The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2632-0487.htm

Received 14 September 2023
Revised 17 November 2023
Accepted 14 December 2023

A conceptual non-stop operation strategy for urban rail transit: a case study of São Paulo's monorail line

Rubens C.N. Oliveira
School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China, and
Zhipeng Zhang
School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China and The State University of New Jersey, Piscataway, New Jersey, USA

Abstract

Purpose – The purpose of this study is to address the extended travel time caused by dwelling time at stations for passengers on traditional rail transit lines. To mitigate this issue, the authors propose the "Non-stop" design, which involves trains comprised of modular vehicles that can couple and uncouple from each other during operation, thereby eliminating dwelling time at stations..

Design/methodology/approach – The main contributions of this paper are threefold: first, to introduce the concept of non-stop rail transit lines, which, to the best of the authors' knowledge, has not been researched in the literature; second, to develop a framework for the operation schedule of such a line; and third, the author evaluate the potential of its implementation in terms of total passenger travel time.

Findings – The total travel time was reduced by 6% to 32.91%. The results show that the savings were more significant for long commutes and low train occupancy rates.

Research limitations/implications – The non-stop system can improve existing lines without the need for the construction of additional facilities, but it requires technological advances for rolling stock.

Originality/value – To eliminate dwelling time at stations, the authors present the "Non-stop" design, which is based on trains composed of locomotives that couple and uncouple from each other during operation, which to the best of the authors' knowledge has not been researched in the literature.

Keywords Dynamic coupling, Modular vehicles, Non-stop, Optimization models, Train scheduling, Travel time minimization

Paper type Conceptual paper

This study was supported by Chinese Scholarship Council (No. 2021DFJ000865) and partially supported by the Shanghai Sailing Program (21YF1420000).

Since submission of this article, the following author has updated their affiliation: Zhipeng Zhang is no longer associated with Rutgers The State University of New Jersey, Piscataway, New Jersey, USA.

Smart and Resilient Transportation
Vol. 6 No. 1, 2024
pp. 32-49
Emerald Publishing Limited
e-ISSN: 2632-0495
p-ISSN: 2632-0487
DOI 10.1108/SRT-09-2023-0006

1. Introduction

Passengers choose their mode of travel based on factors such as fare, comfort, waiting time and in-vehicle travel time. The travel mode choice for a group of passengers can be estimated by assigning weights to each of these factors and comparing the total utility of each mode (Ding and Zhang, 2016). Hence, reducing in-vehicle travel time would improve the utility of a given mode, attracting more passengers and enhancing the satisfaction of current riders.

In rail transit systems, in-vehicle travel time is impacted by the necessity for passengers to stop at every station along the way. To address this issue, some rail transit lines have implemented the Skip-Stop design, where trains stop at every other station. This method is effective for relatively high demand, long distances and when travel demand is evenly distributed along the subway line. Additionally, to further mitigate this issue, some rail transit lines have adopted the express line system, where trains stop only at major stations. This method can be advantageous when travel time saving outweighs waiting time (Suh *et al.*, 2002).

Express lines offer rapid connections between stations; however, they only cater to a portion of riders as the number of possible destinations is reduced compared to regular lines. Implementing an express line is costly, requiring the construction of additional tracks to allow overtaking, a cost not needed for skip-stop implementation. However, in skip-stop lines, a portion of passengers must switch lines during their trip, as their desired station is skipped. This increases in-vehicle travel time, waiting time and walking time, significantly reducing the service's utility for these passengers.

In recent years, companies and researchers have recognized the potential benefits of creating a line design that combines the advantages of both skip-stop and express line designs, enabling passengers to embark and alight at any station without intermediate stops. One promising development, patented by the company Next Future Transportation Inc. in 2022, involves "Selectively Combinable Independent Driving Vehicles," represented in Figure 1. This concept is relatively simple: using vehicles that can couple and uncouple at cruising speeds, passengers can access vehicles from other lines for direct transfers or simply move to a vehicle with a different stop plan on the same line.

On rail lines, these vehicles could be used to implement a design referred to as Non-stop (NS) in this study (see Figure 2). In the NS design, trains consist of locomotives capable of coupling and uncoupling during operation. This allows the front locomotive in a train to skip one station while the rear locomotive stops at it. As depicted in the figure, at time "i" locomotives L1 and L2 move from station ST1 toward station ST2, where there is one parked locomotive, L3. At time "$i+1$," L1 and L2 uncouple: L1 begins to decelerate to stop at ST2, while L2, maintaining a steady pace, intends to skip ST2. Simultaneously, L3 departs from ST2 by accelerating. Finally, at time "$i+2$" L1 halts at ST2, allowing passengers to alight and embark, while locomotives L2 and L3 couple, enabling passengers to move between the locomotives.

Source: www.next-future-mobility.com

Figure 1. MV designed by the Next Future Transportation Inc (Next Future Transportation-Inc., 2021)

This movement between locomotives is what makes the design efficient and non-stop: passengers wishing to skip the next station can simply move to the front locomotive, while those needing to stop at the next station can do so by moving to the back locomotive. The non-stop design offers two significant benefits over the traditional methods cited above.

First, non-stop significantly reduces total travel time, as passengers are bound to skip at least half the stations during their trip – every locomotive skips one station before stopping. Passengers can also skip every single intermediate station by switching locomotives during coupling time, as the front locomotive of a train is bound to skip the next station.

Second, non-stop does not reduce the line's demand. Unlike the express and skip-stop designs, which completely skip certain stations – resulting in some passengers needing to switch lines or modes of transportation to reach their destination – non-stop allows passengers to stop at any station in the line. This is possible because there is always one locomotive stopping at the next station; hence, passengers can move to the stopping locomotive during coupling time.

Non-stop does not require significant investments in infrastructure, such as the construction of new rail tracks, or lines, which can be expensive and less attractive if built deep into the ground, thereby taking longer to reach (Lee *et al.*, 2021). However, it does require locomotives capable of dynamic coupling (DC) – coupling and decoupling at non-zero speed – allowing passengers to safely walk over to another wagon to either skip or stop at the next station.

1.1 Paper contributions

The aim of this work is to explore the applicability of modular vehicles (MVs) capable of DC in railway lines. Much research has been conducted in rail transit timetabling, encompassing various issues such as energy consumption (Scheepmaker *et al.*, 2017), resilience (Ren *et al.*, 2019), safety (Shi *et al.*, 2023), rolling stock routing (Wang *et al.*, 2021), rescheduling (Jiateng *et al.*, 2022) and more. However, most of these approaches did not consider DC or MV. While some recent studies have contemplated these options, to the best of our knowledge, none have applied them to urban rail traffic or considered the non-stop design. In summary, this paper contributes as follows:

- Our work proposes a novel line design for urban rail traffic, leveraging emerging DC technology that enables passengers to skip any station in the line while still being able to stop at their destination for alighting.
- We develop an optimization method for non-stop trains, defining speed trajectories. The model aims to minimize total passenger travel time, accounting for speed limitations, DC and the feasibility of coupling duration to ensure passengers have sufficient time to switch locomotives.

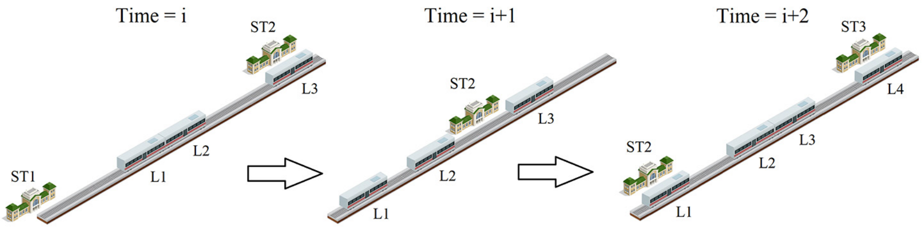


Figure 2.
Operation of a segment with non-stop design

Source: Authors' own work

- We investigate two real-world instances: a small-scale data set from São Paulo's recently opened monorail line 15 and a medium-scale data set from San Francisco's Bay Area Rapid Transit (BART). By comparing our results with current travel times, we demonstrate that the proposed design can reduce travel times by 6%–32.91%.

The paper is organized as follows: Section 2 reviews relevant literature. Section 3 contains the problem description, main concepts, mathematical model and solution method. In Section 4, the model is applied to case studies, and the results are analyzed. Finally, Section 5 discusses the conclusions and ideas for future work.

2. Literature review

The train timetabling problem focuses on finding the optimal timetable concerning passenger travel time, energy consumption, profit, capacity and other factors. The process consists of turning the problem into a mathematical model and solving it with an algorithm that searches for the maximum or minimum value of a function within a given set.

Early works include (Szpigel, 1973) where the problem was formulated as a job-scheduling problem. This formulation sought to minimize travel time and considered trains as jobs and track segments as machines. Branch-and-bound was chosen as the solution method for this non-cyclic train timetabling problem (TTP) that was applied to a Brazilian single-track line.

Although other works followed a similar structure to deal with this problem (Zhou and Zhong, 2007), literature has focused on solving this problem for different objective functions such as maximizing reliability (Jovanovic and Harker, 1991; Carey, 1994), maximizing profit (Brännlund *et al.*, 1998), minimizing energy consumption (Guo and Zhang, 2022; Sun *et al.*, 2023) and maximizing quality of service (Kaspi, 2010).

Regarding time minimization, Bai *et al.* (2019) focus on optimizing train headways at each station to reduce passenger waiting time and platform crowding using a genetic algorithm. Similarly, Zhang *et al.* (2021) research a timetable optimization model for express/local modes to minimize total passenger waiting time at platforms, resulting in a 9.3% savings. Other related research includes a demand-driven timetable assisted by smart card-based automated fare collection systems (Sun *et al.*, 2014) and train scheduling to minimize passenger waiting time (Niu *et al.*, 2015). Niu *et al.* (2015) used a quadratic integer programming model to synchronize passenger loading time windows with train arrival and departure times, aiming to minimize passenger waiting time at stations with time-dependent demand and skip-stop patterns.

Optimization related to skip-stop patterns is relevant to this research due to its similarity to the proposed model. Jiang *et al.* (2017) proposed a model that allows changes in dwelling time and stops to insert additional trains into a highly congested existing timetable to meet demand. Parbo *et al.* (2018) reduced in-vehicle travel time by 5.5% by optimizing stopping patterns. Zhao *et al.* (2021) presented a mixed-integer nonlinear programming (MINLP) model for deciding stop planning, timetables and rolling stock usage to minimize unfairness in passenger waiting time and workload imbalance in rolling stock circulation.

2.1 Literature on modular vehicles

Nold and Corman (2021) conducted a systematic classification of dynamic train unit coupling and decoupling at cruising speeds, identifying two primary types of couplings: mechanical and virtual. Mechanical couplings occur when units physically bind together, while virtual couplings maintain a constant distance through coordinated speed management. Their classification also outlined three types of coupling processes:

1. static manual: couplings requiring physical human assistance;
2. static automatic: couplings not requiring physical human assistance; and
3. automatic dynamic: couplings achievable at speeds higher than zero without human assistance.

According to their classification, the NS design necessitates trains capable of mechanical-automatic DC (or DC) with transfer allowances between trains – classified as the 4.3 generation of unit coupling in operation (UCO). Their numerical evaluation demonstrated a potential travel time reduction of 30%, excluding the novel line designs enabled by this technology.

The implementation of railway lines using 4-gen UCO requires further research. Wu *et al.* (2023) proposed a train controller based on adaptive virtual coupling and a train model to enable DC. The model underwent testing through multiple simulations, demonstrating feasibility. However, in certain instances, the generated jerk from the coupling surpassed the maximum recommended comfort limit.

To the best of our knowledge, very little research has been made on TTP optimization models for train DC, Pei *et al.* (2023) presented a MINLP model to consider passenger flow uncertainty and energy consumption for modular rolling-stock, however, their model only considered couplings at stations, focusing on train sizing for minimal energy consumption. On the other hand, some research has been conducted on bus DC, Zhang *et al.* (2020) presented a model that minimizes the vehicles-miles-traveled while imposing a lower bound to the number of served ride requests. Their model considered enroute transfers and vehicle routing, resulting in a timetable that is completely adapted to passenger's demand. Wu *et al.* (2021) proposed a model with similar prepositions with the objective of minimizing the total number of passenger transfers and the number of MVs maneuvers, such as switching lanes.

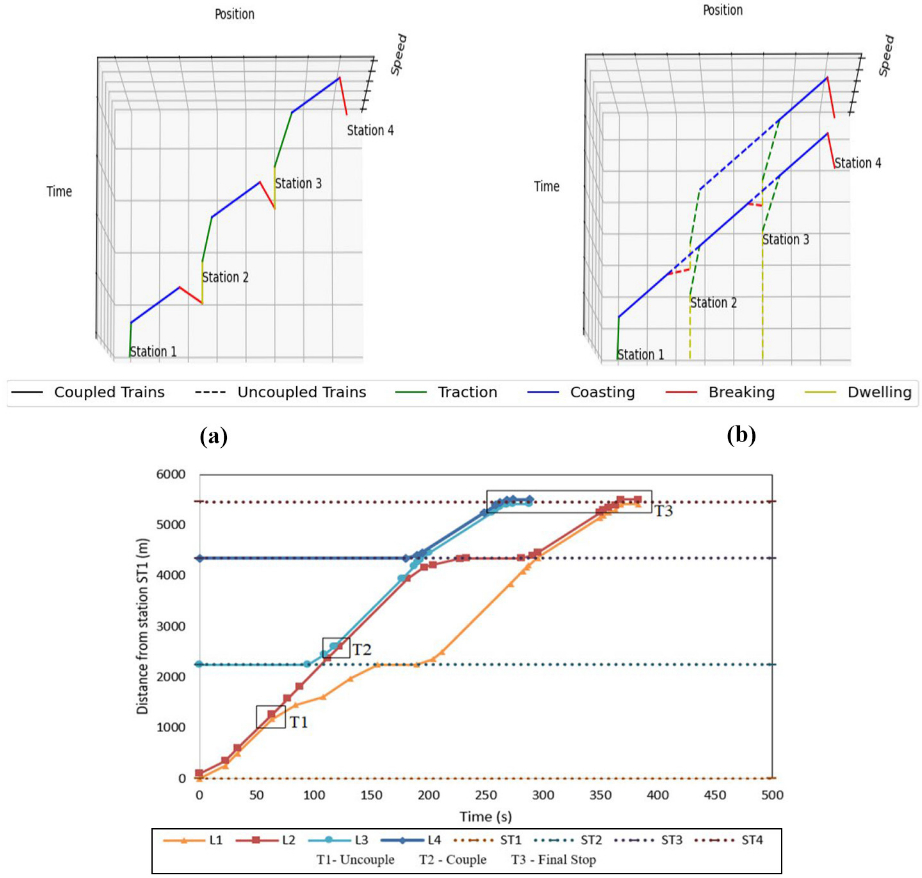
In general, operations research methods can be used to solve a plethora of rail transit optimization problems, and for this reason, it was the chosen method for this paper. To the best of the authors' knowledge, no academic research involving non-stop systems has been conducted. Notably, it explores the feasibility and potential benefits of implementing non-stop systems mathematically. As travel time significantly influences passenger mode choice and capacity, impacting demand and profit, the model concentrates on minimizing passenger in-vehicle travel time.

3. Problem definition and mathematical formulation

3.1 Non-stop design in rail transit system

In the non-stop system, passengers can reach their destination directly without stopping at any station. To better illustrate its functionality, a representation of the non-stop design as a train schedule timetable is presented and compared to the traditional full-stop (FS) line design, in Figure 3. In FS (a), a train departs from the first station, ST1, toward the last station, ST4, with stops at the intermediate stations ST2 and ST3.

Conversely, the non-stop line (b) consists of trains with multiple locomotives enabling independent movement after uncoupling. Hence, in the non-stop design, locomotives depart not only from the first station but also from each subsequent station along the line. Although it may appear that the non-stop design leads to extended waiting times for locomotives L3 and L4, this only occurs for the first train of the line, for instance, at 6 a.m. when the subway system opens. Furthermore, trains uncouple during operation (c), depicted at time T1, where locomotive L1 slows down to stop at station ST1. Simultaneously, locomotive L2 continues moving to couple with locomotive L3 at time T2, allowing passengers from L2 to transfer to L3 and bypass stopping at station ST3. This process also allows passengers from L3 to move to locomotive L2 and disembark at ST3. Finally, at the



(c)

Notes: (a) FS design; (b) NS design; (c) NS design timetable
Source: Authors' own work

Figure 3. Comparison between the operation of a traditional and a non-stop rail transit line

last station of the line, none of the locomotives uncouple, as there are no further stations to skip, as shown at time T3.

As a result, passengers who prefer not to stop at any intermediate station can continuously move to the next locomotive until they reach the one that will stop at their desired destination. However, this movement is restricted by the physical limitation of the door width between locomotives. Consequently, our study considered extreme high-demand scenarios where overcrowding would prevent passengers from switching locomotives. In this situation, the non-stop design could technically transform into a skip-stop design. Even in this worst-case scenario, passengers can still benefit from reduced stop time and consequently lower travel time.

Headway assignment, defining the interval between vehicles moving in the same direction on the same route, becomes a crucial issue in non-stop applications because each train's movement relies on adjacent trains. For instance, if a 180-s headway is assigned for a line, the trains would need to remain at stations for 100–150 s, waiting for the next coupling locomotive to arrive. This is illustrated in Figure 4, where Set 1 L1 represents the rear locomotive of the train departing the station at time "*i*," and Set 2 L2 represents the front locomotive departing at time "*i+headway*."

Note that it takes longer for the rear locomotive L1 to reach station ST1 than for the front locomotive L2. This occurs because L1 must stop at ST1, requiring it to slow down before reaching the station. Therefore, a minimum headway that accounts for this time difference is necessary to make the line feasible. In other words, a headway that is too short or unfeasible could result in locomotives colliding.

In the minimum feasible headway scenario, L1 stops at ST1 for 0 s, making it impossible for passengers to embark or alight. An efficient headway for this section would be 90 s, providing passengers with 33.98 s for alighting/embarking. Conversely, a headway of 180 s would result in an inconveniently long waiting time of 123.98 s.

A train schedule was created for visualization purposes, as depicted in Figure 5. The figure displays locomotives departing from ST1 between 11:20 and 11:25 to illustrate how

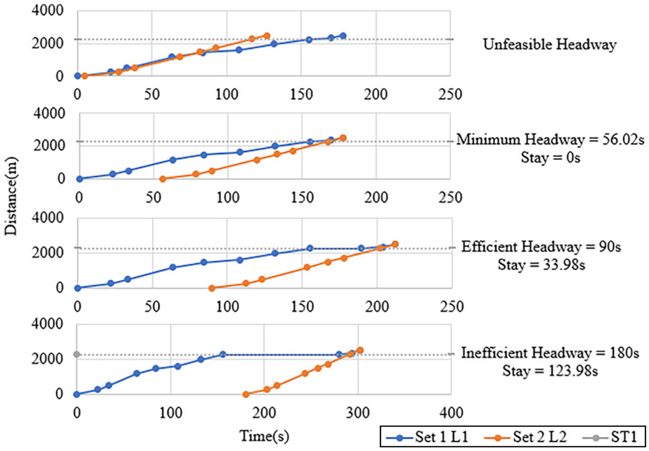


Figure 4.
Stay and headway relationship

Source: Authors' own work

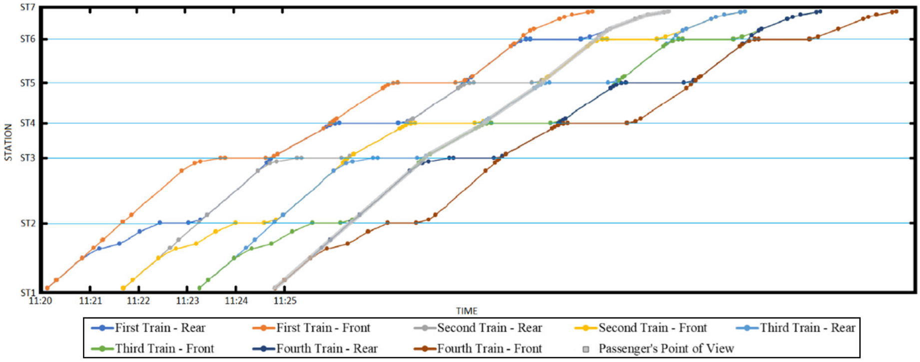


Figure 5.
Non-stop train schedule with passenger's point of view

Source: Authors' own work

waiting time at stations is influenced by the headway and to present the perspective of a passenger moving from station ST1 to ST7, consistently switching locomotives to bypass any intermediate station.

Implementing this design would necessitate significant development in intelligent rolling stock and reliable information systems for railway lines, which have witnessed substantial advancements in recent years due to the integration of autonomous trains. However, the mechanical and electrical advancements required for this model's application are not the primary focus of this study. Instead, the concept is approached from the perspective of operational research. We have formulated a mathematical model to represent its implementation and optimized its operation by minimizing the total passenger travel time. In summary, the problem entails finding a train schedule that minimizes total passenger travel time while satisfying the following constraints:

- Locomotive movements must adhere to basic kinematic equations.
- The average speed, maximum speed and acceleration of the optimized non-stop line cannot exceed those of the original traditional line.
- Locomotives must stop at every other station.
- The duration of a locomotive's stop at a station is determined by the headway and adjacent locomotive movements.
- Each locomotive must couple with its adjacent locomotives (rear and front) at least once to facilitate passenger movement.
- Passengers must use a combination of locomotives to travel from their origin to their destination (defined as paths in this study).
- The maximum passenger capacity of a locomotive must be observed.
- During the coupling phase, the total passengers switching locomotives must not exceed the door's capacity.

3.2 Operation strategy with main concepts

Traditionally, when optimizing travel time for rail transit, the focus of the study does not encompass the train's movement between stations. The conventional idea is rather straightforward – the train should move as fast as possible while adhering to the operational speed in each segment. However, in a non-stop system, it becomes crucial to pinpoint the train's speed and location for the safe coupling and uncoupling of trains. As a result, certain concepts needed to be adopted to transform this aspect into a mathematical model.

The main objective in this study was to simulate a feasible and efficient, based on total passenger travel-time, non-stop line. Consequently, the objective function is the minimization of the time each passenger takes to get from his origin to his destination, which can be achieved in many different ways in the non-stop design, such as skipping every station or just skipping a few, hence, we named these options as *paths*, as it can be seen in (1), where D_p represents the time it takes to complete a path, and N_p, represents the number of passengers who use the path:

$$W_1 = min \sum_{p \in P} D_p \times N_p \tag{1}$$

An important assumption had to be made before running the algorithm. It was necessary to define how much people could move through a single door over a defined period.

According to Oh *et al.* (2016), the passenger flow through a door on urban railways can be mathematically calculated with (2):

$$FT = -2.269 + (0.913 \times B) + (0.492 \times A) + (0.132 \times S) - (0.162 \times DW) - (0.030 \times B \times DW) \qquad (2)$$

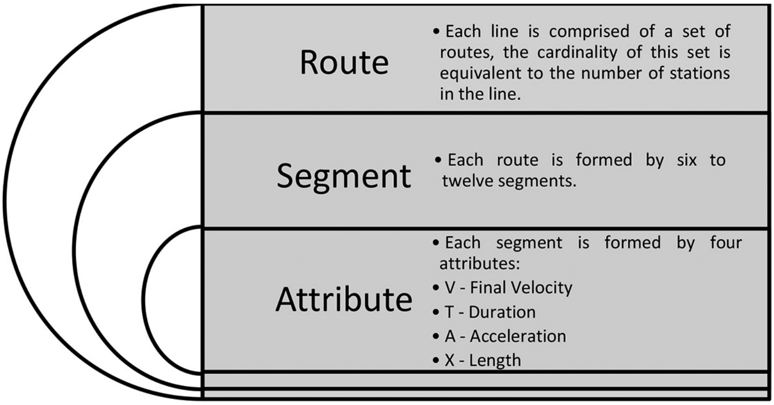
Where FT represents Flow Time, B indicates the number of people boarding, A signifies the number of people alighting, S denotes the number of people standing, and DW stands for door width. According to this formula, we simplified the flow time as one passenger per second.

Additionally, the components of the train schedule were segregated into three layers. First, the most fundamental variables defining the locomotive's movement:

- A – acceleration;
- V – final velocity;
- T – duration; and
- X – length.

Second, the *segment*, which comprises these four variables. Third, the *route*, formed by a set of segments, corresponding in this study to the train's movement from one station to another. Figure 6 illustrates a representation of these layers. The segments play a critical role in constraining the locomotive's movement. For instance, in the final segment of a route, the velocity must reach zero to ensure the train halts at its destination. Furthermore, the total sum of segment lengths within a route should equal the distance between the origin and destination stations.

As length and final velocity depend on duration and acceleration, such constraints impose rigid limitations on the segment. Similar constraints are imposed for the coupling and uncoupling movement of locomotives, necessitating the creation of unconstrained intermediate segments. Therefore, in our implementation, each route was assigned six to twelve segments, dependent on its length.



- Each line is comprised of a set of routes, the cardinality of this set is equivalent to the number of stations in the line.

- Each route is formed by six to twelve segments.

- Each segment is formed by four attributes:
- V - Final Velocity
- T - Duration
- A - Acceleration
- X - Length

Figure 6.
Representation of the layers that define a route

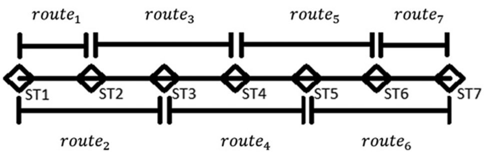
Source: Authors' own work

According to the adopted definitions, a line with seven stations would be split into seven different *routes* as shown on Figure 7. When a locomotive completes $route_1$, for instance, it immediately starts $route_3$. Therefore, a passenger transferring from *L2* to *L3* is equivalent to switching from $route_2$ to $route_3$. It is also important to note that $route_1$ and $route_7$ are shorter than the other routes. This discrepancy occurs only at the first and penultimate stations of the line and ensures that every station is serviced.

Consequently, a passenger boarding at ST2 without changing locomotives will traverse routes 3, 5 and 7, forming a path that skips stations ST3 and ST5. Expanding this reasoning led to the observation that odd-numbered locomotives follow odd-numbered routes, while even-numbered locomotives traverse even-numbered routes. This principle holds significant importance in preventing bottlenecks. For instance, if a passenger is aware that their destination is an even-numbered station, they can proactively switch to an even-numbered locomotive, thereby avoiding a transfer during a potentially crowded train moment.

3.3 Notation

The primary elements used in the mathematical model are displayed in Table 1. By interconnecting the decision variables within the constraints, it became feasible to create train schedules that minimize passenger travel time, guarantee proper locomotive coupling,



Source: Authors' own work

Figure 7. Route assignment to rail transit line

(a) General subscripts	
p	Path indices, $p \in P$, where P is the set of paths
s	Segment indices, $s \in S$, where S is the set of segments of a locomotive
i, j	Station indices, i, $j \in R$, where R is the set of stations
k	Locomotive indices, $k \in$ K, where K is the set of all locomotives, K_e is the set that only includes the first and last locomotives which have shorter routes, and K_n is the set that contains the remaining locomotives. Therefore K = K_e U K_n
(b) Parameters	
$O_{i,j}$	Number of passengers going from station *i* to station *j*
$E_{i,j}$	Distance between station *i* and station *j*
$M_{i,j}$	Maximum speed between station *i* and station *j*
$Q_{i,j}$	Average speed between station *i* and station *j*
(c) Decision variables	
$T_{k,s}$	Duration of segment *s* for locomotive *k* (*seconds*)
$V_{k,s}$	Velocity of segment *s* for locomotive *k* (*m/s*)
$A_{k,s}$	Acceleration of segment *s* for locomotive *k* (m/s^2)
$X_{k,s}$	Length of segment *s* for locomotive *k* (*meters*)
$Stay_k$	Time length a locomotive stop at a station (*seconds*)
$N_{p,i,j}$	Number of passengers going from station *i* to station *j* through path *p*

Source: Authors' own work

Table 1. Mathematical model notation

precise station stops and compliance with speed limits. The following section presents the mathematical model.

3.4 Mathematical model

3.4.1 Comprehensive model for lines with less than ten stations. On top of the objective function which was introduced in (1), the following constraints were used in the model:

$$V_{k,7} = V_{k+1,2} \qquad \forall\ k \in K_n \tag{3}$$

$$V_{k,8} = V_{k+1,3} \qquad \forall\ k \in K_n \tag{4}$$

$$X_{k,8} = X_{k+1,3} \qquad \forall\ k \in K_n \tag{5}$$

$$T_{k,8} = T_{k+1,3} \qquad \forall\ k \in K_n \tag{6}$$

$$\sum_{s \in \check{\mathrm{S}}} X_{k,s} - E_{k-1,k} \leq \sum_{s \in \c{S}} X_{k,s} \qquad \forall k \in K_n \tag{7}$$

$$X_{k,9} \geq X_{k+1,4} + 25 \qquad \forall k \in K_n \tag{8}$$

$$V_{k,s+1} - V_{k,s} = T_{k,s} \times A_{k,s} \qquad \forall k \in \mathrm{K}, \forall\ \ \mathrm{s} \in \mathrm{S} - 1 \tag{9}$$

$$X_{k,s+1} = \frac{{T_{k,s+1}}^2 \times A_{k,s+1}}{2} + V_{k,s} \times T_{k,s} \qquad \forall k \in \mathrm{K}, \forall \mathrm{s} \in \mathrm{S} - 1 \tag{10}$$

$$\sum_{s \in S} X_{k,s} = E_{k,k+2} \qquad \forall\ k \in K_n \tag{11}$$

$$\sum_{s \in S} X_{k,s} = E_{k,k+1} \qquad \forall\ k \in K_e \tag{12}$$

$$Stay_k \geq \sum_{s \in \hat{\mathrm{S}}} T_{k,s} - \sum_{s \in \mathbb{S}} T_{k+1,s} \qquad \forall\ k \in K - 1 \tag{13}$$

$$V_{k,s} \leq M_{i,j} \tag{14}$$

$$\frac{X_{k+1,1} + X_{k+1,2} + X_{k,8} + X_{k,9} + X_{k,10} + X_{k,11} + X_{k,12}}{T_{k+1,1} + T_{k+1,2} + T_{k,8} + T_{k,9} + T_{k,10} + T_{k,11} + T_{k,12}} \leq Q_{i,j} \tag{15}$$

$$D_p = \sum_{k \in K} \sum_{s \in S} T_{k,s} \qquad \forall\ p \in P \tag{16}$$

$$\sum_{p \in \dot{\mathrm{P}}} N_{p,i,j} = O_{i,j} \qquad \forall\ p \in \dot{\mathrm{P}} \tag{17}$$

$$T_{k,s} \geq \sum_{p \in \mathbb{P}} N_p + \sum_{p \in \mathrm{P}} N_p \tag{18}$$

Constraint (3) ensures that locomotives have the same velocity when they couple, similarly, Constraints (4), (5) and (6) guarantee that locomotives stay coupled during the segment.

Constraint (7) ensures that locomotives couple at the same location, where Š represents the set of segments from the first segment up to the locomotive's second coupling, conversely, Ş represents the set of segments from the first segment up to the locomotive's first coupling. For safe uncoupling Constraint (8) forces locomotives to end up 25 meters apart when they uncouple. Basic kinematic equations also had to be implemented, to ensure that the locomotive's movements are feasible, such as Constraint (9) which defines the final velocity of each segment and Constraint (10) which defines the length of each segment. Constraints (11) and (12) ensure that each route ends precisely at the destination station. Constraint (13) prevents infeasible headways, where Ŝ is the set that includes all segments of k up to coupling with $k+1$, and similarly, 𝕊 includes all segments of $k+1$ up to coupling with k. Some restrictions were implemented so that the simulated operation follows the operation guidelines of the actual line. Constraint (14) ensures that no segment is faster than the speed limit for that segment. And Constraint (15) guarantees that the average speed on a section is not higher than the average speed used at the existing line. Constraint (16) defines the decision variable D_p, which is the sum of the duration of all segments in a path. Constraint (17) ensures that the number of passengers going through all the paths departing from one station to another is equal to the demand in between these stations. Ṗ is the set of all paths that depart from station i toward station j. Constraint (18) guarantees that passengers trying to switch from a locomotive to another have time to do so; ℙ is the set of paths where passengers go through coupling at segment s and switch from locomotive $k+1$ to k, representing the people moving backwards; and ℙ is the set of paths where passenger switch from locomotive k to $k+1$.

3.4.2 Adapted model for longer lines. A passenger traveling from the first station to the 20th station in the line faces an array of 2^{19} potential paths, considering the choice of either stopping or skipping each station. This results in more than one million variables for this single origin-destination pair, rendering the initial model impractical for longer lines.

To address this, adaptations were implemented for longer lines. First, only two profiles were devised for origin-destination pairs longer than 14 stations. One profile involves passengers consistently switching locomotives and skipping every station, while the other profile entails passengers never switching locomotives and stopping at every other station. This adjustment significantly reduced the number of variables.

Second, the model was divided into two parts. The first model focuses on minimizing locomotive travel time by solely considering movement-related restrictions. Meanwhile, the second model concentrates on minimizing passenger travel time by optimally assigning passengers to paths determined in the first model.

Although this adapted model is expected to yield slightly inferior results due to a reduction in variables and the disregard for passenger demand when governing locomotive movement, it is also acknowledged that in a real-world application, locomotive movements might not dynamically adapt to changes in passenger demand, and passengers might not optimally switch locomotives. Hence, this adapted model was deemed suitable for the most challenging scenarios.

3.5 Solution method

The proposed model for demand-oriented train speed trajectory represents a non-smooth, non-convex programming problem with the objective function (1) and constraints (3–18). The non-smoothness stems from constraint (15), while the non-convexity arises from the nonlinear, non-convex objective function and constraints (9), (10) and (15). Solving mixed-integer nonlinear problems with these characteristics can be addressed through various approaches. In this study, we used *Filter*, a sequential quadratic programming algorithm solver provided by the *Neos Server* (Czyzyk *et al.*, 1998; Dolan, 2001; Gropp and Moré, 1997).

4. Case study with São Paulo's monorail line

Initially opened in August 2014, São Paulo's monorail line, known as Line 15 or the Silver Line, is projected to span 18 stations once completed. This line was chosen for research due to its current low demand, primarily because it is still under construction and has yet to connect high-demand stations. The non-stop system was considered more suitable for lines with moderate to low demand, largely due to limitations on passenger movement within the train, which restrict their ability to switch locomotives.

As of the data collection in November 2019, only seven stations were operational: Vila Prudente, Oratório, São Lucas, Camilo Haddad, Vila Tolstói, Vila União and Jardim Planalto. The information was collected at noon and produced two origin-destination (O-D) demand matrices, one of which is presented in Table 2.

To prevent the simulated line from exceeding the speed of the existing line, the speed profile was gathered. This data was used to determine the maximum and average speeds per section. The model's locomotive speeds per section were restricted by the speed profile applied in the current line.

The second data set is sourced from BART, providing detailed hourly ridership data based on origin-destination pairs for rail transit in San Francisco and surrounding counties. As this data set lacks locomotive movement information, a speed profile was constructed using estimated travel times from a GPS system and Line 15's speed profile.

BART's data set covers six distinct lines: yellow, blue, orange, green, red and beige. It spans every hour of the year, proving valuable for testing the non-stop design across various scenarios. According to the APTA's Ridership Reports (American Public Transportation Association, 2019), BART records an average weekday ridership of 3,760 passengers per mile, indicating a line with moderate demand by American standards.

4.1 Results and analysis

4.1.1 Operation results of line 15. For Line 15, two simulation cases were conducted, both with a 90-s headway. Case I simulated the VP-JP direction based on the noon O-D demand matrix. In Case II, the JP-VP direction was simulated with increased demand to emulate peak hour scenarios. According to CPTM, 33% of all passengers use the subway system during peak hours (Companhia do Metropolitano de São Paulo, 2019). Using this insight, a new O-D demand matrix was created for the second simulation.

Table 2
OD demand matrix

				Destination			
Origin	VP	OR	SL	CH	VT	VU	JP
Vila Prudente (VP)		12	6	3	8	14	36
Oratório (OR)			0	0	0	0	2
São Lucas (SL)				0	0	0	0
Camilo Haddad (CH)					0	0	0
Vila Tolstói (VT)						0	3
Vila União (VU)							3

Source: Authors' own work

The detailed passenger travel times for Case I are presented in Table 3. As anticipated, longer distances resulted in more time savings, with a few exceptions. For instance, VP-CH showed greater time savings compared to VP-VT. This disparity arose because in the latter, 75% of passengers had to stop at SL, while in the former, every passenger reached their destination directly. For similar reasons, VP-JP demonstrated lower time savings than OR-JP.

In Case II, the total time savings reduced to 27.61%, as indicated in Table 4. This decrease occurred because in Case I, only 14.94% of passengers had to stop at a secondary station during their trip. However, due to limitations on passenger flow through doors, 38.46% of passengers had to stop at a secondary station, and 5.38% had to stop at two secondary stations in Case II.

The observed time change in the initial and final trips verifies that the developed model effectively reduces the route's travel time without elevating the locomotive's speed beyond real-case scenarios. In essence, the results underscore the considerable reduction in passenger travel time on São Paulo's Line 15.

O-D	Total passenger travel time full-stop (seconds)	Total passenger travel time non-stop (seconds)	Time change (%)
VP-OR	1,865.04	1,864.63	−0.02
VP-SL	1,913.52	1,397.27	−26.98
VP-CH	1,279.26	833.69	−34.83
VP-VT	4,314.72	3,315.36	−23.16
VP-VU	9,212.14	5,890.13	−36.06
VP-JP	28,491.48	18,177.04	−36.20
OR-JP	1,212.02	776.10	−35.97
VT-JP	666.27	485.29	−27.16
VU-JP	310.26	310.08	−0.06
Total	49,264.71	33,049.60	−32.91

Source: Authors' own work

Table 3.
Time savings line VP-JP nonpeak period

O-D	Total passenger travel time full-stop (seconds)	Total passenger travel time non-stop (seconds)	Time change (%)
JP-VU	274.50	274.53	0.01
JP-VT	858.68	587.57	−31.57
JP-CH	337.84	219.71	−34.97
JP-OR	1,203.02	759.95	−36.83
JP-VP	43,931.80	32,154.13	−26.81
VU-OR	480.01	420.73	−12.35
VU-VP	8,127.12	5,622.71	−30.82
VT-OR	356.84	253.45	−28.97
VT-VP	8,311.35	5,129.03	−38.29
CH-VP	2,585.52	1,835.01	−29.03
CH-VP	5,525.00	4,259.21	−22.91
SL-VP	2,174.25	2,174.71	0.02
Total	74,165.93	53,690.74	−27.61

Source: Authors' own work

Table 4.
Time savings JP-VP peak period

The correlation between demand and time savings was validated through a comparison of Case I and Case II. With a 49.43% increase in demand, Case II's time savings were reduced by 16.10% in comparison to Case I. As this correlation could become more pronounced in congested scenarios. We considered lines with higher demand in the next section.

4.1.3 BART. Two lines were chosen from BART's data set, yellow line which is 27 stations long and blue line which is 18 stations long. As mentioned in Section 4, the mathematical model applied to these lines was less robust to deal with the intractability present in longer lines, hence, as shown in Table 5 the savings for these lines were smaller than the savings for Line 15.

The results show that the time savings for moderate demand is consistently above 20%, and even for the worst-case scenario where no transfers are allowed, meaning that passengers cannot switch locomotives to skip stations, there was still significant time savings as passengers are bound to skip every other station. Although in practice, they should yield savings similar to Line 15, the longer paths, which only had two profiles available as explained in subsection 3.4.2, yielded the worst results due to the limited number of transfers.

5. Conclusion

Reducing travel time was notably more effective for passengers on longer journeys during non-peak periods. Interestingly, there was still a significant decrease in travel time for those traveling short distances during peak hours, highlighting the substantial impact of dwelling time in stations on total travel time. The results obtained in this study demonstrate that implementing a non-stop system on Line 15 of São Paulo's subway system could reduce total passenger travel time by 27.61% to 32.91% on average demand days. It could also reduce total travel time by 6%–21% even during the busiest hours on BART's yellow and blue lines.

Given that locomotives equipped for a non-stop system are not yet available, this paper primarily focused on the theoretical and mathematical aspects of such an application. However, the introduction of such locomotives in the future would undeniably attract more passengers to the system, given the significant influence of travel time on their choice of transportation.

Table 5.
Time savings for BART

	Savings by path length					Line total	
Demand	≤ 4	>4 ≤ 8	>8 ≤ 12	>12 ≤ 16	>16	Savings	Passengers
(a) Blue line							
Average	−24%	−24%	−29%	−24%	−12%	−22%	209
Busiest hour	−16%	−18%	−11%	−3%	0%	−9%	580
Busiest hour NTA*	−12%	−13%	−4%	+1%	0%	−6%	580
(b) Yellow line							
	≤ 6	>6 ≤ 12	>12 ≤ 18	>18 ≤ 24	>24	Savings	Passengers
Average	−19%	−18%	−21%	−22%	−28%	−21%	490
Busiest hour	−21%	−22%	−16%	−9%	−11%	−16%	1067
Busiest hour NTA*	−12%	−17%	−12%	−8%	−11%	−12%	1067

Note: *No transfers allowed
Source: Authors' own work

Future research plans involve exploring the potential for adjusting the number of locomotives in a train to better accommodate demand. For instance, a station with high demand might have two locomotives stop, provided the departing train has three or more locomotives. This adaptability could improve efficiency and prevent bottlenecks. Additionally, alternating the use of non-stop with other systems like skip-stop and full-stop throughout the day and along the line, based on demand, could be explored. Moreover, the objective function could also consider metrics like energy consumption and resource usage, quantified by the number of locomotives used.

References

American Public Transportation Association (2019), “Ridership reports”, available at: www.apta.com/research-technical-resources/transit-statistics/ridership-report/ridership-report-archives/ (accessed 10 August 2023).

Bai, Y., Hu, Q., Ho, T., Guo, H. and Mao, B. (2019), “Timetable optimization for metro lines connecting to intercity railway stations to minimize passenger waiting time”, *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22 No. 1, pp. 1-12, doi: 10.1109/TITS.2019.2954895.

Brännlund, U., Lindberg, P., Nõu, A. and Nilsson, J.-E. (1998), “Railway timetabling using Lagrangian relaxation”, *Transportation Science*, Vol. 32 No. 4, pp. 358-369, doi: 10.1287/trsc.32.4.358, available at: www.jstor.org/stable/25768834

Carey, M. (1994), “Extending a train pathing model from one-way to two-way track”, *Transportation Research, Part B (Methodological)*, Vol. 28 No. 5, pp. 395-400, doi: 10.1016/0191-2615(94)90038-8.

Companhia do Metropolitano de São Paulo (2019), “Relatório integrado. São Paulo”, available at: https://transparencia.metrosp.com.br/dataset/relat%C3%B3rio-integrado-%E2%80%93-companhia-do-metr%C3%B4-de-s%C3%A3o-paulo (accessed 10 August 2023).

Czyzyk, J., Mesnier, M.P. and Moré, J.J. (1998), “The NEOS server”, *IEEE Computational Science and Engineering*, Vol. 5 No. 3, pp. 68-675.

Ding, L. and Zhang, N. (2016), “A travel mode choice model using individual grouping based on cluster analysis”, *Procedia Engineering*, Vol. 137, pp. 786-795, doi: 10.1016/j.proeng.2016.01.317.

Dolan, E. (2001), “The NEOS server 4.0 administrative guide”, *Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division*, Argonne National Laboratory.

Gropp, W. and Moré, J.J. (1997), “Optimization environments and the NEOS server”, *Approximation Theory and Optimization*, in Buhmann, M.D. and Iserles, A. (Eds), Cambridge University Press, Cambridge, pp. 167-182.

Guo, Y. and Zhang, C. (2022), “Near real-time timetabling for metro system energy optimization considering passenger flow and random delays”, *Journal of Rail Transport Planning and Management*, Vol. 21, p. 100292, doi: 10.1016/j.jrtpm.2021.100292.

Jiang, F., Cacchiani, V. and Toth, P. (2017), “Train timetabling by skip-stop planning in highly congested lines”, *Transportation Research Part B: Methodological*, Vol. 104, pp. 149-174, doi: 10.1016/j.trb.2017.06.018.

Jiateng, Y., Yang, L., D’Ariano, A., Tang, T. and Gao, Z. (2022), “Integrated backup rolling stock allocation and timetable rescheduling with uncertain time-variant passenger demand under disruptive events”, *INFORMS Journal on Computing*, Vol. 34 No. 6, pp. 3234-3258.

Jovanovic, D. and Harker, P.T. (1991), “Tactical scheduling of rail operations – the SCAN-I system”, *Transportation Science*, Vol. 25 No. 1, pp. 46-64, doi: 10.1287/trsc.25.1.46.

Kaspi, M. (2010), “Service oriented train timetabling”, (Master’s Thesis, Tel Aviv University).

Lee, K., Kim, W., Baek, J. and Kim, J. (2021), “A comparison into the factors affecting urban rail systems: local, express, and high-speed rail in tunnels at a great depth in a metropolitan area”, *Sustainability*, Vol. 13 No. 17, p. 9527, doi: 10.3390/su13179527.

Niu, H., Zhou, X. and Gao, R. (2015), "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, doi: 10.1016/j.trb.2015.03.004.

Nold, M. and Corman, F. (2021), "Dynamic train unit coupling and decoupling at cruising speed: systematic classification, operational potentials, and research agenda", *Journal of Rail Transport Planning and Management*, Vol. 18, p. 100241, doi: 10.1016/j.jrtpm.2021.100241.

Oh, S., Kim, S. and Hong, J. (2016), "Comprehensive analysis of the influence of door width on the passenger flow time on Korean urban railways", *J. Rail Rapid Transit*, Vol. 230 No. 6, pp. 1531-1539.

Parbo, J., Nielsen, O. and Prato, C. (2018), "Reducing passengers' travel time by optimising stopping patterns in a large-scale network: a case-study in the Copenhagen region", *Transportation Research Part A: Policy and Practice*, Vol. 113, pp. 197-212, doi: 10.1016/j.tra.2018.04.012.

Pei, M., Xu, M., Zhong, L. and Qu, X. (2023), "Robust design for underground metro systems with modular vehicles", *Tunnelling and Underground Space Technology*, Vol. 132, p. 104865, doi: 10.1016/j.tust.2022.104865.

Ren, X., Yin, J. and Tang, T. (2019), "Quantitative analysis for resilience-based urban rail systems: a hybrid knowledge-based and data-driven approach", pp. 3531-3538, doi: 10.3850/978-981-11-2724-3_0235-cd.

Scheepmaker, G.M., Goverde, R.M. and Kroon, L.G. (2017), "Review of energy-efficient train control and timetabling", *European Journal of Operational Research*, Vol. 257 No. 2, pp. 355-376.

Shi, J., Yang, J., Yang, L., Tao, L., Qiang, S., Di, Z., Amp. and Guo, J. (2023), "Safety-oriented train timetabling and stop planning with time-varying and elastic demand on overcrowded commuter metro lines", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 175, p. 103136, doi: 10.1016/j.tre.2023.103136.

Suh, W., Chon, K.-S. and Rhee, S.-M. (2002), "Effect of skip-stop policy on a Korean subway system", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1793 No. 1, pp. 33-39.

Sun, L., Gang Jin, J., Lee, D.-H., Axhausen, K.W. and Erath, A. (2014), "Demand-driven timetable design for metro services", *Transportation Research Part C: Emerging Technologies*, Vol. 46, pp. 284-299, doi: 10.1016/j.trc.2014.06.003.

Sun, P., Zhang, C., Jin, B., Wang, Q. and Geng, H. (2023), "Timetable optimization for maximization of regenerative braking energy utilization in traction network of urban rail transit", *Computers and Industrial Engineering*, Vol. 183, p. 109448, doi: 10.1016/j.cie.2023.109448.

Szpigel, B. (1973), "Optimal train scheduling on a single track railway", *Proceedings of the IFORS Conference Operational Research* '72, pp. 343-52.

Wang, D., D'Ariano, A., Zhao, J., Zhong, Q. and Peng, Q. (2021), "Integrated rolling stock deadhead routing and timetabling in urban rail transit lines", *European Journal of Operational Research*, Vol. 298 No. 2, doi: 10.1016/j.ejor.2021.05.053.

Wu, J., Kulcsár, B., Selpi, X. and Qu, X. (2021), "A modular, adaptive, and autonomous transit system (MAATS): an in-motion transfer strategy and performance evaluation in urban grid transit networks", *Transportation Research Part A: Policy and Practice*, Vol. 151, pp. 81-98, doi: 10.1016/j.tra.2021.07.005.

Wu, Q., Ge, X., Zhu, S., Cole, C. and Spiryagin, M. (2023), "Physical coupling and decoupling of railway trains at cruising speeds: train control and dynamics", *International Journal of Rail Transportation*, doi: 10.1080/23248378.2023.2169963.

Zhang, Z., Tafreshian, A. and Masoud, N. (2020), "Modular transit: using autonomy and modularity to improve performance in public transportation", *Transportation Research Part E: Logistics and Transportation Review*, Vol. 141, p. 102033, doi: 10.1016/j.tre.2020.102033.

Zhang, R., Yin, S., Ye, M., Yang, Z. and He, S. (2021), “A timetable optimization model for urban rail transit with express/local mode”, *Journal of Advanced Transportation*, Vol. 2021, p. 5589185, doi: 10.1155/2021/5589185.

Zhao, S., Yang, H. and Wu, Y. (2021), “An integrated approach of train scheduling and rolling stock circulation with skip-stopping pattern for urban rail transit lines”, *Transportation Research Part C: Emerging Technologies*, Vol. 128, p. 103170, doi: 10.1016/j.trc.2021.103170.

Zhou, X.S. and Zhong, M. (2007), “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, doi: 10.1016/j.trb.2006.05.003.

Further reading

Next-Future-Transportation-Inc (2021), “Next future transportation”, available at: www.next-future-mobility.com/

Silva, C., Saldiva Paulo Amato-Lourenço, L., Rodrigues-Silva, F. and Miraglia, S. (2012), “Evaluation of the air quality benefits of the subway system in Sao Paulo, Brazil”, *Journal of Environmental Management*, Vol. 101, pp. 191-196, doi: 10.1016/j.jenvman.2012.02.009.

Corresponding author

Zhipeng Zhang can be contacted at: zp.zhang@sjtu.edu.cn