Li, M. M. Dessouky, L. Yang, and Z. Gao, "Joint optimal train regulation and passenger flow control strategy for high-frequency metro lines," *Transportation Research Part B: Methodological*, vol. 99, pp. 113–137, 2017.

[31] H. Zhou, Y. Liu, and Y. Zhao, "Passenger inflow control with hierarchical coordination for overloaded metro lines," *IET Intelligent Transport Systems*, vol. 14, no. 11, pp. 1418–1425, 2020.

[32] X. Wang, J. Wu, X. Yang, X. Guo, H. Yin, and H. Sun, "Multistation coordinated and dynamic passenger inflow control for a metro line," *IET Intelligent Transport Systems*, vol. 14, no. 9, pp. 1068–1078, 2020.

[33] R. Liu, S. Li, and L. Yang, "Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy," *Omega*, vol. 90, Article ID 101990, 2020.

[34] F. Yuan, H. Sun, L. Kang, and S. Zhang, "Joint optimization of train scheduling and dynamic passenger flow control strategy with headway-dependent demand," *Transportation Business*, vol. 10, pp. 627–651, 2022.

[35] G. F. Newell, *Applications of Queueing Theory*, Chapman and Hall Ltd, New York, NY, USA, 1982.
[36] Y. Chen, "A Cramer rule for solution of the general restricted linear equation," *Linear and Multilinear Algebra*, vol. 34, no. 2, pp. 177–186, 1993.
[37] Z. Chen, X. Li, and X. Zhou, "Operational design for shuttle systems with modular vehicles under oversaturated traffic: discrete modeling method," *Transportation Research Part B: Methodological*, vol. 122, pp. 1–19, 2019.
[38] Y. Wu, H. Yang, S. Zhao, and P. Shang, "Mitigating unfairness in urban rail transit operation: a mixed-integer linear programming approach," *Transportation Research Part B: Methodological*, vol. 149, pp. 418–442, 2021.
[39] P. Shang, R. Li, Z. Liu, L. Yang, and Y. Wang, "Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 321–343, 2018.
[40] S. Chen, H. Wang, and Q. Meng, "Solving the first-mile ridesharing problem using autonomous vehicles," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 45–60, 2020.