Omega 144 (2026) 103593

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

A two-stage stochastic programming approach for strategic rolling stock scheduling in passenger railway transport☆

Siqiao Li [a,*], Rolf van Lieshout [b], Patrick Stokkink [c]

[a] Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628CN Delft, Netherlands
[b] Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, 5600MB Eindhoven, Netherlands
[c] Department of Engineering Systems and Services, Faculty of Technology, Policy and Management, Delft University of Technology, 2628BX Delft, Netherlands

ARTICLE INFO

Dataset link: Demand input data for computational experiments (Original data)

Keywords:
Railway planning
Rolling stock scheduling
Shunting
Uncertain demand
Stochastic programming

ABSTRACT

Conventional railway planning follows a sequential approach in which rolling stock scheduling is performed only after timetables are constructed, forcing early-stage line-planning decisions to rely on simplified rolling stock cost approximations. The lack of precise cost information hampers line-plan evaluation and limits the design of cost-efficient services. This study considers rolling stock scheduling in a strategic setting, aiming to estimate fleet size and operational costs for a given line plan without solving the timetabling problem. Since predefined train services are unavailable at this level, unlike in tactical planning, we introduce a network representation incorporating shunting movements like coupling and uncoupling. To account for demand uncertainty, a two-stage stochastic programming approach is proposed, in which first-stage decisions determine fleet acquisition and second-stage decisions optimize rolling stock circulation, including shunting, under each demand scenario. Two models are developed, differing in whether the fleet size is estimated using a flow-based or circulation-based approach. The applicability of the proposed models is demonstrated using instances derived from the Dutch Intercity network. Compared to a trip-based benchmark, the circulation-based model yields more accurate cost estimates, while short and simple cycles retain most of its value at much lower computational cost. Moreover, stochastic planning better captures the procurement–service trade-off under demand uncertainty and produces more robust fleet plans than a deterministic approach. Under strong emphasis on investment costs, fleet choices are dominated by flexible, cost-effective unit types. Allowing shunting enables flexible capacity redistribution, reducing fleet requirements and investment costs while maintaining comparable service levels.

1. Introduction

Public transport plays a critical role in modern societies, providing efficient means of mobility while alleviating issues such as traffic congestion, air pollution, energy consumption, and road accidents [1–3]. Among the various modes of public transport, railway systems stand out for their ability to accommodate high passenger volumes and provide long-distance services, which makes careful planning of these systems especially important. Railway transport planning typically covers different stages, including problems such as line planning, train timetabling, and rolling stock scheduling [4]. Line planning is a strategic problem that aims to design a set of service lines balancing operational efficiency and service quality. Based on a line plan, the timetabling problem determines the departure and arrival times for trains at different stations. Subsequently, the rolling stock scheduling problem seeks to find a set of efficient circulations that cover all the timetabled services while accounting for operational features such as the ability to couple and uncouple train units during the day to form different compositions.

In practice, the planning stages are addressed sequentially to ensure computational tractability, as an integrated approach would be impractical given that each stage is already computationally demanding on its own [5,6]. However, a disadvantage of the sequential approach is that the ultimate objectives of the planning process, such as service quality and costs, can only be evaluated after all planning steps have been solved, and therefore require approximations in the earlier planning steps. For example, the cost of a line plan cannot be evaluated until both the timetabling problem and the rolling stock scheduling problem are solved, which complicates the task of finding cost-efficient line plans.

☆ Area: Transportation and Logistics. This manuscript was processed by Associate Editor Dimitris Paraskevopoulos.

* Corresponding author.
E-mail addresses: S.Li-14@tudelft.nl (S. Li), r.n.v.lieshout@tue.nl (R. van Lieshout), p.s.a.stokkink@tudelft.nl (P. Stokkink).

https://doi.org/10.1016/j.omega.2026.103593
Received 16 April 2026; Accepted 21 May 2026
Available online 27 May 2026

Incorporating rolling stock considerations into line planning is challenging, as its cost depends strongly on operational decisions, such as when and where train units are (un)coupled. Therefore, existing studies typically rely on coarse cost estimations, where the rolling stock cost, when included, is modeled in a highly simplified manner. The most common approach is to assign a line-based cost parameter and multiply it by service frequency (see, e.g., Goossens et al. [7], Goerigk and Schmidt [8] and Canca et al. [9]), rather than derive cost from explicit train circulations. In addition, other important limitations remain. First, these studies assume a single train type per line and do not accommodate multiple compositions. Second, none of the models incorporate shunting movements, despite their central role in realistic rolling stock scheduling. Third, they generally assume static demand and ignore demand variations. These limitations imply that current line planning approaches do not provide sufficiently realistic cost estimates, particularly because key elements of rolling stock scheduling, such as composition choice, train circulation, and shunting, are only modeled at the tactical level. As a result, even the best-designed line plan may carry a substantial gap between its predicted and actual rolling stock cost.

Therefore, a fundamental question that arises in railway planning is: is it possible to estimate the rolling stock cost of a line plan, capturing operational details such as compositions and shunting, without explicitly designing a timetable? We answer this question affirmatively, by proposing the **strategic rolling stock scheduling problem**, which offers a more realistic cost estimation method during the early strategic stage of planning. This problem requires assessing line plans under alternative shunting-preference policies (ranging from extensive use of coupling/uncoupling to avoiding it during the core peak) and retains key features of a rolling stock scheduling problem, such as train unit types, compositions and shunting rules, but can be solved independently of the operated timetable. An additional feature that results from moving rolling stock scheduling from a tactical to a strategic setting is demand uncertainty, which must be accounted for when determining long-term rolling stock decisions.

In practice, the number of required train units, which is a main driver of the cost, is primarily determined by the peak period [10]. Thus, instead of designing a rolling stock schedule for an entire day, we develop a periodic rolling stock circulation for the peak period. We treat this peak period as a representative interval during which the line plan and demand are stable. This allows the frequency information to be translated into period-level service coverage, and hence into an estimate of the required number of train units. Compositions and shunting possibilities are modeled under the adopted shunting policy. To address demand uncertainty, we formulate the problem within a two-stage optimization framework, where the first-stage decisions determine the fleet size, and the second-stage decisions determine the operated circulations. The objective is to minimize the sum of the first-stage cost and the expected second-stage cost. We develop both flow-based and circulation-based models, which differ in whether the circulation set is specified a priori.

We develop both theoretical and numerical results. We show that the circulation-based model always leads to a higher or equal cost estimate than the flow-based model. In addition, we prove that if demand between pairs of stations is symmetric, there always exists an optimal flow solution without shunting operations between different lines. We conduct a series of numerical experiments based on instances generated from the Dutch Intercity Network. We demonstrate the advantages of a stochastic approach in achieving a more effective trade-off between procurement cost and service quality under demand uncertainty. Moreover, we compare the proposed two formulations and demonstrate their applicability.

In summary, this study addresses the aforementioned research gaps with the following contributions: (i) We introduce the strategic rolling stock scheduling problem and propose two two-stage stochastic programming models, distinguished by whether the circulation set is provided a priori. (ii) Through a comparison of the proposed two formulations against a trip-based benchmark, we demonstrate the applicability of the models, identifying the suitable model and determining the appropriate circulation set size for specific practical objectives. (iii) We provide managerial insights into how service priorities, demand uncertainty, shunting flexibility, and fleet heterogeneity shape strategic fleet sizing, cost estimation, and service quality.

The remainder of this paper is organized as follows. In Section 2, we review related literature on rolling stock scheduling and rescheduling and highlight the research gaps that we address. In Section 3, we introduce the strategic rolling stock scheduling problem. In Section 4, we provide two formulations and explore theoretical results to provide insights into their performance and relationships. In Section 5, we present the numerical experiments for two instances in the Netherlands. In Section 6, we draw conclusions and discuss ideas for future work.

2. Literature review

In railway planning, rolling stock scheduling determines how train units are combined into compositions and assigned to timetabled services, whereas rolling stock rescheduling adjusts an existing plan in response to operational disruptions. Existing studies predominantly address the problem at the tactical or operational level, with a predetermined timetable taken as input, and mainly follow either a flow-based or a path-based approach.

Flow-based approaches determine rolling stock allocations at an aggregate level by tracking train units or compositions through a network representation of the timetable. A common variant is to model train unit movements with a graph representation. For example, Budai et al. [11] study short-term rolling stock re-balancing using a network flow formulation with stations, trips, coupling, and uncoupling arcs. For problems with heterogeneous train unit types and composition-dependent shunting rules, flow-based models are often strengthened by explicitly modeling the composition assigned to each trip and their feasible composition changes. In this stream, Alfieri et al. [12] combine a multicommodity flow model with transition graphs whose nodes represent feasible compositions and whose arcs represent transitions. Similarly, Fioole et al. [13] formulate the composition assignment and transition decisions directly in an integer program, while also modeling station inventories and composition changes. Under a rescheduling context, Nielsen et al. [14] extend this formulation to real-time disruption management by fixing the circulations that have already been executed and rescheduling the remaining part of the timetable. Lin and Kwan [15] propose a two-phase strategy for train unit scheduling. A network-level model is first solved, neglecting the ordering of train units. Subsequently, a station-level model is solved to ensure detailed constraints regarding local shunting plans. More recently, Grimm et al. [16] compare two prevalent models utilized in rolling stock scheduling, demonstrating that the composition model is the most compact and fastest model when applied to the Netherlands Railways instances.

A limitation of flow-based formulations is that they do not explicitly represent the complete duty of each individual train unit. As a result, constraints that are naturally defined at the unit level, such as maintenance requirements or other path-dependent restrictions, are difficult to incorporate directly. To address this issue, path-based approaches model the full path of each train unit over the planning horizon. Instead of only determining aggregate flows or trip-level compositions, these formulations keep track of the complete sequence of trips operated by each individual unit. Based on this idea, Haahr et al. [17] compare a flow-based composition model with a path-based formulation, showing that the latter is particularly attractive when unit-specific constraints need to be included. Following this work, Haahr and Lusby [18] show that integrating rolling stock scheduling with train unit shunting can avoid infeasibilities arising from a sequential planning approach. Along

Table 1
Representative studies most relevant to the present work.

Publications	Planning stage	Topology	Unit types	(Un)Coupling	Shunting rules	Ordering
Alfieri et al. [12]	Tactical	Line	Multiple	✓	✓	✓
Fioole et al. [13]	Tactical	Line	Multiple	✓	✓	✓
Lin and Kwan [15]	Tactical	Network	Multiple	✓	–	–
Zhong et al. [19]	Tactical	Network	Multiple	✓	✓	–
Van Lieshout and Bouman [25]	Strategic	Network	Single	–	–	–
Bouman et al. [26]	Strategic	Network	Single	–	–	–
This study	Strategic	Network	Multiple	✓	✓	✓

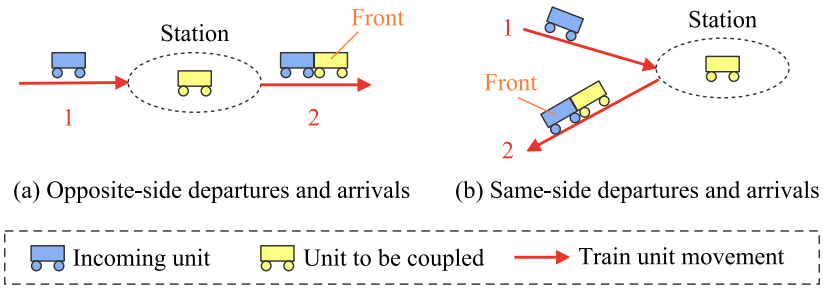


Fig. 1. Illustration of two shunting conditions in the coupling case.

this line, Zhong et al. [19] consider rolling stock scheduling with maintenance requirements, combining a flow-based first stage with a second-stage assignment model that checks whether the generated rolling stock schedules are feasible for individual units under maintenance restrictions. This study highlights the importance of an additional path layer to enforce unit-level feasibility.

Besides exact formulations, studies have also explored heuristic methods, particularly for real-time rescheduling where solution speed is critical. For example, Hoogervorst et al. [20] propose a variable neighborhood search heuristic for rolling stock rescheduling, where neighborhoods operate on train unit duties, single-type path adjustments, and composition changes.

While the above studies focus on self-propelled train units, their counterpart for locomotive-hauled trains corresponds to railcar and locomotive assignment problems. A commonly seen practice in freight transport is a railcar to train routing problem [21], followed by a locomotive assignment [22]. Other variants include the simultaneous assignment of locomotives and cars [23], operational car assignment [24], etc. A key difference is that these studies usually model coupling and decoupling as consist-level switching of locomotives or railcars, whereas self-propelled train-unit scheduling focuses on assigning ordered compositions to trips and enforcing feasible composition changes between consecutive trips.

By contrast, only a limited number of studies consider vehicle scheduling in the absence of a predetermined timetable. Van Lieshout and Bouman [25] study how vehicle schedules can be constructed directly from a line plan with the objective of minimizing the number of vehicles, allowing lines to be merged into larger circulations. They show that the problem is NP-hard even when at most two lines can be integrated and line compatibility is predetermined. Building on this setting, Bouman et al. [26] investigate a sequential approach that first determines a periodic vehicle schedule for a given line plan and only afterwards constructs a timetable. They show that this reordering can reduce computational complexity while still yielding optimal solutions in certain settings.

Table 1 summarizes the studies that are most directly comparable to our work in terms of rolling-stock decision structure, namely the treatment of shunting operations and unit ordering under a train-unit heterogeneity setting.

3. Problem description

This section introduces the problem setting. Section 3.1 defines train units, compositions and shunting movements, Section 3.2 explains the construction of the service network, and Section 3.3 describes how shunting possibilities are modeled based on this network. Finally, Section 3.4 introduces the optimization objective and cost parameters.

3.1. Train units and shunting movements

The ultimate goal of the strategic rolling stock scheduling problem is to estimate the cost and required number of train units of a line plan, while considering demand scenarios, different train unit types, and shunting rules. Maintenance and track allocation considerations are excluded in this strategic formulation, as the focus is on capturing the high-level composition and sizing decisions for the fleet.

A line plan defines a set of bidirectional lines, denoted as L, each serving a sequence of stops between two terminal stations once every H time units. If a line is operated multiple times per period, the line plan includes multiple copies of that line according to its frequency. All lines must be operated using rolling stock. A *train unit* is the smallest indivisible unit that constitutes rolling stock. Let $\mathcal{T}$ denote the set of train unit types, indexed by t. Each train unit type $t \in \mathcal{T}$ consists of a fixed number of carriages and provides a specific number of seats e_t. Compatible train units can be coupled together to form a *composition*, enabling additional capacity when required. Due to shunting regulations, the order of train units within a composition must be tracked; therefore composition ab differs from composition ba.

The composition of a train can be changed through coupling and uncoupling operations, defined as *shunting* movements. Practical rules need to be considered when defining possible composition changes at stations equipped with shunting facilities. In this regard, we assume that such stations permit shunting operations within the same line and between different lines. Generally, train units can only be coupled to the front of an incoming train and uncoupled from the rear of a leaving train [12]. In stations where train departures and arrivals occur on the same side in a direction of a line, the shunting rule is reversed, which allows train units to be coupled to the rear and uncoupled from the front. Fig. 1 illustrates these two cases for coupling movements, where the resulting trains are treated as two distinct compositions because the coupled train units appear in opposite orders. Another general rule is that coupling and uncoupling movements cannot be performed at the same time [11].

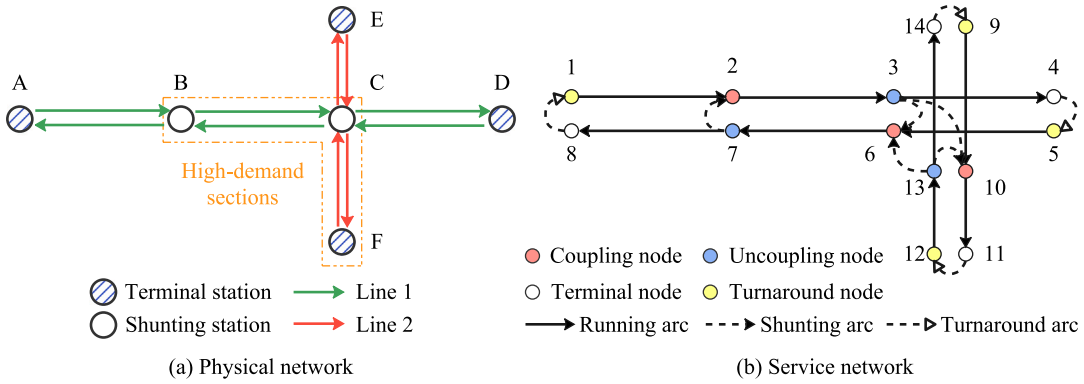


Fig. 2. Illustrative service network.

3.2. Network construction

Based on the line plan, we construct a service network $G = (N, A)$ that represents all services to be performed by the rolling stock. Stations that do not meet the necessary conditions for shunting operations are omitted. Each *station* is associated with a physical location, a set of bidirectional lines to which it is connected, and two travel directions on each line (upstream and downstream). A *node* $i \in N$ is defined as a tuple consisting of a location, a line, and a travel direction, i.e., it represents a station on a given line and in a specific direction, together with the rolling stock operation that may occur at that location. Depending on the operation, nodes are classified into four types: coupling, uncoupling, turnaround, and terminal nodes.

(1) Coupling and uncoupling nodes. At a shunting station, the two directional operations correspond to coupling and uncoupling. For instance, in the physical network shown in Fig. 2a, Line 1 follows the sequence A–B–C–D, where section B–C exhibits higher demand. Consequently, trains running from A toward D may require additional capacity at station B, whereas this is not necessary in the opposite direction. In the corresponding service network shown in Fig. 2b, nodes 2 and 7 both represent station B but in opposite directions: node 2 is a coupling node, as uncoupling is unnecessary when operating from A toward C, while node 7 is an uncoupling node, as coupling is unnecessary when operating from C toward A.

(2) Turnaround and terminal nodes. For terminal stations at the ends of a line, departures and arrivals occur on the same side of the line (e.g., Fig. 1b), corresponding to a turnaround movement, after which the train composition is reversed. Without loss of generality, we assume that terminal stations satisfy the necessary shunting conditions. For turnaround and shunting movements at a terminal location, we create two nodes: a turnaround node that captures the composition change and a terminal node where the composition remains unchanged.

The network contains three types of arcs: running arcs represent train movements between two stations, shunting arcs represent transitions of rolling stock from one composition to another, and turnaround arcs represent the reversal of rolling stock at the end of a line. Coupling nodes have incoming shunting arcs, while uncoupling nodes have outgoing shunting arcs.

The nodes and arcs in G are associated with a number of sets and parameters. The time duration of arc (i, j), which is either travel, shunting, or turnaround time, is denoted as τ_{ij}. The set of all shunting arcs is denoted as A^{S}. The set of outgoing arcs from node i is represented by A_i^+, while the set of incoming arcs to node i is denoted as A_i^-. Further, each running arc is associated with a number of required seats. Since the future demand is uncertain, we define a set of scenarios Ψ, and let d_{ij}^{φ} denote the required seats on arc (i, j) in scenario $\varphi \in \Psi$.

Due to platform-length limitations, train composition is restricted by an arc-specific upper bound on train length, expressed as the maximum number of train units that may be coupled on that arc. Let P_{ij} denote the set of all possible compositions that satisfy this upper-bound constraint for arc (i, j), and let n_p^t denote the number of t-type train units in composition $p \in P_{ij}$. Accordingly, the number of seats provided by composition p is $e_p := \sum_{t \in \mathcal{T}} e_t n_p^t$.

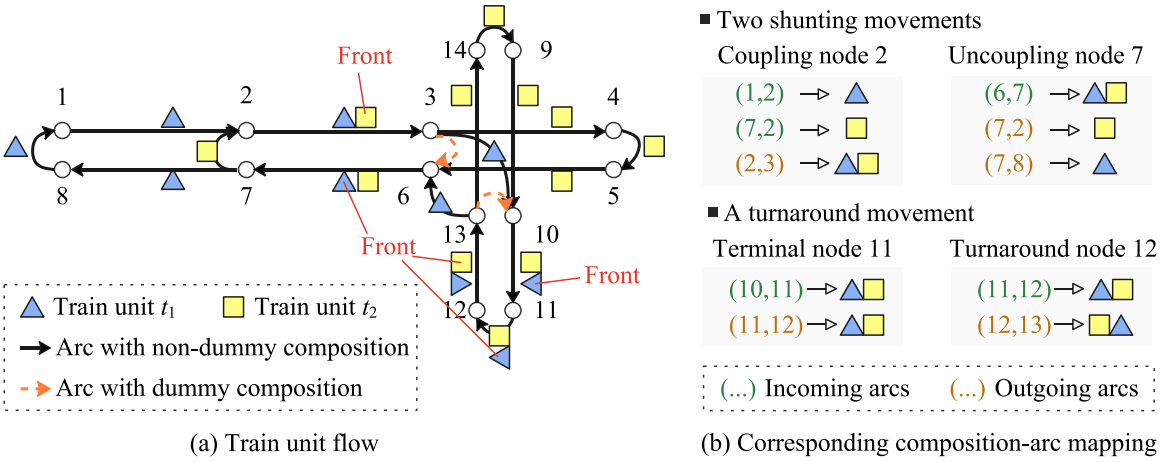
3.3. Shunting possibilities

To model shunting possibilities, we introduce a set of allowed *composition changes* Q_i for all $i \in N$, which is determined based on local shunting rules. Each composition change $q \in Q_i$ specifies the compositions assigned to the incoming and outgoing arcs of node i. At a coupling node, the composition change q defines compositions on all incoming arcs and a composition on the outgoing arc. At an uncoupling node, it defines the compositions on all outgoing arcs and one composition on the incoming arc. At turnaround and terminal nodes, q defines one composition for the incoming arc and one for the outgoing arc. Operator-specific shunting preferences can be captured through the station-wise sets Q_i, for example by allowing only in-line shunting movements or by restricting Q_i to identity changes (i.e., the same composition on the incoming and outgoing arcs) to represent a no-shunting policy.

Fig. 3 depicts a feasible rolling stock circulation for the instance in Fig. 2, along with the composition–arc mapping associated with the feasible flow. At node 2, a t_2 train unit coming from arc $(7, 2)$ is coupled to the front of a t_1 train unit, leading to a longer train performing arc $(2, 3)$. At node 3, the t_1 train unit is uncoupled from the rear. The above shunting movements are set based on the general shunting rules, and different shunting operations will be involved depending on the shunting rules. For the turnaround arc $(11, 12)$, the composition stays unchanged at node 11, after which it is reversed at node 12.

To capture the mapping between composition changes and the compositions assigned to adjacent arcs, we introduce a binary parameter $\mu_{p,ij}^{q+}$ which equals one if, after the implementation of a composition change q at node i, the train composition transitions to p when traversing outgoing arc (i, j). Similarly, we introduce a binary parameter $\mu_{p,ji}^{q-}$ which equals one when the composition of a train traversing an incoming arc (j, i) is p before a composition change q is applied at node i. This captures the feasible compositions and the corresponding composition changes dictated by shunting rules.

In addition to modeling station-specific shunting rules, another advantage of the composition-change concept is that it allows us to track the number of shunting movements. In particular, we introduce the parameter s_q for each composition change $q \in Q_i$ to denote the number of associated shunting operations, which takes the value 0 or 1. Specifically, $s_q = 1$ if a shunting movement occurs at node i, and $s_q = 0$ otherwise. Since it is possible that no shunting arc is used, i.e., the composition remains unchanged across consecutive arcs, we define a dummy composition p_0 to represent an unused shunting arc. Accordingly, we set $s_q = 0$ whenever all shunting arcs incident to node i are assigned to p_0. For example, in the feasible solution shown in Fig.



(a) Train unit flow (b) Corresponding composition-arc mapping

Fig. 3. Feasible rolling stock circulation and its associated composition–arc mapping.

3, a dummy composition is assigned to shunting arcs $(3,6)$ and $(13,10)$. Nevertheless, the number of shunting operations at nodes 6 and 13 equals one, because each node still has one shunting arc, $(3,10)$ and $(13,6)$, respectively, carrying a non-dummy composition.

3.4. Objective components and cost structure

The strategic rolling stock scheduling problem determines the fleet composition required to operate the services prescribed by a given line plan and, for each scenario, a feasible rolling stock circulation that minimizes the total expected cost. It should be noted that all costs are reported in monetary units for convenience of comparison; however, coefficients are stylized and used for preference analysis rather than operator-specific accounting.

Over-investing may lead to unnecessary expenses and low utilization of train units, while under-investing may result in seat shortages and reduced service quality. To balance these risks and to minimize expected costs under demand uncertainty, we account for both rolling stock costs and seat shortages. In particular, the main decision in the strategic rolling stock scheduling problem is how many train units of each type should be acquired. Here, parameter w_t^{P} is introduced to denote the purchase cost of a t-type train unit in the first stage.

To capture second-stage costs in each scenario, parameter w^{S} is introduced to denote the shunting cost per shunting operation, reflecting the operators' preference for fewer shunting movements and simpler operations. In addition, parameter $w_{t,ij}^{\mathrm{O}}$ is introduced to represent the operational cost of operating a t-type train unit on arc (i,j). To preserve service quality, seat shortages are penalized through parameter w_{ij}^{Q}, which assigns a quality-related penalty cost to each passenger affected on arc (i,j). The penalty reflects that either the discomfort from on-board crowding or the cost of switching to an alternative transport mode increase with travel distance and duration. Passengers traveling longer distances generally face higher burdens, whether due to extended time spent standing in a crowded carriage or higher fares for long-distance replacement services, than those on shorter trips. Consequently, arcs with longer travel times are assigned higher penalties. Finally, parameter λ is included as a weighting factor that specifies the relative importance of the first- and second-stage cost components.

4. Methodology

This section presents the proposed methodology, in which two variants of the strategic rolling stock scheduling problem are formally defined, differing in whether a predefined circulation set is employed. A flow-based model **M1** is introduced in Section 4.1, followed by a circulation-based model **M2** in Section 4.2, which yields a more accurate solution. Both models also cover the single-type fleet setting by restricting the set of available unit types. While both models presented in this section focus on the complete fleet from scratch, Section S1 of the supplementary material presents an additional-acquisition variant that accounts for a given initial inventory. The notation for models **M1** and **M2** is delineated in Tables 2 and 3.

Table 2
Generic notation for models **M1** and **M2**.

Notation	Description
Set	
L	Set of lines
Ψ	Set of scenarios, indexed by φ
N	Set of nodes, indexed by i,j
A	Set of arcs, indexed by (i,j)
A^{S}	Set of shunting arcs
$\mathcal{T}$	Set of train unit types, indexed by t
Q_i	Set of allowed composition changes for node i, indexed by q
P_{ij}	Set of allowed compositions for arc (i,j), indexed by p
Parameter	
λ	Weighting factor for the first-stage cost
w_t^{P}	Purchase cost of a t-type train unit
w^{S}	Shunting cost per shunting movement
$w_{t,ij}^{\mathrm{O}}$	Operational cost for a t-type train unit on arc (i,j)
w_{ij}^{Q}	Quality-related penalty for one seat shortage on arc (i,j)
e_t	Number of seats provided by a t-type train unit
e_p	Number of seats provided by a train with composition p
d_{ij}^{φ}	Number of seats required for arc (i,j) in scenario φ
τ_{ij}	Travel time for arc (i,j), measured in minutes
H	Time period, measured in minutes
s_q	Number of shunting operations involved in composition change q
n_p^t	Number of t-type train units coupled in a train of composition p
$\mu_{p,ij}^{q+}/\mu_{p,ij}^{q-}$	Binary parameter equal to 1 if composition change q defines a composition p for outgoing/incoming arc (i,j), and 0 otherwise
Decision variable	
$y_{p,ij}^{\varphi}$	Binary variable equal to 1 if the composition of a train performing arc (i,j) is p, and 0 otherwise
$v_{q,i}^{\varphi}$	Binary variable equal to 1 if composition change q is operated at node i, and 0 otherwise
n_t	Integer variable denoting the number of t-type train units purchased in the first stage

4.1. A flow-based model formulation

This section introduces a flow-based model that accounts for multiple train unit types and explicitly model the ordering of train units

$$\min \quad \lambda \sum_{t\in\mathcal{T}} w_t^{\mathrm{P}} n_t + \sum_{\varphi\in\Psi} \frac{1}{|\Psi|} \left(\sum_{i\in N} \sum_{q\in Q_i} w^{\mathrm{S}} s_q v_{q,i}^{\varphi} + \sum_{(i,j)\in A} \sum_{p\in P_{ij}} \sum_{t\in\mathcal{T}} w_{t,ij}^{\mathrm{O}} n_p^t y_{p,ij}^{\varphi} + \sum_{(i,j)\in A} w_{ij}^{\mathrm{Q}} \epsilon_{ij}^{\varphi} \right), \tag{1a}$$

$$\text{s.t.} \quad \sum_{p\in P_{ij}} e_p y_{p,ij}^{\varphi} + \epsilon_{ij}^{\varphi} \geq d_{ij}^{\varphi} \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi, \tag{1b}$$

$$\sum_{p\in P_{ij}} y_{p,ij}^{\varphi} = 1 \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi, \tag{1c}$$

$$\sum_{q\in Q_i} v_{q,i}^{\varphi} = 1 \qquad \forall i\in N,\ \forall \varphi\in\Psi, \tag{1d}$$

$$y_{p,ij}^{\varphi} = \sum_{q\in Q_i} \mu_{p,ij}^{q+} v_{q,i}^{\varphi} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{1e}$$

$$y_{p,ij}^{\varphi} = \sum_{q\in Q_j} \mu_{p,ij}^{q-} v_{q,j}^{\varphi} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{1f}$$

$$\sum_{(i,j)\in A} \sum_{p\in P_{ij}} n_p^t \frac{\tau_{ij}}{H} y_{p,ij}^{\varphi} \leq n_t \qquad \forall t\in\mathcal{T},\ \forall \varphi\in\Psi, \tag{1g}$$

$$y_{p,ij}^{\varphi} \in \{0,1\} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{1h}$$

$$v_{q,i}^{\varphi} \in \{0,1\} \qquad \forall i\in N,\ \forall q\in Q_i,\ \forall \varphi\in\Psi, \tag{1i}$$

$$n_t \in \mathbb{Z} \qquad \forall t\in\mathcal{T}, \tag{1j}$$

$$\epsilon_{ij}^{\varphi} \in \mathbb{Z} \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi. \tag{1k}$$

Box I.

Table 3
Additional notation for **M2**.

Notation	Description
C	Set of circulations, indexed by c
C_{ij}	Set of circulations containing arc (i,j)
A_c	Set of arcs included in circulation c
θ	Maximum number of lines permitted in any circulation
τ_c	Round trip time of circulation c
k_c	Number of trains required to operate circulation c in every period
$z_{t,c}^{\varphi}$	Integer variable denoting the number of t-type train units for each composition executing circulation c in scenario φ
ϵ_{ij}^{φ}	Integer variable denoting the seat shortage in scenario φ on arc (i,j)

in compositions. Section 4.1.1 first presents the model formulation. Subsequently, Section 4.1.2 examines its structural properties.

4.1.1. Model formulation

Integer variable n_t is introduced to denote the number of t-type train units purchased in the first stage. Three groups of variables are involved to model the second-stage cost for all scenarios in the future. Specifically, to capture the change in the position of train units incurred by shunting operations, we introduce two groups of binary variables: $y_{p,ij}^{\varphi}$, which equals one if the composition of a train executing arc (i,j) is p in scenario φ, and $v_{q,i}^{\varphi}$, which equals one if composition change q is operated at node i in scenario φ. Additionally, we denote the seat shortage in scenario φ in the second stage by ϵ_{ij}^{φ}.

Model **M1** is formulated as follows (see the Eqs. (1a)–(1k) in Box I):

The objective (1a) minimizes the sum of the λ-weighted first-stage purchase cost and the average expected second stage cost across all scenarios. Constraint (1b) enforces demand satisfaction, stating that the demand on each arc is either served by trains in the assigned composition or left unserved. Constraint (1c) guarantees that exactly one composition is assigned to each arc, with the practical rules defining feasible train formations embedded in the construction of the composition set. Constraint (1d) ensures that only one composition change is selected at each node, where the local shunting rules are implicitly incorporated through the construction of the composition-change set. Constraints (1e) and (1f) relate the selected compositions to the selected composition changes. Specifically, the composition of a train executing the incoming and outgoing arcs of a coupling or uncoupling node must be consistent with the composition change applied at that node. Taken together, Constraints (1c)–(1f) ensure train unit flow balance across the service network in each scenario: Constraint (1c) determines the flow on each arc; Constraint (1d) determines the incoming and outgoing flows at each node; and Constraints (1e) and (1f) enforce consistency between arc flows and node flows. Constraint (1g) restricts the number of used train units in each scenario to the number of purchased train units for each type. It is important to note that the left-hand side of this constraint provides a lower bound on the number of t-type train units needed to operate the schedule once per period. Variable definitions are provided by constraints (1h)–(1k).

4.1.2. Structural properties of the flow-based model

This section examines the structural properties of the flow-based model formulation that arise under certain network conditions. Note that the results in this section are derived for the strategic problem considered in this paper, where the rolling stock schedule is constructed on a periodic service network representation and is not tied to a specific timetable. Consequently, feasibility is evaluated with respect to the service network and the associated shunting rules.

In particular, cross-line shunting operations tightly couple different lines, making rolling stock planning more intertwined and hindering the possibility of decomposing the problem by line. Since such movements introduce operational complexity and safety concerns, operators typically prefer to avoid them whenever possible. The following analysis shows that, under symmetric demand patterns, an optimal flow-based plan naturally avoids cross-line shunting altogether.

Symmetry arises naturally when upstream and downstream services exhibit identical demand patterns. Before presenting the main result, Definition 1 introduces the notions of conjugate nodes and arcs.

Definition 1 (*Conjugate Nodes and Conjugate Arcs*)**.** For each node i, let $\bar{i}$ denote its *conjugate*: the node representing the same line and station but serving the opposite direction. Similarly, the *conjugate* of an arc (i, j) is the arc $(\bar{j}, \bar{i})$, denoted $\overline{(i, j)}$, whenever this arc exists.

According to Definition 1, running arcs always have conjugates, and inter-line shunting arcs have conjugates corresponding to the reverse connection between the two lines. Turnaround arcs and within-line shunting arcs do not have conjugates. Using these concepts, symmetric instances are next defined in Definition 2.

Definition 2 (*Symmetric Instances and Solutions*)**.** An instance is said to be *symmetric* if the passenger demand on every running arc equals the passenger demand on its conjugate. A solution is *symmetric* if, for every train unit type, the number of train units operating on any running arc is equal to the number operating on its conjugate arc.

Proposition 1 shows that, under symmetry, the flow-based model always admits symmetric optimal solutions.

Proposition 1. *If an instance is symmetric, the flow-based model* **M1** *admits an optimal solution that is symmetric.*

The proof is given in Appendix A.

Building on the property introduced by Proposition 1, Proposition 2 demonstrates that cross-line shunting arcs are never required in optimal solutions of symmetric instances.

Proposition 2. *If an instance is symmetric, the flow-based model* **M1** *admits an optimal solution that does not use any shunting arcs between different lines.*

The proof is given in Appendix A, which show that in a symmetric instance any use of cross-line shunting arcs can be symmetrically rerouted through same-line shunting arcs without increasing cost. In a timetable-based setting, additional timing constraints could prevent such rerouting; this possibility is outside the scope of the strategic abstraction studied here.

Overall, the result implies that when a flow-based model is used, optimal plans under symmetric settings never require cross-line shunting. This allows the decision makers to exclude cross-line shunting arcs from the network entirely, simplifying the formulation. For asymmetric cases, the above propositions do not generally hold, as directional demand imbalances can be exploited by assigning larger compositions to high-demand directions and smaller ones to low-demand directions across lines, yielding better demand–capacity matching.

4.2. A circulation-based model

The model presented in Section 4.1 is a flow-based formulation that approximates the number of train units required to cover a line plan by aggregating the operating times of all train units across the service network. In this section, we take an alternative approach by introducing a predefined set of circulations, allowing us to move from an aggregate flow representation to a detailed level of train paths. Section 4.2.1 defines the concept of a circulation and explains how circulations are generated, and Section 4.2.2 presents the resulting model formulation.

4.2.1. Definition and generation of circulations

A *circulation* is a simple cycle in the train service network with no arc traversed more than once, that can be operated periodically by a train unit. Compared with the flow-based model, a circulation-based formulation provides a more accurate estimate of the number of train units required, as it explicitly models the path followed by each train through the network.

Denote the set of circulations as C, and denote those circulations containing arc (i, j) as C_{ij}. Denote the set of arcs included in $c \in C$ as A_c. The round trip time of circulation c, denoted as τ_c, is computed by summing the travel times for all arcs that comprise c, i.e., $\tau_c := \sum_{(i,j)\in A_c} \tau_{ij}$. For circulations with a round trip time larger than H, multiple trains need to be assigned to ensure that the arcs are covered in each period. The number of trains needed to operate circulation c in every period, denoted as k_c, can be calculated by $k_c := \left\lceil \frac{\tau_c}{H} \right\rceil$.

We use an example illustrated in Fig. 4 to show the difference in calculating the number of trains between flow-based and circulation-based solutions, as well as the effect of imposing a no-shunting restriction. In the service network depicted in Fig. 4a, each arc has a travel time of 1 time unit, and every service should be operated once every $H = 5$ time units. Fig. 4b–c show a flow-based solution and a circulation-based solution that use the same number of train units on each arc. However, the flow-based solution underestimates the number of required train units by one. Fig. 4d shows the no-shunting solution, where three train units are required, but each unit must remain on a fixed service path, leading to higher unit-kilometer operating cost.

Apart from the advantage of accurately estimating the unit number, introducing a circulation set enables operators to define feasible and viable circulations, which may, for example, include only one or two lines.

Recall that L denote the set of lines in the complete network, and let θ be the maximum number of lines permitted in any circulation. Algorithm 1 outlines the procedure for constructing the candidate circulation set by iterating over each subset of lines containing θ line. For each such subset, we consider the subgraph induced by these lines. All elementary cycles in this subgraph are enumerated using Johnson's algorithm [27], a depth-first search-based enumeration procedure that uses a blocking-unblocking mechanism to list every elementary directed circuit exactly once. Its running time is $O((c + 1)(n + e))$, where c, n, and e denote the numbers of cycles, nodes, and edges in the graph, respectively.

Algorithm 1 Generation of the circulation set

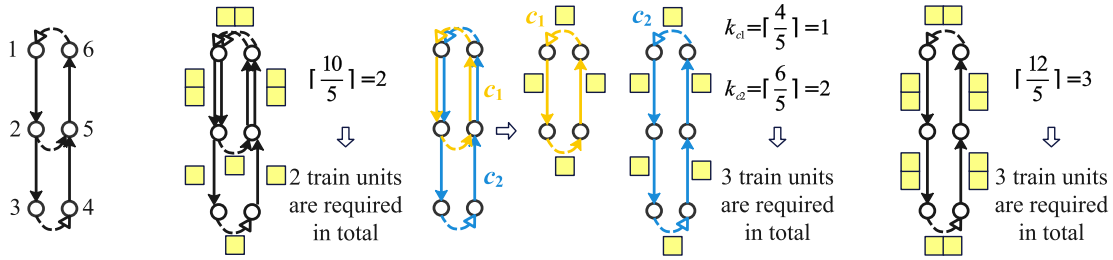
```
Input: Complete service network G, set of lines L, parameter θ
Output: Circulation set C
C ← ∅
for all subsets L̃ ⊆ L with |L̃| = θ do
    Construct the subgraph G̃ induced by lines in L̃
    C(L̃) ← set of all elementary circuits in G̃ obtained by Johnson's algorithm
    C ← C ∪ C(L̃)
end for
return C
```

4.2.2. Model formulation

Given a circulation set, this section introduces model **M2**, which provides a more accurate representation of rolling stock operations. Apart from variables $y^{\varphi}_{p,ij}$, $v^{\varphi}_{q,i}$, n_t, and ϵ^{φ}_{ij} already defined in model **M1**, an additional integer variable $z^{\varphi}_{t,c}$ is introduced to denote the number of t-type train units for each composition executing circulation c in scenario φ. This variable enables the disaggregation of train units flowing on each arc into the specific circulations that cover that arc, thereby providing a more detailed and operationally meaningful assignment. The proposed model is formulated as follows (see the Eqs. (2a)–(2m) in Box II):

Constraints (2a)–(2f) are identical to those expressed by constraints (1a)–(1f) in **M1**. Constraint (2g) ensures that the number of t-type train units defined by circulations matches the quantity specified by compositions. The inventory constraint (2h) calculates the number of t-type train units in scenario φ by summing up all the t-type train units used across all circulations. The domains of decision variables are defined by constraints (2i)–(2m).

Comparing the two proposed formulations, Proposition 3 states that the objective value of the circulation-based model always provides an upper bound on the objective value of the flow-based model.



(a) Service network with τ=1 and H=5 (b) Flow-based solution requiring two train units (c) Circulation-based solution requiring three train units (d) No-shunting solution requiring three train units

Fig. 4. Comparison of flow-based and circulation-based solutions.

$$\min \quad \lambda \sum_{t\in\mathcal{T}} w_t^{\mathrm{P}} n_t + \sum_{\varphi\in\Psi} \frac{1}{|\Psi|} \left(\sum_{i\in N} \sum_{q\in Q_i} w^{\mathrm{S}} s_q v_{q,i}^{\varphi} + \sum_{(i,j)\in A} \sum_{p\in P_{ij}} \sum_{t\in\mathcal{T}} w_{t,ij}^{\mathrm{O}} n_p^t y_{p,ij}^{\varphi} + \sum_{(i,j)\in A} w_{ij}^{\mathrm{Q}} \epsilon_{ij}^{\varphi} \right), \tag{2a}$$

$$\text{s.t.} \quad \sum_{p\in P_{ij}} e_p y_{p,ij}^{\varphi} + \epsilon_{ij}^{\varphi} \geq d_{ij}^{\varphi} \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi, \tag{2b}$$

$$\sum_{p\in P_{ij}} y_{p,ij}^{\varphi} = 1 \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi, \tag{2c}$$

$$\sum_{q\in Q_i} v_{q,i}^{\varphi} = 1 \qquad \forall i\in N,\ \forall \varphi\in\Psi, \tag{2d}$$

$$y_{p,ij}^{\varphi} = \sum_{q\in Q_i} \mu_{p,ij}^{q+} v_{q,i}^{\varphi} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{2e}$$

$$y_{p,ij}^{\varphi} = \sum_{q\in Q_j} \mu_{p,ij}^{q-} v_{q,j}^{\varphi} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{2f}$$

$$\sum_{c\in C_{ij}} z_{t,c}^{\varphi} = \sum_{p\in P_{ij}} n_p^t y_{p,ij}^{\varphi} \qquad \forall (i,j)\in A,\ \forall t\in\mathcal{T},\ \forall \varphi\in\Psi, \tag{2g}$$

$$\sum_{c\in C} k_c z_{t,c}^{\varphi} \leq n_t \qquad \forall t\in\mathcal{T},\ \forall \varphi\in\Psi, \tag{2h}$$

$$y_{p,ij}^{\varphi} \in \{0,1\} \qquad \forall (i,j)\in A,\ \forall p\in P_{ij},\ \forall \varphi\in\Psi, \tag{2i}$$

$$v_{q,i}^{\varphi} \in \{0,1\} \qquad \forall i\in N,\ \forall q\in Q_i,\ \forall \varphi\in\Psi, \tag{2j}$$

$$z_{t,c}^{\varphi} \in \mathbb{Z} \qquad \forall c\in C,\ \forall t\in\mathcal{T},\ \forall \varphi\in\Psi, \tag{2k}$$

$$n_t \in \mathbb{Z} \qquad \forall t\in\mathcal{T}, \tag{2l}$$

$$\epsilon_{ij}^{\varphi} \in \mathbb{Z} \qquad \forall (i,j)\in A,\ \forall \varphi\in\Psi. \tag{2m}$$

Box II.

Proposition 3. *Let $Z_{\mathbf{M}i}$ denote the optimal objective value of model* **M***i, where $i\in\{1,2\}$. For any instance, it holds that $Z_{\mathbf{M1}} \leq Z_{\mathbf{M2}}$.*

The proof is given in Appendix A, which demonstrates that every feasible **M2** solution can be mapped to a no-worse feasible **M1** solution.

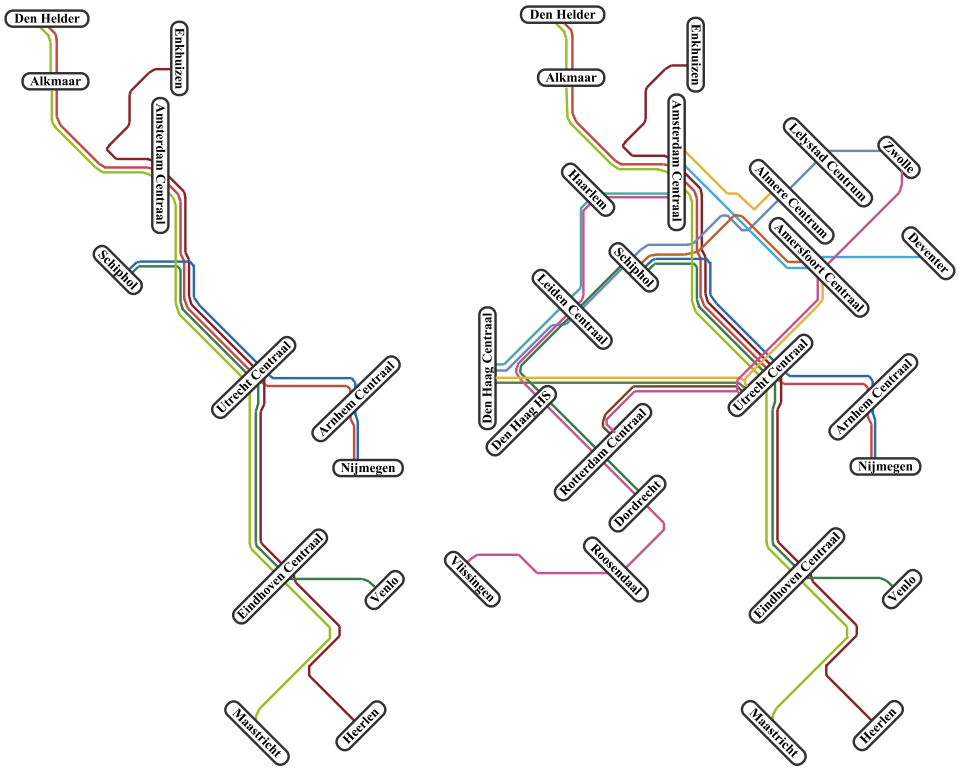
It should be noted that Propositions 1 and 2 apply only to the flow-based model; in the circulation formulation, combining arcs from different lines may lower the rounded-up train-unit requirement, meaning that **M2** can exploit cross-line circulations to obtain a strictly smaller inventory cost.

5. Computational results

In this section, we present the results of a series of numerical experiments to demonstrate the effectiveness of the proposed formulations. Section 5.1 introduces the test instances. Section 5.2 compares the two strategic models against a trip-based benchmark to assess the accuracy of their cost estimates. Section 5.3 further compares the two strategic models. Section 5.4 evaluates the benefit of using a stochastic programming approach by comparing it with a deterministic formulation. Section 5.5 explores the benefit of introducing multiple train unit types. Section 5.6 analyzes how strategic decisions change under different planning preferences regarding first-stage procurement and second-stage operational performance. Finally, Section 5.7 examines the benefit of allowing shunting movements for strategic procurement and service outcomes.

5.1. Instances

Two line plans with different sizes are generated based on the 2024 Intercity service network operated by Netherlands Railways (Nederlandse Spoorwegen, NS), the largest railway operator in the Netherlands [28]. These line plans are depicted in Fig. 5. The expanded network includes 16 lines, whereas the partial network includes 5 lines. The period length is set to 30 min. This choice is motivated by the fact



(a) Partial network. (b) Expanded network.

Fig. 5. Selected part of intercity network operated by NS.

that the considered Intercity line plans operate predominantly at two trains per hour [28], implying that $H = 30$ directly corresponds to one planned departure per line per period. The number of lines included in a circulation is initially limited to 2 lines, as public transport operators prefer short circulations involving one or two lines to simplify operations and reduce dependencies within the network [29,30].

To create a variety of demand scenarios accounting for its uncertainty we consider a total of 120 scenarios among which a reasonable sampled scenario set size is further analyzed in Section 5.4.1. The baseline demand requirement is specified for each arc under three demand patterns: a symmetric pattern and two asymmetric patterns corresponding to morning and evening peak hours. These baseline values are represented at a coarse level as the required number of carriages and are back-engineered from the rolling stock schedule used by NS. The carriage-based demand is then converted into seat-based demand for the stochastic models using a fixed seat-per-carriage factor. Under each demand pattern, 40 scenarios are then generated by introducing random variations to the baseline demand, with fluctuations ranging from −25% to +25% of the original values, consistent with transport-planning settings in which demand fluctuations of 20%–30% are commonly examined [31,32]. The demand data used in the experiments are publicly available at https://github.com/siqiaoli0293/strategic-rolling-stock-data.

We consider four types of train units that are currently used in the Netherlands: ICM-3, ICM-4, VIRM-4, and VIRM-6, abbreviated as I3, I4, V4, and V6, respectively. The first two types are single-deck train units, while the latter two are double-deck train units. Their characteristics are summarized in Table 4. The cost parameters for each train unit type are obtained by multiplying the number of seats by a specified cost per seat. Specifically, let $\bar{w}^{\mathrm{P}}$ and $\bar{w}^{\mathrm{O}}$ denote the unit purchase cost per seat and the unit operational cost per seat per minute, respectively. The purchase and operational cost parameters are then given by $w^{\mathrm{P}}_t = \bar{w}^{\mathrm{P}} e_t$ and $w^{\mathrm{O}}_{t,ij} = \bar{w}^{\mathrm{O}} \tau_{ij} e_t$. The seat shortage penalty for arc (i, j) is computed as $w^{\mathrm{Q}}_{ij} = \bar{w}^{\mathrm{Q}} \tau_{ij}$, where $\bar{w}^{\mathrm{Q}}$ is the unit penalty per passenger per minute. In the baseline setting, we set $\bar{w}^{\mathrm{P}} = 0.01$ and $\bar{w}^{\mathrm{O}} = 0.00017$. A service-first policy is adopted with $\bar{w}^{\mathrm{Q}} = 100$, indicating that satisfying passenger demand is prioritized, and investment and operating costs are minimized subject to this priority through a sufficiently large penalty on seat shortages. The weighting parameter λ is set to 1 in the baseline configuration, a choice that is further supported by the results in Section 5.6, where the solution is shown not to be overly sensitive to this parameter.

Table 4
Characteristics of train unit types.

Type	# Seats (e_t)	# Carriages	w^{P}_t	$\bar{w}^{\mathrm{O}} e_t$
VIRM4	406	4	4.0	0.07
VIRM6	597	6	6.0	0.10
ICM3	228	3	1.5	0.04
ICM4	299	4	2.0	0.05

Trains can only be formed by coupling train units of the same deck type; thus, types I3 and I4 can be coupled, as can V4 and V6, but ICM units cannot be coupled with VIRM units. Another requirement for forming trains is the maximum length of a train, which is set to 12 carriages, following common practice in the literature [12,29,33]. The above restrictions result in 21 candidate compositions for P_{ij}. Under the baseline setting of $\bar{w}^{\mathrm{Q}} = 100$, any composition for an arc with insufficient capacity is never optimal. Thus, to reduce computational burden, for each scenario φ, the composition set P_{ij} is replaced by a reduced subset P^{φ}_{ij} if there exists at least one composition $p \in P_{ij}$ such that $e_p \geq d^{\varphi}_{ij}$. In this case, $P^{\varphi}_{ij} := \{ p \in P_{ij} \mid e_p \geq d^{\varphi}_{ij} \}$. Otherwise, if no such composition exists, we set $P^{\varphi}_{ij} = P_{ij}$, and correspondingly $\epsilon^{\varphi}_{ij} > 0$. The full composition set is retained in the deterministic–stochastic comparison experiment in Section 5.4, where the shortage-penalty parameter is varied to explicitly examine the resulting cost–service trade-off.

Table 5
Sensitivity of the strategic objective estimate to the turnaround time.

τ^T	**M1**						**M2**					
	Obj.	Gap^{Obj}	Z^P	Z^S	Z^O	Time (s)	Obj.	Gap^{Obj}	Z^P	Z^S	Z^O	Time (s)
7	147.1	9.7%	90	12	45.1	0.1	153.1	6.0%	96	12	45.1	0.1
9	147.7	9.3%	90	12	45.7	0.1	153.7	5.6%	96	12	45.7	0.1
11	153.6	5.7%	96	12	45.6	0.1	154.3	5.2%	96	12	46.3	0.1
13	154.9	4.9%	96	12	46.9	0.1	154.9	4.9%	96	12	46.9	0.1
15	155.5	4.5%	96	12	47.5	0.1	155.5	4.5%	96	12	47.5	0.1
17	156.1	4.1%	96	12	48.1	0.1	156.1	4.1%	96	12	48.1	0.2
19	156.7	3.7%	96	12	48.7	0.1	162.7	0.1%	102	12	48.7	0.1
Benchmark model M3: objective = 162.8, Z^P = 108, Z^S = 0, Z^O = 54.8, runtime = 1.4 s.												

Notes: All models are solved to optimality. τ^T = turnaround time. Gap^{Obj} is computed relative to the objective of the benchmark M3. Z^P = first-stage purchase cost. Z^S = second-stage shunting cost. Z^O = second-stage operational cost.

All computational experiments were conducted on the DelftBlue supercomputer [34]. Gurobi 12.0.0 is used as the optimization solver with a 1-hour time limit and a termination gap of 0.1%, a setting that balances computational tractability with the level of accuracy required for strategic decision making.

5.2. Comparison against a trip-based benchmark

This section compares the two strategic models with a higher-resolution benchmark in which rolling stock schedules are optimized on an explicit timetable. Intercity Line 3500 between Schiphol and Venlo in Fig. 5 is used as the case study, since extending the benchmark to the full network would require timetable synchronization across interacting lines together with a combinatorial number of predecessor–successor trip connections for coupling and uncoupling movements. Exact studies that jointly capture both network-level interactions and such composition-change decisions remain limited in the literature.

In the benchmark, a regular timetable is constructed with a 30-minute headway, obtained by dividing the period length H by the line frequency. All trips traversing the same arc are assumed to face the same demand requirement as in the strategic setting. The minimum operational time for a shunting movement is set to 30 min. Given this regular timetable, the rolling stock schedule is determined for each trip on each arc using formulation **M3** introduced in Appendix C. For the strategic circulation model **M2**, all feasible circulations are enumerated, yielding three circulations in total.

The main distinction between the strategic models, **M1** and **M2**, and the trip-based benchmark **M3** lies in how trip connections are represented. In **M1** and **M2**, trip connections at shunting and terminal stations are approximated by shunting and turnaround arcs with fixed travel times. In **M3**, these connections are modeled explicitly for each trip pair under the timetable, leading to heterogeneous connection times across stations. To assess the impact of this approximation, **M1** and **M2** are solved with turnaround time τ^T ranging from 7 min, i.e., the baseline value, to 19 min, corresponding to the largest connection time in the benchmark timetable. The results are reported in Table 5, where all three models are solved to optimality.

In terms of cost estimation, **M2** is closer to **M3** than **M1** for most values of τ^T, in line with Proposition 3. This supports the view that tracking complete unit trajectories yields estimates that are closer to the trip-based benchmark. The smallest estimation error is attained at $\tau^T = 19$, indicating that both the circulation-based representation and the assumed connection time affect estimation accuracy. The value of τ^T that enables the strategic model to attain the smallest estimation error, however, is not universal, since the objective difference between a strategic model (**M1** or **M2**) and the benchmark **M3** also depends on the constructed timetable. This dependence is particularly relevant at the network level, where realized connection times may vary due to interactions across lines. Hence, when detailed timetable information is unavailable, using a conservative turnaround-time estimate is a practical choice to avoid overstating rolling stock flexibility. Under such conservative choices, a strategic model always leads to an underestimate of the trip-based benchmark cost.

Comparing the two strategic models from the perspective of how their objective values vary with τ^T, the objective of **M1** increases gradually, whereas that of **M2** exhibits a piecewise behavior. This difference arises because, in **M1**, the rounding-up effect of cycle times is aggregated over arcs, leading to a smoother change in the required number of units, while in **M2**, it is determined separately for individual circulations, which induces threshold-based changes. In this case, such a threshold is observed between $\tau^T = 17$ and 19.

In terms of runtime, both strategic models are solved to optimality within 0.1 s, whereas the benchmark model **M3** requires substantially more time. This runtime difference demonstrates the computational advantage of the strategic models at the line level. The advantage is expected to become even more pronounced when scaling up to network-level settings or when timetable construction is also taken into account.

Overall, these results support the use of strategic models as efficient screening tools for estimating the cost of candidate line plans before applying more computationally intensive timetabling and tactical rolling stock scheduling models. Among the two strategic models, **M2** is preferable when estimation accuracy matters, as it more closely reproduces the benchmark costs by explicitly capturing unit circulations, provided that the approximated connection time is chosen carefully.

5.3. Comparison of strategic models *M1* and *M2*

The benchmark comparison in Section 5.2 shows, on a single line, that **M2** provides a closer cost estimate than **M1**. This section extends the comparison of the two strategic models to the network level to assess the deviation induced by the flow-based representation when complete unit trajectories are not tracked. At the same time, the maximum number of lines allowed in one circulation θ is varied from 1 to 3 to examine whether enriching the circulation set in **M2** leads to a noticeable change in the objective value. The circulation set remains restricted to elementary simple cycles. Table 6 reports the objective value, runtime, and MIP gap of **M2** for each value of θ. Since θ only affects the feasible circulation set in **M2**, the corresponding values of **M1** are reported separately as a fixed reference.

The results show that, at the network level, the objective difference between **M1** and **M2** becomes more pronounced. Under the baseline setting $\theta = 2$, the objective value of the flow-based model **M1** differs from that of **M2** by 3.65% and 4.41% for the partial and expanded network, respectively. In addition, for **M2**, enlarging the circulation set from $\theta = 1$ to 3 only slightly changes the objective value, with a total reduction of 0.61% and 0.04% for the partial and expanded network, respectively. By contrast, the computational burden rises substantially as θ increases, as reflected by the increase in runtime and MIP gap. This indicates that a restricted circulation set can already provide a relatively accurate cost estimate without incurring excessive computational effort.

Table 6
Comparison of objective values and computational performance under $\theta = 1, 2, 3$.

Model	θ	Partial network			Expanded network		
		Runtime (s)	MIP gap	Objective	Runtime (s)	MIP gap	Objective
M1	–	[a]3600	1.10%	880.2	70	0.10%	1771.9
M2	1	[a]3600	1.69%	914.8	19	0.03%	1854.0
M2	2	[a]3600	2.11%	913.5	172	0.10%	1853.5
M2	3	[a]3600	1.37%	909.3	[a]3600	0.12%	1853.2

Notes:
[a] 1-hour limit reached.

Overall, the results indicate that not tracking complete unit trajectories leads to a more pronounced deviation in **M1** at the network level, while further enriching the circulation set beyond $\theta = 2$ yields only limited additional benefit in **M2** relative to its computational cost. Given that **M2** provides a more detailed strategic representation, the results for **M2** are retained in the main paper when both models exhibit similar patterns, whereas those for **M1** are provided in Section S2 of the supplementary material.

5.4. Advantage of using stochastic programming approaches

This section examines the value of the stochastic approach relative to the deterministic approach. Section 5.4.1 first justifies the adopted scenario-set size. Section 5.4.2 compares the two approaches under different planning preferences over service quality and costs.

5.4.1. Calibration of scenario-set size

This section first examines the effect of scenario-set size on the quality of the stochastic solution under the Sample Average Approximation (SAA) framework. The analysis consists of two steps: (i) generating multiple first-stage solutions using independently sampled scenario sets of different sizes, and (ii) evaluating these solutions on a common reference scenario set.

Step 1: Generating first-stage solutions using sampled scenario sets. Denote the original full scenario set as Ψ, based on which sampled scenario sets with varying sizes ranging from 10 to 120 with a step size of 10 are generated. Denote a sampled scenario set as $\bar{\Psi}$, and $|\bar{\Psi}| \in \{10, 20, \ldots, 120\}$. For each $|\bar{\Psi}|$, six independent scenario samples are generated. For each sample, the two-stage stochastic program is solved to obtain a first-stage decision $\boldsymbol{n}(|\bar{\Psi}|)$.

Step 2: Evaluating the quality of each stochastic solution. To assess the quality of each first-stage decision, we evaluate it on the full scenario set Ψ as the reference set. For each scenario $\psi \in \Psi$, we solve a deterministic second-stage recourse problem with $\boldsymbol{n}(|\bar{\Psi}|)$ fixed, and obtain the corresponding second-stage objective value, denoted by $Z_2(\boldsymbol{n}(|\bar{\Psi}|), \psi)$. The performance measure of solution $\boldsymbol{n}(|\bar{\Psi}|)$ is defined as the mean evaluated second-stage objective over the full reference scenario set Ψ:

$$\mu_{|\bar{\Psi}|} = \frac{1}{|\Psi|} \sum_{\psi \in \Psi} Z_2(\boldsymbol{n}(|\bar{\Psi}|), \psi).$$

For each sample size $|\bar{\Psi}|$, independently sampled scenario sets are generated, yielding multiple values of $\mu_{|\bar{\Psi}|}$. The boxplots in Fig. 6 show the distribution of these values.

The results shows that for small scenario sets, e.g., $|\bar{\Psi}| = 10$ or $|\bar{\Psi}| = 20$, the performance of the resulting stochastic solutions exhibits substantial dispersion across the samples. This indicates high sensitivity to random scenario sampling and suggests that such small scenario sets provide insufficient statistical stability. As $|\bar{\Psi}|$ increases, both the median performance and the interquartile range improve substantially, indicating diminishing variability across scenario samples. Beyond $|\bar{\Psi}| = 50$, both the mean and the variability exhibit only marginal improvement. A comparison between the partial and the expanded network further shows that the partial network reaches this stability at smaller values of $|\bar{\Psi}|$, indicating that the number of scenarios required for statistical stability increases with the size of the network.

Taken together, the results suggest that a scenario-set size of approximately $|\bar{\Psi}| = 50$ provides a sufficiently accurate approximation. Furthermore, computation times with 50 scenarios are also reasonable, with only a 5.6% increase relative to a sample size of 10 for the expanded network. This scenario-set size is therefore adopted in the other numerical experiments.

5.4.2. Comparison with deterministic approach

With the scenario-set size fixed as above, this section compares the stochastic and deterministic approaches under different preferences of service quality relative to procurement and operating costs. Different shortage-penalty values w^{Q} are considered, spanning settings from a service-last to a service-first policy.

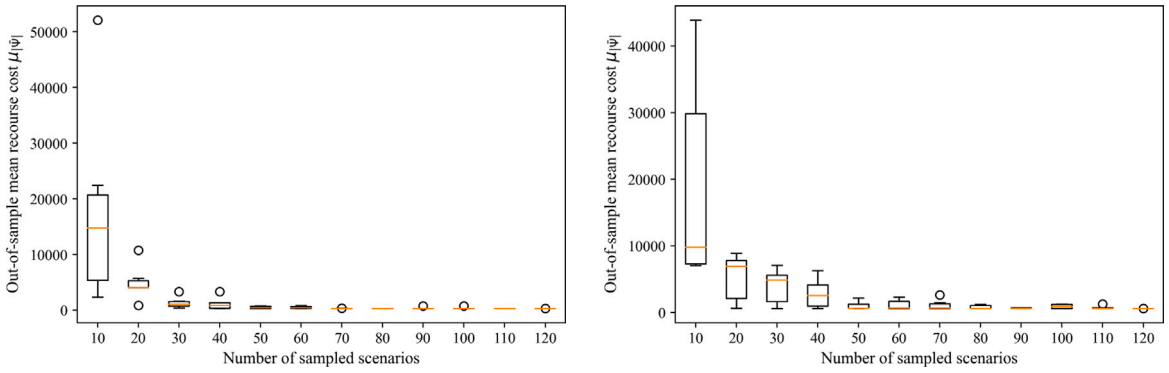
For each value of w^{Q}, a deterministic benchmark is constructed by replacing the uncertain demand on each arc in models **M1** and **M2** with its expected value. This yields a single-scenario deterministic model, denoted by $\bar{\Psi}$, whose solution determines the corresponding first-stage decision $\boldsymbol{n}(\bar{\Psi})$. For the stochastic approach, the first-stage decision is obtained by solving the stochastic model on a sampled scenario set of size 50 and is subsequently evaluated on the same full scenario set. To preserve the intended trade-off between service quality and cost, the full composition set P_{ij} is retained in this experiment. The resulting first-stage purchase cost, second-stage shunting and operational cost, seat shortages, and procured units for model **M2** are reported in Table 7.

The results show that, when w^{Q} is very small, shortages are penalized too weakly and neither model procures rolling stock. Once w^{Q} becomes large enough to make seat shortages relevant, the results show a procurement–service trade-off. The deterministic approach tends to adopt a tighter fleet, whereas the stochastic approach procures a more conservative fleet plan and therefore preserves service quality more effectively.

This trade-off is reflected in most tested settings by positive values of SR. In addition, the reduction in shortages is not monotone in w^{Q}. It is largest at intermediate penalty levels, where the deterministic solution still tends to limit fleet size, while the stochastic solution more actively accounts for the risk associated with high-demand scenarios. As w^{Q} becomes very large, both approaches are pushed toward near-zero shortages, so the shortage gap narrows. Overall, the stochastic approach changes the procurement decision in a way that offers stronger protection against scenario variability when shortages start to matter.

The results also show that the effect of stochastic planning depends on the network setting. In the partial network, the main effect of the stochastic approach is to procure more rolling stock in exchange for fewer shortages. When the network provides limited room to redistribute capacity, the value of the stochastic approach lies primarily in constructing a more protective fleet plan. This explains why the improvement in shortages in the partial network is often accompanied by negative values of SU and, in many cases, higher second-stage cost as well. From a managerial perspective, the stochastic approach then acts mainly as an insurance mechanism: the operator pays more upfront in order to reduce expected service loss under uncertain demand.

In the expanded network, however, the same insurance effect interacts with a richer operating structure. Here, a better anticipation of



(a) Partial network. (b) Expanded network.

Fig. 6. Distribution of out-of-sample mean recourse cost for model **M2**.

Table 7
Comparison between stochastic and deterministic approaches for Model **M2**.

Network	$\bar{w}^{Q}$	Stochastic approach				Deterministic approach				SR	SU
		Z^{P}	Z^{SO}	SS	PU	Z^{P}	Z^{SO}	SS	PU		
Partial	10^{-4}	0	0	27 979	0	0	0	27 979	0	0	0
Partial	10^{-3}	275	218	2 968	156	262	210	3 521	139	553	−17
Partial	10^{-2}	544	303	129	146	436	284	687	115	558	−31
Partial	10^{-1}	578	319	1	129	462	303	495	115	494	−14
Partial	10^{0}	588	319	0	127	506	315	155	104	155	−23
Partial	10^{1}	588	319	0	127	506	315	155	104	155	−23
Partial	10^{2}	580	323	0	131	506	315	155	104	155	−27
Partial	10^{3}	598	317	0	123	506	315	155	104	155	−19
Expanded	10^{-4}	0	0	62 807	0	0	0	62 807	0	0	0
Expanded	10^{-3}	589	441	6 686	324	583	441	6 854	316	168	−8
Expanded	10^{-2}	1127	605	119	279	984	611	625	285	506	6
Expanded	10^{-1}	1200	622	2	266	1064	642	228	264	226	−2
Expanded	10^{0}	1228	622	0	257	1064	647	227	264	227	7
Expanded	10^{1}	1228	623	0	263	1064	647	228	264	228	1
Expanded	10^{2}	1228	622	0	258	1064	647	228	264	228	6
Expanded	10^{3}	1228	622	0	256	1064	647	228	264	228	8

Notes: Z^{P} = First-stage purchase cost. Z^{SO} = Second-stage shunting and operational cost. SS = Seat shortages. PU = Purchased units. SR = Shortage reduction, calculated as SS (deterministic) − SS (stochastic). SU = Saved units, calculated as PU (deterministic) − PU (stochastic).

demand uncertainty not only influences how many units are procured, but also improves how these units are positioned and utilized across the network. This is reflected in the second-stage cost: for moderate-to-high values of w^{Q}, the stochastic approach often achieves lower Z^{SO} than the deterministic approach in the expanded network. Hence, the additional first-stage investment is not merely buying reserve capacity; it also supports a deployment pattern that fits the scenario-dependent demand more effectively. In several expanded-network settings, this improvement is strong enough that the stochastic solution reduces shortages while using no more units, and in some cases even fewer units, than the deterministic one. At high penalty levels, this reduction in the total number of units is accompanied by a higher purchasing cost, indicating that the stochastic approach also changes the fleet composition toward more effective but more expensive unit types.

Taken together, the results indicate that the main benefit of the stochastic approach is that it exposes and manages the procurement–service trade-off under demand uncertainty more explicitly. When service reliability is important, the stochastic approach offers a fleet plan that is less vulnerable to adverse demand realizations; and when the network is sufficiently rich, part of this benefit can be recovered through more efficient second-stage deployment, and in some settings even through a reduction in required units.

5.5. Benefit of train unit type diversity

Introducing additional train unit types increases the variety of feasible compositions and composition changes, thereby offering greater flexibility in responding to heterogeneous demand. At the same time, a larger number of unit types increases system complexity and computational effort. From a strategic perspective, operators must make informed trade-offs between the number of unit types, achievable efficiency gains, and the overall complexity of the system. This section investigates whether the increased diversity in unit types is justified by comparing results obtained under different numbers of train unit types.

Specifically, three cases are tested with varying $|\mathcal{T}|$: (i) $|\mathcal{T}| = 1$, using only V4 units, as using V3, V6, or I4 units alone cannot satisfy demand under the 12-carriage train-length limit in most scenarios; (ii) $|\mathcal{T}| = 2$, using V4 and V6 units; (iii) $|\mathcal{T}| = 4$, using V4, V6, I3, and I4 units (the baseline). When the available train unit types are reduced from four to a single type, the numbers of feasible compositions and composition changes on the expanded network decrease by 81% and 88%, respectively, indicating a substantial reduction in problem complexity. The problem complexity, objective values, and cost components for model **M2** are reported in Table 8.

The findings show that incorporating multiple train unit types reduces total cost. In particular, increasing $|\mathcal{T}|$ from 1 to 2, results in a 1% and 3% reduction in total cost on the partial and expanded network, respectively, despite the substantial increase in the number of composition changes. This improvement is largely driven by the availability of the V6 units, which provide additional capacity flexibility at a relatively favorable cost.

However, extending the unit set from $|\mathcal{T}| = 2$ to $|\mathcal{T}| = 4$ by adding the single-deck I3 and I4 units results in almost no further reduction

Table 8
Objectives and cost components with various numbers of unit types.

$\|\mathcal{T}\|$	$\|P\|$	Partial network				Expanded network			
		$\|Q\|$	Objective	Z^P	Z^{SO}	$\|Q\|$	Objective	Z^P	Z^{SO}
1	4	294	914.3	580	334.3	1 536	1917.7	1256	661.7
2	7	638	912.7	594	318.7	3 186	1853.9	1228	625.9
4	21	2298	913.5	594	319.5	12 742	1853.5	1228	625.5

Notes: Z^P = first-stage purchase cost. Z^{SO} = second-stage cost.

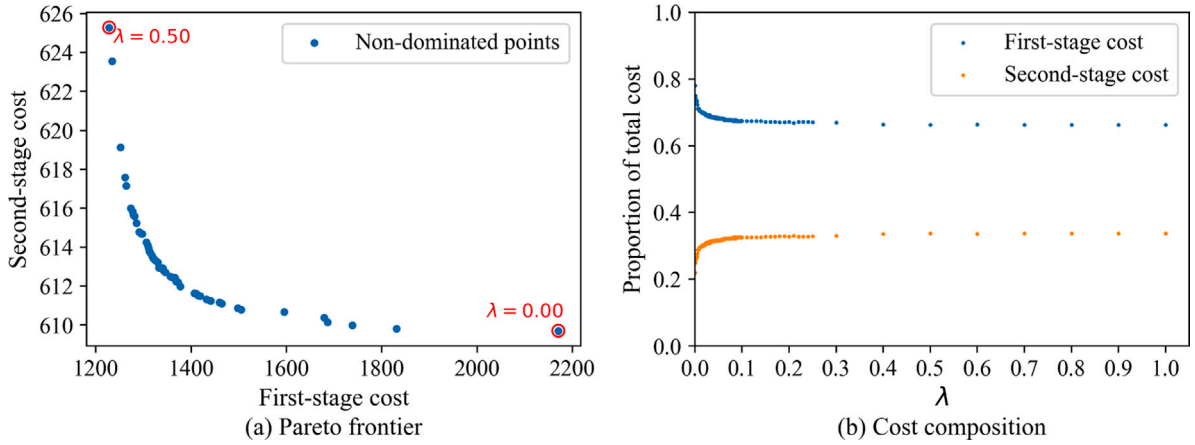


Fig. 7. Evolution of first-stage and second-stage costs.

in total cost. This limited gain can be traced back to the cost structure used in the experiments: both purchase and operational costs are nearly proportional to the seat capacity, and the purchase cost accounts for more than half of the total cost. Under such a cost specification, the ICM units (I3 and I4) do not offer a clear economic advantage relative to V4 and V6, which restricts their usefulness in the optimization.

Overall, the performance difference between the two- and four-type settings provides a clear managerial insight: under a cost structure that is approximately linear in seat capacity, the V4 and V6 units are already sufficiently cost-effective and versatile to cover the required services. Further research is needed to examine cases in which the availability of a specific unit type, such as the V4, is limited.

5.6. Trade-off between first-stage and second-stage costs

In the objectives of the proposed models, the parameter λ determines the relative emphasis placed on first-stage purchase cost versus second-stage operational cost. A higher value of λ corresponds to a planning preference in which rolling stock procurement is weighted more strongly, whereas a lower value places relatively more emphasis on second-stage operational performance across scenarios. As strategic planning often involves comparing alternatives under different planning preferences rather than fixing a single universally accounting trade-off, we use λ as a transparent decision lever to examine how fleet investment and operational outcomes change as procurement is emphasized to different degrees.

We consider the expanded network and employ model **M2** to gradually increase λ from 0 to 1.0, covering the transition from operations-emphasized to procurement-emphasized preferences. Forty-nine non-dominated solutions are identified, and the resulting Pareto frontier is shown in Fig. 7a. The corresponding proportions of first-stage and second-stage costs are shown in Fig. 7b.

The Pareto frontier shows that when λ is very small, modest increases in procurement emphasis substantially reduce purchase cost while causing noticeable changes in second-stage operating cost. Once λ exceeds approximately 0.1, the frontier becomes much flatter, and the cost composition becomes more stable. It indicates that procurement already receives sufficient emphasis to prevent excessive fleet choices and that further increases mainly affect second-stage fine-tuning within a procurement-dominant regime. This is consistent with practical settings in which investment is typically the dominant cost component in strategic rolling stock planning. Since our baseline setting uses $\lambda = 1$, the baseline conclusions correspond to a procurement-emphasized planning preference.

A similar stabilization appears in the purchased fleet composition, as shown in Fig. 8. When λ is small, more train units of all types are purchased, whereas for $\lambda \geq 0.1$ both the number and the types of purchased units become stable. Within $0 \leq \lambda \leq 0.1$, purchases of V4 units increase slightly while purchases of the other types decrease, indicating that V4 units are relatively cost-effective and can replace other types economically. This is related to their operational flexibility: under the 12-carriage train-length limit, V4 and V6 units, especially V4 alone, allow a wider range of feasible compositions with different capacities, making them better suited to accommodate demand variation across scenarios. For example, V4 units alone can form three feasible compositions consisting of one, two, or three units, with the three-unit composition providing the largest capacity among all feasible compositions generated by the four unit types. In addition, units of types I3 and I4 disappear for $\lambda \geq 0.4$, suggesting that once procurement is sufficiently emphasized, these types become less attractive than more flexible and cost-effective alternatives.

Overall, the results indicate that in the practically relevant procurement-emphasized regime, the optimal fleet composition is driven by cost-effective and operationally flexible unit types such as V4, underscoring the importance of addressing the problem at the strategic planning level.

5.7. Benefit for allowing shunting movements

Building on the analysis in Section 5.6, this section examines whether allowing shunting movements affects the resulting strategic procurement decisions. The proposed model is compared with a restricted no-shunting variant in which neither in-line nor cross-line shunting movements are permitted. The formulation of this variant is provided in Appendix B. Both models are solved using the same stochastic demand scenarios on the expanded network, and the resulting final

Table 9
Comparison of shunting-enabled baseline and no-shunting variant.

Network	$\bar{w}^{Q}$	Shunting-enabled				No-shunting				SR	SU
		Z^{P}	Z^{SO}	SS	PU	Z^{P}	Z^{SO}	SS	PU		
Partial	10^{-4}	0	0	27 802	0	0	0	27 802	0	0	0
Partial	10^{-3}	275	217	3 004	156	280	205	3 192	165	188	9
Partial	10^{-2}	544	304	134	146	570	288	255	134	121	−12
Partial	10^{-1}	578	321	1	129	660	313	0	139	−1	10
Partial	10^{0}	588	321	0	127	660	314	0	139	0	12
Partial	10^{1}	588	321	0	127	660	314	0	139	0	12
Partial	10^{2}	580	324	0	131	660	314	0	139	0	8
Partial	10^{3}	598	317	0	123	660	314	0	139	0	16
Expanded	10^{-4}	0	0	62 932	0	0	0	62 932	0	0	0
Expanded	10^{-3}	589	441	6 864	324	586	416	7 731	329	867	5
Expanded	10^{-2}	1127	610	122	279	1197	591	213	281	91	2
Expanded	10^{-1}	1200	626	5	266	1310	615	8	278	3	12
Expanded	10^{0}	1228	625	0	257	1350	619	0	279	0	22
Expanded	10^{1}	1228	625	0	263	1356	619	0	279	0	16
Expanded	10^{2}	1228	625	0	258	1356	619	0	281	0	23
Expanded	10^{3}	1228	626	0	256	1356	620	0	280	0	24

Notes: Z^{P} = First-stage purchase cost. Z^{SO} = Second-stage shunting and operational cost. SS = Seat shortages. PU = Purchased units. SR = Shortage reduction, calculated as SS (no-shunting) − SS (shunting-enabled). SU = Saved units, calculated as PU (no-shunting) − PU (shunting-enabled).

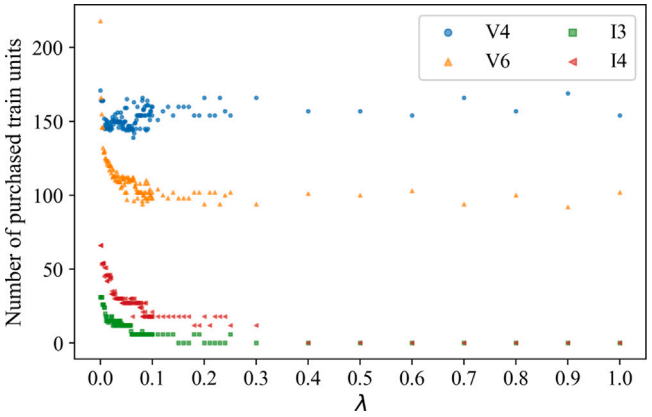


Fig. 8. Distribution of the number of purchased train units.

solutions are compared across different seat-shortage penalty values w^{Q}. The results for model **M2** is reported in Table 9.

When w^{Q} is very small, neither approach purchases additional train units, resulting in substantial seat shortages. Once w^{Q} exceeds this threshold and imposes a meaningful penalty on seat shortages, both models invest in additional rolling stock. Since different combinations of train unit types may be selected, the corresponding changes in seat-shortage reduction and saved units are not necessarily monotonic. The advantage of allowing shunting is twofold:

(1) In terms of investment cost and procured resources, the shunting-enabled model requires on average 14 fewer purchased units, corresponding to a reduction of 7.2% in investment cost. As a result, it also yields a lower total cost even after accounting for shunting-related operating costs. When procurement costs are monetized in a more realistic manner and therefore constitute the dominant component of the objective, the cost advantage associated with shunting-driven procurement savings is expected to be even more pronounced. This finding confirms the advantage of redistributing capacity along the network, as illustrated in Fig. 4, reducing the rolling stock required to achieve a given service level.

(2) In terms of service quality, the shunting-enabled model achieves comparable service levels at moderate values of w^{Q} and substantially lower seat shortages at larger values of w^{Q}. On average, allowing shunting reduces seat shortages by approximately 100 passengers across the tested settings. This effect is especially evident in the expanded network, where operational heterogeneity across lines and directions is greater and the flexibility created by shunting can be exploited more effectively.

Overall, the results demonstrate that allowing shunting movements improves both resource efficiency and service performance, enabling comparable or better service levels with markedly fewer procured units.

6. Conclusions

In this study, we focused on a strategic rolling stock scheduling problem with demand uncertainty, aiming to estimate the cost of implementing rolling stock to execute a line plan. We proposed and compared two stochastic programming models with a two-stage structure and varying levels of complexity. In the first stage, operators decide the number of train units, and in the second stage the rolling stock circulations, based on passenger demand.

The two strategic models are compared against a trip-based benchmark, which shows that the circulation-based model provides the more accurate cost estimate. Within that model, restricting attention to short and simple cycles captures most of the benefit while avoiding the substantial computational burden of a richer circulation set.

Compared with a deterministic approach, stochastic planning makes the procurement–service trade-off under demand uncertainty more explicit. When service reliability matters, it yields fleet plans that are less vulnerable to demand fluctuations and, in richer networks, may even reduce unit requirements through more efficient deployment.

Furthermore, an integrated use of multiple train unit types yields a 1% and 3% reduction in total cost for the partial and expanded networks, respectively, highlighting the value of unit types that are both economical and flexible in forming compositions tailored to scenario-specific demand.

The analysis of planning preferences shows that, in the practically relevant procurement-emphasized regime, investment dominates the cost structure and the preferred fleet is driven by cost-effective and operationally flexible unit types, with V4 standing out as a particularly attractive option.

Finally, allowing shunting movements gives planners additional flexibility to redistribute capacity across the network, thereby reducing the fleet required to achieve a given service level, with the benefit being particularly visible in richer network settings.

Future research could extend this framework to explicitly incorporate crew scheduling, allowing the joint determination of rolling stock purchases and crew requirements. Such an extension would quantify the number of drivers needed to operate the planned rolling stock

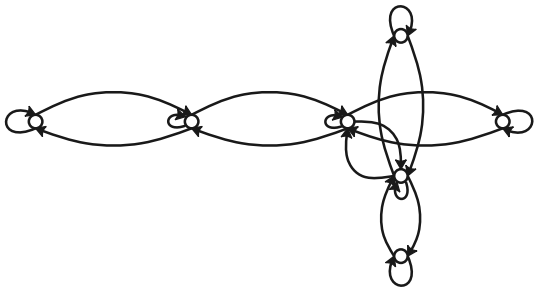


Fig. A.1. Contracted graph corresponding to the service network in Fig. 2b.

schedule, account for constraints such as legal rest periods, and provide a more complete estimate of total system cost. Another possible direction is to consider additional uncertainties of track blockages caused by infrastructure failures and weather conditions to cater to the practical needs under disruptions.

CRediT authorship contribution statement

Siqiao Li: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Rolf van Lieshout:** Conceptualization, Methodology, Validation, Writing – review & editing. **Patrick Stokkink:** Methodology, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Proofs

Proposition 1. If an instance is symmetric, the flow-based model **M1** admits an optimal solution that is symmetric.

Proof of Proposition 1. A flow solution is given by the collection $\mathbf{x} = \left(x_{ij}^{\varphi,t}\right)_{(i,j)\in A,\, t\in T,\, \varphi\in\Psi}$, where $x_{ij}^{\varphi,t} := \sum_{p\in P_{ij}} n_p^t\, y_{p,ij}^{\varphi}$ denotes the number of t-type train units on arc (i,j) in scenario φ. According to Definition 2, a solution $\mathbf{x}$ is *symmetric* if $x_{ij}^{\varphi,t} = x_{\bar{i},\bar{j}}^{\varphi,t}$ for all train unit types t, all scenarios φ, and all arcs (i,j) that have a conjugate.

We begin by considering the rolling stock schedule for a fixed train type t in any scenario φ, and we suppress the superscripts for brevity. Consider any optimal solution $\mathbf{x}^* = (x_{ij}^*)$ and $x_{ij}^* := \sum_{p\in P_{ij}} n_p^t\, y_{p,ij}^{\varphi}$. Construct the auxiliary graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ by merging every $i \in N$ with its conjugate $\bar{i}$. Fig. A.1 provides an example of this construction. By construction, $\mathbf{x}^*$ also defines a feasible flow on $\mathcal{G}$.

Next, compute a *residual* solution $\mathbf{x}'$ by setting

$$x'_{ij} := x_{ij}^* - \min\{x_{ij}^*, x_{\bar{i},\bar{j}}^*\} \qquad \forall (i,j) \in \mathcal{A}.$$

The same value is subtracted from each arc and its conjugate, so $\mathbf{x}'$ still satisfies the flow conservation constraints on $\mathcal{G}$. If the original solution $\mathbf{x}^*$ is asymmetric, then the residual solution $\mathbf{x}'$ must be non-zero. Since $\mathbf{x}'$ is a flow, this implies that there exists a circulation in $\mathcal{G}$ with non-zero flow.

Consider such a simple circulation c in $\mathcal{G}$ and its conjugate circulation $\bar{c}$, and the corresponding sets of arcs in the original network. By symmetry of the instance (Definition 2), every arc on c and its conjugate arc on $\bar{c}$ share identical parameters, including demand, penalty coefficients for unserved demand, and rolling stock cost coefficients.

Starting from the optimal solution $\mathbf{x}^*$, we may adjust the flow along c and $\bar{c}$ so as to reduce their imbalance, i.e., by decreasing the flow on c and increasing the flow on $\bar{c}$ by the same amount, or vice versa, while keeping all other flows unchanged. For each such adjustment, the composition and slack variables (y, ϵ) on the affected arcs are re-optimized to satisfy the demand constraints and minimize the resulting objective value. Due to the symmetry of costs and penalties, the marginal cost trade-off associated with modifying the flow on c is identical to that on $\bar{c}$. Hence, aligning the flow on c with that on $\bar{c}$ cannot increase the objective value.

Since $\mathbf{x}^*$ is optimal, the adjusted solution must also be optimal. Moreover, this operation strictly reduces the difference between the flows on at least one pair of conjugate arcs. Repeating this procedure for all simple circulations in the residual flow (and for all train types t and scenarios φ) yields, after finitely many steps, an optimal solution in which no residual flow remains. Consequently, we obtain an optimal solution satisfying $x_{ij}^{\varphi,t} = x_{\bar{i},\bar{j}}^{\varphi,t}$ for all train types t, all scenarios φ, and all arcs that have a conjugate, i.e., an optimal symmetric solution. □

Proposition 2. For a symmetric instance, the flow-based model **M1** admits the optimal solution that does not use any shunting arcs between different lines.

Proof of Proposition 2. It follows from Proposition 1 that only symmetric solutions need to be considered. Now suppose that a solution contains positive flow on a cross-line shunting arc from line A to line B. By symmetry, the solution contains the same flow on the conjugate arc, which is the cross-line shunting arc from line B to line A. Therefore, this flow can be rerouted by using the two same-line shunting arcs, from line A to line A, and from line B to line B, which does not increase the costs. Therefore, there exists an optimal solution that does not use cross-line shunting arcs. □

It should be noted that the rerouting step in the above proof is interpreted within the periodic service network representation; under a fixed timetable, additional event-time constraints may restrict such transformations.

Proposition 3. Let $Z_{\mathbf{M}i}$ denote the optimal objective value of model **M**i, where $i \in \{1, 2\}$. For any instance, it holds that $Z_{\mathbf{M1}} \le Z_{\mathbf{M2}}$.

Proof of Proposition 3. Consider any feasible solution $(\bar{\mathbf{y}}, \bar{\mathbf{v}}, \bar{\mathbf{z}}, \bar{\mathbf{n}}, \bar{\epsilon})$ to **M2**. We will show that the solution $(\bar{\mathbf{y}}, \bar{\mathbf{v}}, \bar{\mathbf{n}}, \bar{\epsilon})$ is feasible for **M1** and that its objective value is no larger than that of the original **M2** solution.

First, the variables $(\bar{\mathbf{y}}, \bar{\mathbf{v}}, \bar{\epsilon})$ satisfy Eqs. (1b)–(1f), (1h), (1i), and (1k), which are shared by both models. Since $\bar{\mathbf{z}}$ corresponds to circulations, Eq. (2g) guarantees that the flow induced on each arc by $\bar{\mathbf{z}}$ is consistent with that defined by $\bar{\mathbf{y}}$.

We next examine the inventory constraint. By definition, k_c is no smaller than the total running time of circulation c measured in units of H, i.e., $k_c \ge \frac{\sum_{(i,j)\in A_c} \tau_{ij}}{H}$. Therefore, for any scenario φ,

$$\bar{n}_t \ge \sum_{c\in C} k_c \bar{z}_{t,c}^{\varphi} \ge \sum_{c\in C} \sum_{(i,j)\in A_c} \frac{\tau_{ij}}{H} \bar{z}_{t,c}^{\varphi} = \sum_{(i,j)\in A} \sum_{p\in P_{ij}} n_p^t \frac{\tau_{ij}}{H} \bar{y}_{p,ij}^{\varphi}, \tag{A.1}$$

which directly leads to Eq. (1g) of **M1**. Hence, $(\bar{\mathbf{y}}, \bar{\mathbf{v}}, \bar{\mathbf{n}}, \bar{\epsilon})$ is feasible for **M1**.

Finally, note that **M1** and **M2** share the same objective function, and in both models the inventory cost is nondecreasing in n_t. Eq. (1g) only requires n_t in **M1** to be at least the flow-induced quantity on the left-hand side, whereas Eqs. (2h) and (A.1) show that the values $\bar{n}_t$ used in **M2** are no smaller than this quantity. Hence, an optimal solution to **M1** satisfies $n_t \le \bar{n}_t$ for all t, implying that the objective value of **M1** is no larger than that of the original **M2** solution. □

$$\min \quad \lambda \sum_{t\in\mathcal{T}} w^{\mathrm{P}}_t n_t + \sum_{\varphi\in\Psi} \frac{1}{|\Psi|}\left(\sum_{r\in R}\sum_{i\in N^{\mathrm{sh}}_r}\sum_{q\in Q_{ir}} w^{\mathrm{S}} s_q v^{\varphi}_{q,ir} + \sum_{(r,i,j)\in A}\sum_{p\in P_{ij}}\sum_{t\in\mathcal{T}} w^{\mathrm{O}}_{t,ij} n^t_p y^{\varphi}_{p,r,ij} + \sum_{(r,i,j)\in A} w^{\mathrm{Q}}_{ij}\epsilon^{\varphi}_{r,ij}\right), \tag{C.1a}$$

$$\text{s.t.} \quad \sum_{p\in P_{ij}} e_p y^{\varphi}_{p,r,ij} + \epsilon^{\varphi}_{r,ij} \geq d^{\varphi}_{r,ij} \qquad \forall (r,i,j)\in A,\ \forall\varphi\in\Psi, \tag{C.1b}$$

$$\sum_{p\in P_{ij}} y^{\varphi}_{p,r,ij} = 1 \qquad \forall (r,i,j)\in A,\ \forall\varphi\in\Psi, \tag{C.1c}$$

$$\sum_{q\in Q_{ir}} v^{\varphi}_{q,ir} = 1 \qquad \forall r\in R,\ \forall i\in N^{\mathrm{sh}}_r,\ \forall\varphi\in\Psi, \tag{C.1d}$$

$$y^{\varphi}_{p,r,ij} = \sum_{q\in Q_{ir}} \mu^{q+}_{p,r,ij} v^{\varphi}_{q,ir} \qquad \forall (r,i,j)\in A,\ \forall p\in P_{ij},\ \forall\varphi\in\Psi, \tag{C.1e}$$

$$y^{\varphi}_{p,r,ij} = \sum_{q\in Q_{jr}} \mu^{q-}_{p,r,ij} v^{\varphi}_{q,jr} \qquad \forall (r,i,j)\in A,\ \forall p\in P_{ij},\ \forall\varphi\in\Psi, \tag{C.1f}$$

$$\sum_{(r,i,j)\in A^{\mathrm{src}}}\sum_{p\in P_{ij}} n^t_p y^{\varphi}_{p,r,ij} \leq n_t \qquad \forall t\in\mathcal{T},\ \forall\varphi\in\Psi, \tag{C.1g}$$

$$y^{\varphi}_{p,r,ij} \in \{0,1\} \qquad \forall (r,i,j)\in A,\ \forall p\in P_{ij},\ \forall\varphi\in\Psi, \tag{C.1h}$$

$$v^{\varphi}_{q,ir} \in \{0,1\} \qquad \forall r\in R,\ \forall i\in N^{\mathrm{sh}}_r,\ \forall q\in Q_{ir},\ \forall\varphi\in\Psi, \tag{C.1i}$$

$$n_t \in \mathbb{Z} \qquad \forall t\in\mathcal{T}, \tag{C.1j}$$

$$\epsilon^{\varphi}_{r,ij} \in \mathbb{Z} \qquad \forall (r,i,j)\in A,\ \forall\varphi\in\Psi. \tag{C.1k}$$

Box III.

Appendix B. Model adaptation to a no-shunting variant

We consider a no-shunting variant in which train compositions cannot be changed en route. Consequently, all shunting arcs are removed from the network. For the circulation-based model **M2**, the circulation set C only contains a complete cycle for each line.

For both models, the second-stage shunting cost term is removed from the objectives (1a) and (2a). The composition-change decision variables $v^{\varphi}_{q,i}$ are removed, and the associated composition-change constraints (1d)–(1f) and (2d)–(2f) are deleted. Moreover, since each arc belongs to exactly one line, the arc-level "one-composition-per-arc" constraint, (1c) and (2c), becomes redundant and is removed. To enforce that each line uses a single composition, we introduce a line-level selection variable $x^{\varphi}_{p,\ell} \in \{0,1\}$ and add:

$$\sum_{p\in P_\ell} x^{\varphi}_{p,\ell} = 1 \qquad \forall \ell\in L,\ \forall\varphi\in\Psi, \tag{B.1}$$

$$y^{\varphi}_{p,ij} = x^{\varphi}_{p,\ell} \qquad \forall \ell\in L,\ \forall (i,j)\in A_\ell,\ \forall p\in P_\ell,\ \forall\varphi\in\Psi. \tag{B.2}$$

Here, P_ℓ denotes the set of feasible train compositions for line ℓ, and A_ℓ denotes the set of arcs associated with line ℓ.

Appendix C. A higher-resolution benchmark formulation

This section constructs a benchmark formulation **M3** that serves as a higher-resolution counterpart for one representative line, considering the constraints enforced by a tactical model proposed by Fioole et al. [13]. The benchmark is not intended as a fully tactical rolling stock circulation model. Instead, it serves as an intermediate extension between the strategic model and a fully tactical formulation, and is used to assess how the proposed strategic approximation behaves when the time dimension is represented explicitly at the trip level.

To avoid introducing tactical complexity beyond the scope of the strategic model, all available units are assumed to be parked at the two terminals, where inventory constraints are imposed. At intermediate stations, unit reuse is restricted through designated successor-trip connections, such that an uncoupled unit can only be coupled to the first outgoing trip in the opposite direction.

Let a trip denote one complete journey between the two terminal stations. Model **M3** takes a timetable-defined trip set R as additional input. Compared with models **M1** and **M2**, model **M3** assigns compositions to trip arcs (r,i,j) with $r\in R$, and models shunting movements when trip r stops at shunting stations $i\in N^{\mathrm{sh}}_r$. The formulation is given below (see the Eqs. (C.1a)–(C.1k) in Box III)

Constraints (C.1b)–(C.1f) extend their counterparts in the strategic model by introducing the trip index r. Constraint (C.1g) introduces A^{src} as the set of source trip arcs whose entering compositions are not induced by any predecessor transition, so that fleet size is determined by the initial terminal inventories required to activate these arcs.

With the formulation in place, the remaining issue is how to convert the benchmark outcome into a periodic measure that is comparable to the strategic results. To this end, we select a set of representative services from the timetable and use the associated shunting, operational, and shortage costs to construct a periodic cost induced by the higher-resolution benchmark. Specifically, let $\bar{H} = BH$ be the smallest integer-multiple-period window such that one complete service between the two terminals can be finished within this interval, i.e., $\bar{H} = BH > \sum_{(i,j)\in\bar{A}} \tau_{ij}$, where B is a positive integer and $\bar{A}$ denotes the set of running arcs in one direction, upstream or downstream. All services whose departure times from the terminals fall within this window are taken as representative services. The timetable horizon is chosen sufficiently long so that these representative services already exhibit a periodic pattern, that is, each of them has both predecessor and successor trips that are also completed within the modeled horizon. Accordingly, for benchmark model **M3**, the purchase cost is evaluated in the same way as in Eq. (C.1a), whereas the second-stage cost is evaluated only over the representative services, as given in Eq. (C.2).

$$\bar{C} = \sum_{\varphi\in\Psi} \frac{1}{B|\Psi|}\left(\sum_{r\in\bar{R}}\sum_{i\in N^{\mathrm{sh}}_r}\sum_{q\in Q_{ir}} w^{\mathrm{S}} s_q v^{\varphi}_{q,ir} + \sum_{(r,i,j)\in\bar{A}}\sum_{p\in P_{ij}}\sum_{t\in\mathcal{T}} w^{\mathrm{O}}_{t,ij} n^t_p y^{\varphi}_{p,r,ij} + \sum_{(r,i,j)\in\bar{A}} w^{\mathrm{Q}}_{ij}\epsilon^{\varphi}_{r,ij}\right). \tag{C.2}$$

For the Line 3500 case study, the first downstream service is designed to depart at time 0, while the first upstream service departs 15 min later. The one-way travel time, i.e., $\sum_{(i,j)\in\bar{A}} \tau_{ij}$, is 119 min. Under the 30-minute period setting, this implies a representative window of length 120 min. Hence, B is set to $\bar{H}/H = 4$.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.omega.2026.103593.

Data availability

Research Link Provided

Demand input data for computational experiments (Original data) (GitHub)

References

[1] Cipriani Ernesto, Gori Stefano, Petrelli Marco. Transit network design: A procedure and an application to a large urban area. Transp Res C 2012;20(1):3–14.

[2] Farahani Reza Zanjirani, Miandoabchi Elnaz, Szeto Wai Yuen, Rashidi Hannaneh. A review of urban transportation network design problems. European J Oper Res 2013;229(2):281–302.

[3] Manser Patrick, Becker Henrik, Hörl Sebastian, Axhausen Kay W. Designing a large-scale public transport network using agent-based microsimulation. Transp Res A: Policy Pr 2020;137:1–15.

[4] Lusby Richard M, Larsen Jesper, Ehrgott Matthias, Ryan David. Railway track allocation: Models and methods. OR Spectrum 2011;33:843–83.

[5] Schöbel Anita. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. Transp Res C 2017;74:348–65.

[6] Schiewe Philine, Schöbel Anita. Integrated optimization of sequential processes: General analysis and application to public transport. EURO J Transp Logist 2022;11:100073.

[7] Goossens Jan-Willem, Van Hoesel Stan, Kroon Leo. A branch-and-cut approach for solving railway line-planning problems. Transp Sci 2004;38(3):379–93.

[8] Goerigk Marc, Schmidt Marie. Line planning with user-optimal route choice. European J Oper Res 2017;259(2):424–36.

[9] Canca David, De-Los-Santos Alicia, Laporte Gilbert, Mesa Juan A. Integrated railway rapid transit network design and line planning problem with maximum profit. Transp Res E 2019;127:1–30.

[10] Abbink Erwin, Van den Berg Bianca, Kroon Leo, Salomon Marc. Allocation of railway rolling stock for passenger trains. Transp Sci 2004;38(1):33–41.

[11] Budai Gabriella, Maróti Gábor, Dekker Rommert, Huisman Dennis, Kroon Leo. Rescheduling in passenger railways: the rolling stock rebalancing problem. J Sched 2010;13:281–97.

[12] Alfieri Arianna, Groot Rutger, Kroon Leo, Schrijver Alexander. Efficient circulation of railway rolling stock. Transp Sci 2006;40(3):378–91.

[13] Fioole Pieter-Jan, Kroon Leo, Maróti Gábor, Schrijver Alexander. A rolling stock circulation model for combining and splitting of passenger trains. European J Oper Res 2006;174(2):1281–97.

[14] Nielsen Lars Kjær, Kroon Leo, Maróti Gábor. A rolling horizon approach for disruption management of railway rolling stock. European J Oper Res 2012;220(2):496–509.

[15] Lin Zhiyuan, Kwan Raymond SK. A branch-and-price approach for solving the train unit scheduling problem. Transp Res B 2016;94:97–120.

[16] Grimm Boris, Hoogervorst Rowan, Borndörfer Ralf. A comparison of two models for rolling stock scheduling. Transp Sci 2025.

[17] Haahr Jørgen, Wagenaar Joris C, Veelenturf Lucas P, Kroon Leo G. A comparison of two exact methods for passenger railway rolling stock (re)scheduling. Transp Res E 2016;91:15–32.

[18] Haahr Jørgen, Lusby Richard M. Integrating rolling stock scheduling with train unit shunting. European J Oper Res 2017;259(2):452–68.

[19] Zhong Qingwei, Lusby Richard M, Larsen Jesper, Zhang Yongxiang, Peng Qiyuan. Rolling stock scheduling with maintenance requirements at the Chinese High-Speed Railway. Transp Res B 2019;126:24–44.

[20] Hoogervorst Rowan, Dollevoet Twan, Maróti Gábor, Huisman Dennis. A variable neighborhood search heuristic for rolling stock rescheduling. EURO J Transp Logist 2021;10:100032.

[21] Jha Krishna C, Ahuja Ravindra K, Şahin Güvenç. New approaches for solving the block-to-train assignment problem. Networks: An Int J 2008;51(1):48–62.

[22] Piu Francesco, Speranza M Grazia. The locomotive assignment problem: A survey on optimization models. Int Trans Oper Res 2014;21(3):327–52.

[23] Cordeau Jean-François, Soumis François, Desrosiers Jacques. Simultaneous assignment of locomotives and cars to passenger trains. Oper Res 2001;49(4):531–48.

[24] Lingaya Norbert, Cordeau Jean-Françcois, Desaulniers Guy, Desrosiers Jacques, Soumis Françcois. Operational car assignment at VIA rail Canada. Transp Res B 2002;36(9):755–78.

[25] Van Lieshout Rolf, Bouman Paul. Vehicle scheduling based on a line plan. In: 18th workshop on algorithmic approaches for transportation modelling, optimization, and systems. ATMOS 2018, Vol. 65, 2018, p. 15:1–15:14.

[26] Bouman Paul, Schiewe Alexander, Schiewe Philine. A new sequential approach to periodic vehicle scheduling and timetabling. In: 20th symposium on algorithmic approaches for transportation modelling, optimization, and systems. ATMOS 2020, Schloss Dagstuhl-Leibniz-Zentrum für Informatik; 2020.

[27] Johnson Donald B. Finding all the elementary circuits of a directed graph. SIAM J Comput 1975;4(1):77–84.

[28] NS. Spoorkaart 2024: Hier te downloaden. 2023, URL https://nieuws.ns.nl/spoorkaart-2024-hier-te-downloaden/. [Accessed 5 November 2025].

[29] Kroon Leo, Huisman Dennis, Abbink Erwin, Fioole Pieter-Jan, Fischetti Matteo, Maróti Gábor, Schrijver Alexander, Steenbeek Adri, Ybema Roelof. The new dutch timetable: The OR revolution. Interfaces 2009;39(1):6–17.

[30] Van Lieshout Rolf. Integrated periodic timetabling and vehicle circulation scheduling. Transp Sci 2021;55(3):768–90.

[31] European Parliament. The future of european long-distance transport: Scenario-based analysis. Technical report, European Parliament, Directorate-General for Internal Policies; 2008, URL https://www.europarl.europa.eu/RegData/etudes/etudes/join/2008/417471/IPOL-JOIN_ET(2008)417471_EN.pdf.

[32] National Transport Authority. Alternative future scenario for travel demand. Technical report, National Transport Authority, Dublin; 2020, URL https://www.nationaltransport.ie/wp-content/uploads/2021/03/Alternative-Scenario-Development-Note-v-6.1_Final.pdf.

[33] Huisman Dennis, Kroon Leo G, Lentink Ramon M, Vromans Michiel JCM. Operations research in passenger railway transportation. Stat Neerl 2005;59(4):467–97.

[34] Delft High Performance Computing Centre (DHPC). DelftBlue supercomputer (phase 2). 2024, https://www.tudelft.nl/dhpc/ark:/44463/DelftBluePhase2. [Accessed 5 November 2025].