The Intermodal Railroad Blocking and Railcar Fleet-Management Planning Problem

Julie Kienzle
Serge Bisaillon
Teodor Gabriel Crainic
Emma Frejinger

November 2025

The Intermodal Railroad Blocking and Railcar Fleet-Management Planning Problem

Julie Kienzle[1,2*], Serge Bisaillon[1], Teodor Gabriel Crainic[1,3], Emma Frejinger[1,2]

1. Interuniversity Research Centre on Enterprise Networks Logistics and Transportation (CIRRELT)
2. Department of Computer Science and Operations Research, Université de Montréal
3. Department of Analytics, Operations, and Information Technologies, Université du Québec à Montréal

Abstract: Rail is a cost-effective and relatively low-emission mode for transporting intermodal containers over long distances. This paper addresses tactical planning of intermodal railroad operations by introducing a new problem that simultaneously considers three consolidation processes and the management of a heterogeneous railcar fleet. We model the problem with a scheduled service network design with resource management (SSND-RM) formulation, expressed as an integer linear program. While such formulations are challenging to solve at scale, we demonstrate that our problem can be tackled with a general-purpose solver when provided with high-quality warm-start solutions. To this end, we design a construction heuristic inspired by a relax-and-fix procedure. We evaluate the methodology on realistic, large-scale instances from our industrial partner, the Canadian National Railway Company: a North American Class I railroad. The computational experiments show that the proposed approach efficiently solves practically relevant instances, and that solutions to the SSND-RM formulation yield substantially more accurate capacity estimations compared to those obtained from simpler baseline models. Managerial insights from our study highlight that ignoring railcar fleet management or container loading constraints can lead to a severe underestimation of required capacity, which may result in costly operational inefficiencies. Furthermore, our results show that the use of multi-platform railcars improves overall capacity utilization and benefits the network, even if they can locally lead to less efficient loading as measured by terminal-level slot utilization performance indicators.

Keywords: Intermodal rail transportation, block planning, railcar fleet management, load planning, service network design, integer programming.

Results and views expressed in this publication are the sole responsibility of the authors and do not necessarily reflect those of CIRRELT.

Les résultats et opinions contenus dans cette publication ne reflètent pas nécessairement la position du CIRRELT et n'engagent pas sa responsabilité.

_____

* Corresponding author: kienzlejulie@gmail.com

1   Introduction

Railroads move large quantities of a wide range of valuable commodities over long distances. Rail is considered a cost-effective and relatively low-emission mode for long-distance ground transportation and can therefore play an important role in enhancing the sustainability of freight transportation.

Railroads are a key component of national and global supply chains and *intermodal transportation* that is steadily and strongly growing in volume and worth. Rail intermodal transportation means moving containers loaded in single or double stacks on particular railcars. Given its economic importance, intermodal transport has become a major component of railroad operations, planning and, management. It is well known that, in order to operate efficiently and profitably, railroads face challenging decision-making problems at all levels of planning, strategic, tactical, and operational. Intermodal services require particular attention in this context due to both customer-related factors (e.g., the need to synchronize rail–ship activities in ports and stricter-than-usual time requirements) and operational concerns (e.g., double-stack loaded intermodal railcars generally bypass automated sorting in classification yards). As a result, these services typically call for dedicated planning processes.

We focus on tactical planning of intermodal railroad operations in the North American market. This is a large market with infrastructure that enables double-stacking containers on railcars. While this increases the number of containers that can be transported within a given train length, it also makes the planning problem more complex. Indeed, there are several container and railcar types with specific loading constraints (Mantovani et al. 2017). We introduce the *Intermodal Railroad Blocking and Railcar Fleet-Management* (*IBRM*) problem. Taking a train service schedule as given by the railroad, we aim to build an economically and customer-service efficient plan that simultaneously considers the selection of extra services, the loading of containers on railcars, the blocking of loaded and empty railcars, the selection of blocks, the distribution of demand flows through the service network, and the management of the railcar fleet. The plan is built for a cyclic schedule of given length (e.g., a week), intended to be repeatedly executed over the tactical-planning horizon (e.g., a season).

Particularly challenging issues in addressing the IBRM are (i) simultaneously considering three consolidation processes: containers to railcars, railcars to blocks, and blocks to trains; (ii) differentiating railcar and container types and representing, in a computationally efficient way, the loading of containers onto railcars within a tactical model; and (iii) integrating the management of a limited heterogeneous railcar fleet with the design of the blocking and service network.

While there is an extensive literature on railroad planning, relatively few studies focus on integrated planning and intermodal rail transport. For example, Morganti et al. (2020) is the study closest to our work, as they focus on intermodal block planning; however, they do not consider railcar fleet management nor extra train service selection.

This paper offers several contributions. *First*, we introduce the IBRM problem, which is new to the literature, and propose an enhanced *Scheduled Service Network Design with Resource Management* (*SSND-RM*) model (Crainic and Hewitt 2021). The SSND-RM model takes the form of an integer linear programming model (ILP) based on a cyclic four-layer space time network representation (namely, the container, car, block and train layers). This approach enables the use of a continuous-time network representation, where the time structure is defined by the arrival and departure times of the train services considered at the terminals on their respective routes.

Second, we propose a construction heuristic inspired by a relax-and-fix procedure. We use the heuristic to compute high-quality warm-start solutions for general-purpose ILP solvers. This simple approach is appealing from a practical point of view, as a general-purpose solver can be used as a black box. Using this approach, we show that we can solve large-scale realistic

instances from our industrial partner, the Canadian National Railway Company, in a reasonable time.

Third, by comparing solutions with those obtained by solving simpler baseline models, we show the significance of the problem we introduce. Based on extensive numerical experiments and analyses, we offer several managerial insights. In brief, ignoring loading constraints and railcar fleet management can result in severe underestimation of required capacity; we link characteristics of demand (number of containers of different types) to an adequate composition of the railcar fleet; and we highlight the importance of multi-platform cars from a capacity usage point of view.

The structure of the paper unfolds as follows: Section 2 describes the problem in detail and Section 3 gives a review of the related literature. We describe the network representation in Section 4 followed by the SSND-RM formulation in Section 5. Section 6 outlines the solution approach, and Section 7 describes the computational results based on large-scale instances from our industrial partner. Finally, Section 8 concludes the paper.

2 Problem Description

We consider a tactical planning problem facing railroads. The tactical plan is cyclic over a given schedule length T, for example, a week, and is repeated over a medium-term planning horizon (e.g., three months). The aim is to guide intermodal rail operations over a physical network. To describe the problem in detail, we start by introducing core concepts – scheduled train service supply, container demand, and railcars – followed by a description of loading rules, blocks, and the decision-making problem. We summarize the related notation in Table 1.

Train services. Railroads operate an infrastructure network made up of rail tracks connecting terminals $\theta \in \Theta$ that are dedicated, totally or partially, to intermodal traffic. We consider a set of scheduled train services $\sigma \in \Sigma$. The services are dedicated to intermodal traffic and each train service σ has an origin terminal o_σ with departure time α_σ, and a destination terminal d_σ with arrival time β_σ. It also has a set of intermediate stops $\Theta_\sigma^{\text{INT}}$, each $i(\sigma) \in \Theta_\sigma^{\text{INT}}$ with arrival $\beta_{i(\sigma)}$ and departure $\alpha_{i(\sigma)}$ times. We define a leg as the route between two consecutive stops. The service capacity is expressed in train length (we use feet, shorthand ft) that may vary by leg. We note that intermodal traffic is typically light compared to general cargo, the limiting factor is therefore train length rather than tonnage.

We consider two types of scheduled train services: First, regular services, Σ_{INIT}, designed to transport the majority of the recurrent demand. Second, potential extra train services, Σ_{EXTRA}, that are to be selected, and operated at fixed cost, $f_\sigma, \sigma \in \Sigma_{\text{EXTRA}}$, when needed. An extra service is activated only when loaded at least at $U_\sigma\%$, $\sigma \in \Sigma_{\text{EXTRA}}$, of its capacity on its leg set. Extra services can include using capacity on general cargo trains, or dedicated extra intermodal trains. The scheduled services $\Sigma = \Sigma_{\text{INIT}} \cup \Sigma_{\text{EXTRA}}$ are given, extra ones being potential only. The selection of a subset of extra trains is therefore part of the tactical planning problem.

Containers. Intermodal containers come in different sizes and types. For example, the North American market has high- and low-cube containers of various sizes (e.g., 20-, 40-, and 53-ft). Unlike operational load planning problems (Mantovani et al. 2017), only high-level information is available at the tactical planning stage. As we further detail below, crucial in this context are aspects that impact the usage of railcars, and hence the capacity usage of the train. Let $k \in \mathcal{K}$ denote intermodal demand, defined as a number v_k of containers of a specific type $\tau_k \in \mathcal{T}$ arriving at origin terminal o_k at time α_k, loaded on railcars and moved to their destination terminal d_k before a certain due-date β_k. We consider two container types $\mathcal{T} = \{\tau_{40}, \tau_{53}\}$, as two 20-ft containers can be defined as one 40-ft-long box, while containers longer than 40 ft (e.g., 45 ft) may be treated as 53-ft-long ones as they follow similar loading rules.

Railcars. The North American fleet contains several different railcar types. Mantovani et al. (2017) discuss these in depth. Here, we describe the information that is relevant to tactical planning. The types of railcars $\gamma \in \Gamma$ differ in the number of platforms, 1, 3, or 5, and their respective size (length) $\pi \in \Pi$, and can accommodate single- or double-stacked containers. The most common platform types in the North American market are 40 ft and 53 ft long. Each platform is shaped like a well to lower the center of mass of loaded containers. The space is therefore physically restricted by the platform size. Let η_π^γ denote the number of platforms of type $\pi \in \Pi$ on railcar type $\gamma \in \Gamma$ (all the platforms that make up a railcar are of the same type), and let Γ_π denote the set of railcar types made up of platform type $\pi \in \Pi$. Let λ_γ be the length of railcar type $\gamma \in \Gamma$.

For the tactical planning problem, it is important to note that, for a given platform size π, the ratio $\lambda_\gamma/\eta_\pi^\gamma$ decreases with increasing number of platforms. This means that, for example, a five-platform railcar uses less train capacity per platform, than a one-platform railcar. However, railcars with more than one platform have additional loading restrictions. We attend to those next.

Loading Rules. Each platform may be double-stacked (two container slots, one at the bottom and one on top) or single-stacked (one slot) and loading rules dictate which configurations of container types are allowed on which railcar type (Mantovani et al. 2017). Here, we focus on rules pertaining to container size, as this is the only information available at the tactical planning level.

First we consider the bottom slot as the loading rule simply dictates that the platform needs be of sufficient size. That is, independently of the railcar type, a 40-ft container can be loaded in the bottom slot on either a 40-ft or 53-ft platform, whereas a 53-ft container requires a 53-ft platform.

The top slot size is not physically restricted by a well. So in principle, a 53-ft or 40-ft container could be loaded in the top slot, independently of the container size loaded in the bottom slot. This is true for 40-ft railcars with one platform and all 53-ft railcar types (see the bottom part of Figure 1 for an example of a loading that respects the rules). However, for 40-ft railcars with more than one platform, the limited distance between platforms does not allow to load 53-ft containers in all top slots. Therefore, 40-ft containers can be loaded in any top slot, but 53-ft containers can only be loaded in every second slot (see the top part of Figure 1 for an example of a loading that does not respect the rules). This means that the maximum number of 53-ft containers on railcars with three 40-ft platforms is two, the corresponding figure for five 40-ft platforms is three, etc.

These loading rules imply that several demands with possibly different container types and time characteristics can be loaded together on a same railcar as long as the loading rules are respected.
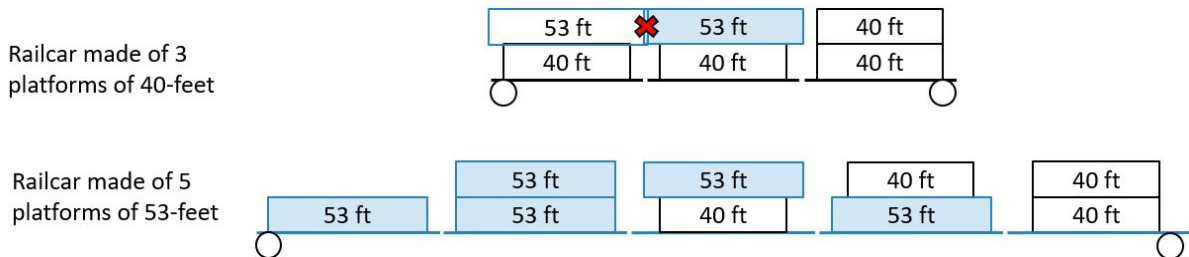

Figure 1: Illustrative examples of loading rules (not respected in the top figure, respected in the bottom figure)

Blocks. A block $b \in \mathcal{B}$ is a consolidation of railcars that move as a single unit between a pair of terminals taking advantage of economies of scale and reducing terminal handling costs. Blocks are assigned to a sequence of train services. Blocking of intermodal traffic differs from its general cargo counterpart. Whereas it is customary to sort most railcars at so-called classification yards, this is less common for intermodal railcars. For instance, double-stack railcars cannot be sorted in hump yards for physical reasons (it could destabilize the load). This means that the intermodal railcars that form a given block all have the same destination, which is the destination of the block.

As such, a block $b \in \mathcal{B}$ is defined by an origin terminal o_b where it is built by grouping the classified (sorted) railcars assigned to it, the time of departure at the origin terminal α_b, the destination terminal d_b where it is dismantled, the time of arrival at destination β_b, the sequence of trains moving the block from its origin to its destination, and the terminals where they are being transferred from one train to another with the respective arrival and departure times given by the associated train-service schedules. A block has a maximum capacity u_b that is limited by the train capacity on each train leg.

To illustrate the basic concept, we show a stylized example in Figure 2. It depicts three train services σ_1, σ_2 and σ_3 (arrows with dashed, solid, and dotted lines, respectively). There are five terminals represented on the vertical axis ($\theta_1, \ldots, \theta_5$) and time is represented on the horizontal axis. Each train has an origin and a destination terminal and are respectively composed of three, two, and one legs. We represent four demands at the bottom part of the figure, each with an origin and a destination. Note that demands k_2 and k_3 share the same origin, but they have different destinations. For the sake of simplicity, we do not explicitly represent containers and railcars in this figure. The demand is transported on four blocks (see highlighting next to the trains with the block identification at the origin): b_1 is transported by trains σ_1 and σ_2, b_2 by σ_2 and σ_3, whereas b_3 and b_4 are transported by a single train (σ_2 and σ_1, respectively).
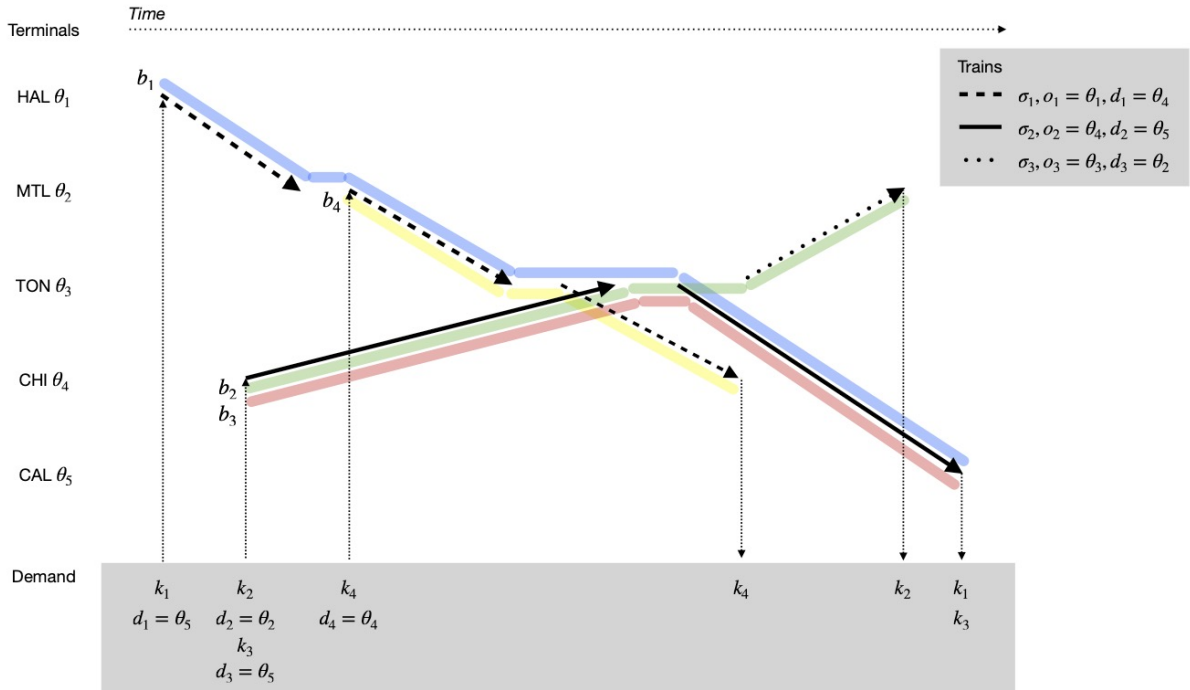

Figure 2: Illustrative example of trains, blocks and demands

We note that a given demand k can be split on multiple railcars, and even on multiple blocks. Let $\mathcal{B}_k \subseteq \mathcal{B}$ denote the set of blocks that can transport demand $k \in \mathcal{K}$ given its time

characteristics, while $\mathcal{K}_b$ is the set of demands that can be assigned to block $b \in \mathcal{B}$. Let c_{bk} be the unit cost of transporting a container of demand k on block $b \in \mathcal{B}_k$, and $c_{b\gamma}$ the unit cost of using railcar type γ in block $b \in \mathcal{B}$. The tactical plan selects the subset of blocks in $\mathcal{B}$ that should be built. Fixed costs f_b, $b \in \mathcal{B}$, reflect the time and resources required to build a block at its origin terminal and to transfer it between services, when relevant. The transportation of demand that cannot be satisfied by the service network defined by the tactical plan is outsourced at a relatively high cost c_k, $k \in \mathcal{K}$.

Railcar fleet management. We consider a heterogeneous fleet of railcars where we denote by H^γ the total number of railcars of type $\gamma \in \Gamma$. Because there is a limited railcar supply in the network, it is essential to efficiently manage these resources to bring the right type of railcars to the right location in the network at the right time. This is a challenging aspect of the problem as demand may be unbalanced between the origin-destination pairs, and the right mix of railcar types depends on the specific mixes of container types (see discussion on loading rules). Some railcars might, thus, have to travel empty so as to be distributed where they are needed (e.g., at terminals where a railcar shortage is occurring). When the fleet composition is modified due to, e.g., seasonal changes or new railcar acquisitions, railroad companies have to decide at which terminals the railcars should be first stationed. In this case, in addition to providing a railcar circulation (i.e., determining the flow of railcars within the network to support the planned services), the railcar fleet management plan should also give the initial allocation of railcars to the different terminals in the network. Let $c_{\gamma\theta}$ denote the unit cost of initially allocating railcar type $\gamma \in \Gamma$ to terminal $\theta \in \Theta$.

Decision-making problem. Given a train schedule, a set of demands, and a heterogeneous fleet of railcars, the goal of the IBRM problem is to determine simultaneously a scheduled blocking plan that includes container-to-railcar (loading problem) and railcar-to-block (blocking problem) assignments, a railcar circulation to support the selected blocks with, when wanted, an initial railcar allocation to terminals (railcar fleet management problem), the demand itineraries, including, when needed, the volumes assigned to general cargo trains or trucks, and finally, when desired, a selection of extra trains to add to the schedule to satisfy demand; all of this while minimizing the total operational costs. As a benefit, managerial strategic insights, such as determining an effective railcar fleet composition, can also be derived from the provided tactical plan.

3 Literature Review

There is a rich and long history of successful Operations Research developments and contributions targeting railroad planning. A comprehensive review of this literature goes far beyond the scope of this paper. The interested reader may turn to a series of surveys that synthesize this story and contributions, including Assad (1980b), Dejax and Crainic (1987), Crainic (1988), Crainic and Laporte (1997), Cordeau et al. (1998), Crainic (2000), Newman et al. (2002), Crainic (2003), Ahuja et al. (2005a), Crainic and Kim (2007), Bektaş and Crainic (2008), Crainic (2009), Yaghini and Akhavan (2012), Piu and Speranza (2014) and Chouman and Crainic (2021). The Service Network Design methodology and contributions are synthesized and reviewed in Crainic and Hewitt (2021), Crainic (2025b,a) and Crainic and Rei (2025). Several observations and trends stand out when examining this literature and allow one to situate our contribution within the field.

First, few studies target intermodal rail transportation planning. We discuss them in the last part of the section, following a rapid overview of developments relevant to the problem we address.

Infrastructure	
Θ	Set of terminals, $\theta \in \Theta$
t^{TRANS}	Block transfer time at any terminal

Train services	
Σ	Set of scheduled train services, $\sigma \in \Sigma$
Σ_{INIT}	Set of regular scheduled train services
Σ_{EXTRA}	Set extra (non regular) scheduled train services
o_σ	Origin terminal of $\sigma \in \Sigma$
α_σ	Departure time of $\sigma \in \Sigma$
d_σ	Destination terminal of $\sigma \in \Sigma$
β_σ	Arrival time of $\sigma \in \Sigma$
$\Theta_\sigma^{\text{INT}}$	Set of intermediate stops of $\sigma \in \Sigma$
$\beta_{i(\sigma)}$	Arrival time at intermediate stop (node) $i(\sigma) \in \Theta_\sigma^{\text{INT}}$, $\sigma \in \Sigma$
$\alpha_{i(\sigma)}$	Departure time at intermediate stop (node) $i(\sigma) \in \Theta_\sigma^{\text{INT}}$, $\sigma \in \Sigma$
f_σ	Fixed cost of using $\sigma \in \Sigma_{\text{EXTRA}}$
U_σ	Capacity usage threshold for $\sigma \in \Sigma_{\text{EXTRA}}$

Demand	
$\mathcal{K}$	Set of demands, $k \in \mathcal{K}$
o_k	Origin terminal of $k \in \mathcal{K}$
α_k	Arrival time at origin of $k \in \mathcal{K}$
d_k	Destination terminal of $k \in \mathcal{K}$
β_k	Due date at destination of $k \in \mathcal{K}$
υ_k	Number of containers of $k \in \mathcal{K}$
$\mathcal{T}$	Set of container types, $\tau \in \mathcal{T}$; $\mathcal{T} = \{\tau_{40}, \tau_{53}\}$ in current setting
τ_k	Container type of $k \in \mathcal{K}$
c_k	Unit outsourcing cost $k \in \mathcal{K}$

Railcars	
Γ	Set of railcar types, $\gamma \in \Gamma$
λ_γ	Length of railcar type $\gamma \in \Gamma$
Π	Set of platform sizes, $\pi \in \Pi$ $\Pi = \{\pi_{40}, \pi_{53}\}$ in current setting
Γ_π	Set of railcar types made up of platform type $\pi \in \Pi$
η_π^γ	Number of platforms of type π for railcar $\gamma \in \Gamma$
H^γ	Number of railcars of type $\gamma \in \Gamma$ available in the system
$c_{\gamma\theta}$	Unit cost for intially allocating railcar type $\gamma \in \Gamma$ to terminal $\theta \in \Theta$

Blocks	
$\mathcal{B}$	Set of blocks, $b \in \mathcal{B}$
$\mathcal{B}_k$	Set of blocks which may transport demand $k \in \mathcal{K}$, $\mathcal{B}_k \subseteq \mathcal{B}$
o_b	Origin terminal of $b \in \mathcal{B}$
α_b	Departure time of $b \in \mathcal{B}$
d_b	Destination terminal of $b \in \mathcal{B}$
β_b	Arrival time of $b \in \mathcal{B}$
f_b	Cost of building $b \in \mathcal{B}$
u_b	Capacity of $b \in \mathcal{B}$
$\mathcal{K}_b$	Set of demands that can be assigned to $b \in \mathcal{B}$
c_{bk}	Unit cost of moving a container of demand k on $b \in \mathcal{B}$
$c_{b\gamma}$	Unit cost of moving a railcar of type γ on $b \in \mathcal{B}$

Table 1: Summary of notation related to core concepts and problem parameters

Second, one observes that most early contributions focus on single problems or combinations of a limited number of them, blocking, service selection, and resource management in particular. Blocking (see, e.g., Bodin et al. 1980, Newton et al. 1998, Ahuja et al. 2007, Jha et al. 2008, Yaghini and Akhavan 2012, Yaghini et al. 2021) is assumed to be addressed first (in contrast, Morganti et al. 2020, assumes a given service schedule), followed by service selection and makeup (see, e.g., Morlok and Peterson 1970, Assad 1980a, Nozick and Morlok 1997, Newman and Yano 2000, Yaghini and Akhavan 2012). In most cases, network design type formulations (Magnanti and Wong 1984, Crainic et al. 2021) are proposed to select blocks or services, from given sets, respectively.

The management of resources is then addressed as a different planning activity, focusing on distribution and routing given a set of services. To address the challenge of unbalanced trade, resources that complete their assignments and are needed in terminals different from their current locations are repositioned to undertake the next round of activities. The activity is often characterized as "operational", even though resources support services and, accordingly, the planning horizon of the previous two activities is shared. In general, network flow optimization methods are proposed in this context, evolving from transportation/assignment models (see, e.g., Bomberault and White 1966, White 1968, White and Bomberault 1969, White 1972) to multicommodity integer flow time-space formulations integrating various practical rules and limitations (see, e.g., Florian et al. 1976, Joborn et al. 2004, Ahuja et al. 2005b, Vaidyanathan et al. 2008a,b, Balakrishnan et al. 2016, Bouzaïene-Ayari et al. 2016, Piu et al. 2015, Ortiz-Astorquiza et al. 2021, Miranda et al. 2020).

Third, a trend can be observed in railroad tactical planning toward comprehensive models that integrate the main system and operational components, e.g., selection, scheduling and makeup of services, railcar classification and block design, resource assignment and management, and demand routing through the selected service network. Service Network Design (SND) appears as the methodology of choice to address these problems (Chouman and Crainic 2021). Our work belongs to this railroad planning and OR methodological development area.

Crainic et al. (1984) propose what is probably the first railroad tactical planning SND model (Crainic and Rousseau 1986, generalizes it for consolidation-based multicommodity multimode freight transportation systems). The model integrates service selection, service frequency optimization, car classification and blocking, train makeup, and freight routing. The nonlinear path-based formulation also accounts for congestion and delays in terminals and on the network's tracks, as well as for the (re-)positioning of empty railcars through one or several OD demand matrices (generated through demand-distribution models from the surplus and penury levels at terminals derived from the loaded demand).

Haghani (1989) presents a SSND model combining train routing, scheduling, and make-up, as well as empty car distribution, on a space-time network. A heuristic is used to address a somewhat simplified version of the model and illustrate the interest of integrated planning. The model proposed by Keaton (1989, 1992) aims to determine the pairs of terminals to connect by direct services, whether to offer more than one train a day, as well as the routing of freight and the blocking of railcars. Gorman (1998) starts from the previous model to design, using a tabu-enhanced genetic search metaheuristic, a scheduled operating plan that follows as much as possible the particular operation rules of a major North-American railroad. All these contributions model blocking through classification costs, rather than explicit blocking decision variables.

Zhu et al. (2014) propose what appears to be the first comprehensive SSND railroad planning model, integrating the selection of scheduled services, selection of blocks, service makeup, and railcar classification and routing. The model is built on a cyclic, multi-layer (Crainic et al. 2022, Crainic 2024) time-space network corresponding to the discretized schedule length. The layers corresponds to the railcar terminal activities (demand entry and exit, and railcar classification and blocking), block terminal activities, and train service operations. The authors propose a

matheuristic combining dynamic block generation and slope scaling, Tabu Search and ellipsoidal search methods. The authors do not address resource management issues.

Pedersen and Crainic (2007), Pedersen et al. (2009) focus on integrating the management of one type of resource, namely, locomotives or intermodal shuttles (i.e., locomotive and fixed set of container-carrying railcars), into tactical-planning SSND formulations. The authors formalize the concept of design-balanced SSND, forcing equal numbers of selected services arriving at or leaving from any terminal, at all relevant time moments. Pedersen and Crainic (2007), Vu et al. (2013), Chouman and Crainic (2015) present metaheuristics to address the resulting SSND.

Andersen et al. (2009a) enlarge the scope of the previous problem setting and devise SSND with resource management models to address the coordination and synchronization of services and fleets of several railroads and navigation companies at particular junction points (train services operating in several countries need locomotive changes at some borders due to differences in electric tensions). Andersen et al. (2009b) then observe that resources that support services or that are repositioned operate according to cycle schedules, returning to their home bases after certain periods of time. The authors compare SSND arc and cycle-based formulations and show that the latter provide more modeling flexibility, in terms of cycle duration for example, and computational efficiency, provided that cycles are generated a priori. Andersen et al. (2011) propose a branch-and-price with column generation algorithm to address the latter issue. Crainic et al. (2014, 2018) extend this work in the general SSND context for consolidation-based freight carriers (with one unit of resource per service only) in two ways: 1) a larger range of resource management concerns, e.g., multiple resource types, acquisition and allocation/reallocation of resources at terminals, outsourcing services, and cycle duration rules; 2) a matheuristic that incorporates slope scaling and network flow-based cycle generation heuristics.

To conclude this brief review, we return to the limited number of contributions aimed at planning intermodal railroad transport. Nozick and Morlok (1997) minimizes the total moving cost of flatcars and trailers (1 trailer loaded on a flatcar) given a set of services. Newman and Yano (2000) proposes a day-of-week uncapacitated (in the number of trains one can make up in a yard and operate on a line) frequency SND model, to determine whether homogeneous container OD demand should be moved by a direct or indirect (through a unique main yard) service. Müller et al. (2021) propose a SSND with stochastic demand for the case of a homogeneous set of services (called "vehicles" in the paper) managed through design-balancing constraints. A very simple loading rule (one container) characterize these three contributions, which do not consider the selection and formation of services and blocks.

Morganti et al. (2020) start to address the loading issue when formulating their SSND model for the block planning problem for intermodal rail transport with double-stack railcars. Operational loading requires detailed information about containers, such as their individual weights (Mantovani et al. 2016), and is hence too detailed for tactical planning. Morganti et al. (2020) therefore propose a first approximation of those rules, based on the length of the main container and railcar platform types, together with constraints enforcing the length of selected blocks and of the services to which they are assigned. The space-time network is discretized according to the schedules of the given set of services, their respective times of arrival and departure at terminals yielding the time moments of the network. We adopt the same modeling of time in our work. Unlike us, Morganti et al. (2020) do not consider service selection or resource management.

In conclusion, we note that the current state-of-the-art in railroad tactical planning does not address the main challenges of intermodal systems. The methodology we propose contributes to filling this gap, integrating within a comprehensive SSND model service and block selection, realistic tactical-level container-to-railcal loading constraints, and the management of multiple sets of railcars that may be combined, loaded or empty, when building the blocks.

4 Network Modeling

The multi-layer network structure is presented in Section 4.1, introducing the time representation and the train-service layer. The other layers and the inter-layer arcs are then described by following a demand itinerary, from the initial processing at the origin terminal (Section 4.2), to moving through the network (Section 4.3), and the final handling at destination (Section 4.4). Recall that the tactical plan defines the demand itineraries, the blocks to build, the extra services to run, and the management of the railcar fleet. Table 2 summarizes the time-space network notation.

4.1 General Structure and Train Layer

We propose a four-layer (Crainic et al. 2022, Crainic 2024) cyclic time-space network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ of length T to represent the system dynamics, the activities, and the decisions of the IBRM problem. Nodes $i \in \mathcal{N}$ represent events at specific moments in time $t(i)$ and terminals $\theta(i)$. Arcs stand for activities taking place in time and space between these nodes.

The layers correspond to the main components of the system, train services, blocks, railcars, and containers. *Intra-layer* nodes and arcs model the activities corresponding to each layer at appropriate moments in time, while *inter-layer* arcs capture the interactions among the components, e.g., loading/unloading of containers onto/from railcars, consolidating the latter into blocks, attaching/detaching the latter to/from train services, and dismantling blocks at destination. Due to the cyclic nature of the tactical plan, activities initiated before the end of the schedule may end after T, that is, during the next application of the plan. This is modeled by having the corresponding arcs wraparound, times computed $modulo(T)$ (Chouman and Crainic 2021, Crainic and Hewitt 2021). The network is illustrated in Figure 3 for a terminal $\theta \in \Theta$ and a representative time interval featuring a train arrival and a train departure. We refer to this figure throughout the section.
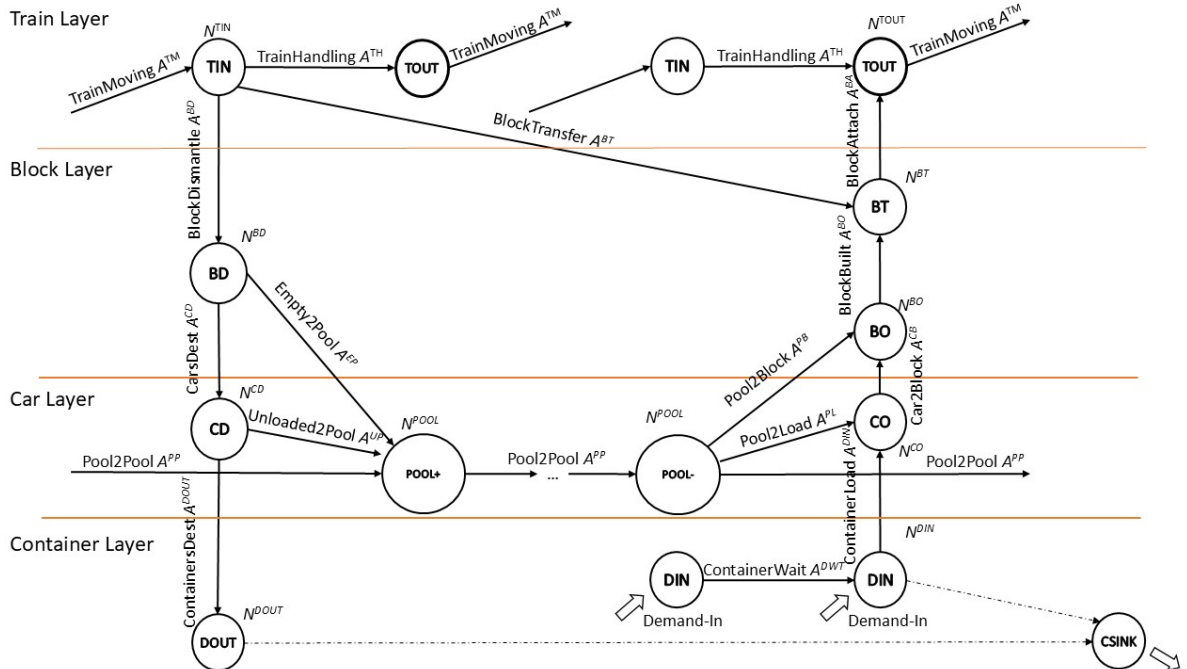

Figure 3: Four-layer (train, block, car, container) time-space network illustration

The *container layer* represents the arrival of each demand $k \in \mathcal{K}$ at its origin yard, the possible waiting before being loaded onto railcars, the actual loading operation, the unloading

at destination, and the subsequent exit from the system. The *car layer* connects the container-based demand to the system and governs how containers are loaded on railcars, that are grouped into blocks and then into trains. It is also in the *car layer* that the fleets of the various types of railcars are managed through the assignment to container-loading activities, the empty movements to balance the needs within the network, the railcar-pool inventory counts at terminals, as well as the initial fleet size and allocation at terminals. Blocks are handled on the *block layer*: building at origin by consolidating empty and loaded railcars, attachment to a train, transfer between two train services, and detachment from the last train at destination for dismantling. Detailed descriptions of the network components and modeling of these three layers are given in the next subsections.

The *train layer* plays a special role in the proposed formulation. It obviously represents the movements between terminals and the time spent there to pick up or drop off the blocks of regular and extra train services. It also defines the time instances of the time-space network and, thus, the pacing of activities and decisions. Recall that, most time-space networks proposed in the service network design literature, including models targeting rail planning (see, e.g., Chouman and Crainic 2021, Crainic and Hewitt 2021), generate time-space networks by 1) discretizing the schedule length according to a given granularity and 2) duplicating each node of the physical network at each of the resulting periods. We take a different approach and define a *continuous-time* network based on the schedules of the given train-service set Σ.

The arrival and departure times of each train service to/from each terminal on its route yield corresponding arrival (TIN) and departure (TOUT) nodes, defining the time structure of the entire network (Figure 3). Let $\mathcal{N}_\theta^{\text{TOUT}}$ and $\mathcal{N}_\theta^{\text{TIN}}$ be the sets of TOUT and TIN nodes at terminal $\theta \in \Theta$, sorted in increasing time order, with $\mathcal{N}^{\text{TOUT}} = \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{TOUT}}$ and $\mathcal{N}^{\text{TIN}} = \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{TIN}}$. Then, any node $i \in \mathcal{N}$, representing an event on any layer, has its time stamp $t(i)$ equal to either a train arrival or departure moment at terminal $\theta \in \Theta$, i.e., $i \in \mathcal{N}_\theta^{\text{TIN}}$ or $i \in \mathcal{N}_\theta^{\text{TOUT}}$, respectively. In other words, container, railcar, and block activities are synchronized with train departures and arrivals, the associated waiting being represented on the inter-layer arcs.

Notice that, more than one service could arrive at or leave a terminal at any given time moment, generating several simultaneous nodes. Such an event is rare, however. Hence, to simplify the presentation, and without loss of generality, we assume a single service arrival or departure at any time moment.

4.2 Embarking on a Train

We describe in the following the "embarking" side of the container, car, and block layers. We discuss the initial allocation of the railcar fleets to terminals in the car layer subsection.

4.2.1 The Container Layer

models demand entering (and existing) the system with the set of nodes $\mathcal{N}^{\text{DIN}} = \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{DIN}}$. A demand $k \in \mathcal{K}$ leaves its origin terminal o_k at the earliest when the first train, following its own arrival time at the terminal, departs. As such, a node is generated in the container layer every time a train leaves (instead of every time a demand enters). Accordingly, a demand $k \in \mathcal{K}$ arriving to o_k at time α_k connects to nodes $i \in \mathcal{N}_{o_k}^{\text{DIN}}$ such that $t(i) \geq \alpha_k$. Recall that demands can be divided and delivered using different blocks and trains. This is modeled with a set of *ContainerWait* arcs $\mathcal{A}^{\text{DWT}} = \cup_{\theta \in \Theta} \{a = (i,j), j = i+1, i \in \mathcal{N}_\theta^{\text{DIN}}, \theta \in \Theta\}$ where $j = i+1$ represents the next-in-time DIN node in terminal θ.

The operation of loading demand onto railcars at the origin terminal is represented by inter-layer *ContainerLoad* arcs $\mathcal{A}^{\text{DIN}} = \cup_{\theta \in \Theta} \mathcal{A}_\theta^{\text{DIN}} = \{a = (i,j) | t(i) = t(j), i \in \mathcal{N}_\theta^{\text{DIN}}, j \in \mathcal{N}_\theta^{\text{CO}}, \theta \in \Theta\}$ linking DIN nodes in the container layer to container-railcar loading nodes CO in the car layer. The container-to-railcar assignments and loading are performed according to the rules introduced in Section 2.

$\mathcal{G} = (\mathcal{N}, \mathcal{A})$	Time-space network, sets of nodes $\mathcal{N}$ and arcs $\mathcal{A}$
$\theta(i), t(i)$	Terminal and time moment of node $i \in \mathcal{N}$
T	Schedule length
u_a	Capacity of arc $a \in \mathcal{A}$

Train Layer

$\mathcal{N}^{\text{TIN}}$	Train-service arrival nodes & times at terminals (TIN), $= \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{TIN}}$
$\mathcal{N}^{\text{TOUT}}$	Train-service departure nodes & times at terminals (TOUT), $= \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{TOUT}}$
$\mathcal{A}_\sigma^{\text{TM}}$	*TrainMoving* (TOUT, TIN) moving arcs of service $\sigma \in \Sigma$
$\mathcal{A}_\sigma^{\text{UTM}}$	Leg set of extra train service $\sigma \in \Sigma_{\text{EXTRA}}$ for which a minimum load U_σ is required in order for the service to be considered for selection
$\mathcal{A}_\sigma^{\text{TH}}$	*TrainHandling* (TIN, TOUT) handling arcs at terminals serviced by $\sigma \in \Sigma$
$\mathcal{B}_a$	Set of blocks on train-service arc $a \in \mathcal{A}_\sigma^{\text{TM}} \cup \mathcal{A}_\sigma^{\text{TH}}$

Block Layer

$\mathcal{N}_\theta^{\text{BO}}$	Block forming nodes (BO)
$\mathcal{N}_\theta^{\text{BD}}$	Block-at-destination nodes (BD)
$\mathcal{N}_\theta^{\text{BT}}$	Block transfer and assembly nodes (BT)
$\mathcal{B}_\theta^+$	Blocks formed at terminal $\theta \in \Theta$
$\mathcal{B}_{\theta i}^+$	Blocks formed at terminal $\theta \in \Theta, t(i) = \alpha_b, i \in \mathcal{N}_\theta^{\text{POOL}-}$
$\mathcal{B}_{\theta i}^-$	Blocks dismanteled at terminal $\theta \in \Theta, t(i) = \beta_b, i \in \mathcal{N}_\theta^{\text{POOL}+}$
$\mathcal{A}^{\text{BB}}$	*BlockBuild* arcs moving new blocks to the assembly node (BT)
$\mathcal{A}^{\text{BT}}$	*BlockTransfer* inter-layer arcs, bring transferred blocks to the assembly node
$\mathcal{A}^{\text{BA}}$	*BlockAttach* inter-layer arcs, move collected blocks to the outgoing train
$\mathcal{A}^{\text{BD}}$	*BlockDismantle* inter-layer arcs, take blocks off trains at destination

Car Layer

$\mathcal{N}^{\text{POOL}}$	Railcar-inventory nodes (POOL), $= \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{POOL}}$
$\mathcal{N}_\theta^{\text{POOL}}$	Terminal $\theta \in \Theta$ railcar-inventory nodes, $= \mathcal{N}_\theta^{\text{POOL}+} \cup \mathcal{N}_\theta^{\text{POOL}-} = \cup_{\gamma \in \Gamma} \{ \mathcal{N}_{\gamma\theta}^{\text{POOL}+} \cup \mathcal{N}_{\gamma\theta}^{\text{POOL}-} \}$
$\mathcal{N}_{\gamma\theta}^{\text{POOL}+}$	Terminal $\theta \in \Theta$ inventory nodes linked to the arrival of trains with possibly increasing railcar-type $\gamma \in \Gamma$ inventory
$\mathcal{N}_{\gamma\theta}^{\text{POOL}-}$	Terminal $\theta \in \Theta$ inventory nodes linked to the departure of trains with possibly decreasing railcar-type $\gamma \in \Gamma$ inventory
$\mathcal{N}_\theta^{\text{CO}}$	Container-to-railcar loading nodes (CO), $\theta \in \Theta$
$\mathcal{N}_\theta^{\text{CD}}$	Container-from-railcar unloading nodes (CD), $\theta \in \Theta$
$\mathcal{A}^{\text{PP}}$	*Pool2Pool* railcar inventory carrying arcs
$\mathcal{A}^{\text{PL}}$	*Pool2Load* empty-railcar-to-container assignment & loading arcs
$\mathcal{A}^{\text{PB}}$	*Pool2Block* empty-railcar inter-layer blocking arcs
$\mathcal{A}^{\text{CB}}$	*Cars2Blocks* loaded-railcar inter-layer blocking arcs
$\mathcal{A}^{\text{EP}}$	*Empty2Pool* empty-railcar at destination inter-layer disassemble arcs
$\mathcal{A}^{\text{CD}}$	*CarsDest* loaded-railcar at destination inter-layer disassemble arcs
$\mathcal{A}^{\text{UP}}$	*Unloaded2Pool* arcs moving unloaded (empty) railcars to their inventory pool

Container Layer

$\mathcal{N}^{\text{DIN}}$	Demand-arrival DIN nodes, $= \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{DIN}}$
$\mathcal{N}_{o_k}^{\text{DIN}} \subseteq \mathcal{N}_\theta^{\text{DIN}}$	Possible DIN nodes for demand $k \in \mathcal{K}$
$\mathcal{N}^{\text{DOUT}}$	Demand-at-destination DOUT nodes, $= \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{DOUT}}$
$\mathcal{A}^{\text{DIN}} = \cup_{\theta \in \Theta} \mathcal{A}_\theta^{\text{DIN}}$	*ContainerLoad* container-to-car origin inter-layer arcs
$\mathcal{A}^{\text{DWT}}$	*ContainerWait* arcs
$\mathcal{A}^{\text{DOUT}}$	*ContainerDest* car-to-container destination inter-layer arcs

Table 2: Time-space network notation

4.2.2 The Car Layer

Four main activities are modeled in this layer: container loading (and unloading), railcar-to-block assignment, railcar fleet management, and the initial dimensioning of the fleet and its assignment to terminals.

We explicitly represent the railcar circulation within the cyclic scheduled service network to control the inventories and flows of each railcar type. For this purpose, we introduce POOL nodes to model inventory changes to each railcar type at each terminal and relevant time moment. Railcar inventories at terminals *decrease* when railcars are assigned loaded or empty to a block. We define the associated set of inventory nodes linked to departing trains $\mathcal{N}_\theta^{\text{POOL}-} = \cup_{\gamma \in \Gamma} \mathcal{N}_{\gamma\theta}^{\text{POOL}-} := \{i \mid t(i) = \alpha_\sigma, \alpha_{j(\sigma)}, j(\sigma) = 1, \ldots, |\Theta_\sigma^{\text{INT}}|, \sigma \in \Sigma\}$, $\gamma \in \Gamma$, $\theta \in \Theta$. We model the assignment of loaded and empty railcars to blocks with intra-layer arcs *Pool2Load* $\mathcal{A}^{\text{PL}} := \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{POOL}-}, j \in \mathcal{N}_\theta^{\text{CO}}, \theta \in \Theta\}$ and inter-layer arcs *Pool2Block* $\mathcal{A}^{\text{PB}} := \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{POOL}-}, j \in \mathcal{N}_\theta^{\text{BO}}, \theta \in \Theta\}$, respectively, connecting the pool to the block-forming BO node in the block layer. Let $\mathcal{B}_\theta^+ = \cup_{i \in \mathcal{N}_\theta^{\text{POOL}-}} \mathcal{B}_{\theta i}^+ = \{b \in \mathcal{B} \mid o_b = \theta(i), \alpha_b = t(i), i \in \mathcal{N}_\theta^{\text{POOL}-}, \theta \in \Theta\}$ stand for the blocks, containing these empty and loaded railcars, formed at their origin terminal θ at the corresponding train-departure time.

Analogously, inventories *increase* when blocks are dismantled (see Section 4.4), with associated set of inventory nodes, linked to the arrival of a train, $\mathcal{N}_\theta^{\text{POOL}+} = \cup_{\gamma \in \Gamma} \mathcal{N}_{\gamma\theta}^{\text{POOL}+} = \{i \mid t(i) = \beta_\sigma, \beta_{j(\sigma)}, j(\sigma) = 1, \ldots, |\Theta_\sigma^{\text{INT}}|, \sigma \in \Sigma\}$, $\gamma \in \Gamma, \theta \in \Theta$. We denote the set of all POOL nodes $\mathcal{N}_\theta^{\text{POOL}} := \cup_{\gamma \in \Gamma}\{\mathcal{N}_{\gamma\theta}^{\text{POOL}+} \cup \mathcal{N}_{\gamma\theta}^{\text{POOL}-}\}$ for each terminal $\theta \in \Theta$, and $\mathcal{N}^{\text{POOL}} := \cup_{\theta \in \Theta} \mathcal{N}_\theta^{\text{POOL}}$.

Pool2Pool arcs $\mathcal{A}^{\text{PP}} := \cup_{\theta \in \Theta}\{a = (i,j), j = i + 1, i, j \in \mathcal{N}_\theta^{\text{POOL}}\}$ connect successive inventory nodes at each terminal, and carry the flows of available empty railcars. To ensure the cyclic nature of the service network and schedule, each arc $a \in \mathcal{A}^{\text{PP}}$, starting from the node $i \in \mathcal{N}_\theta^{\text{POOL}}, \theta \in \Theta$, with the closest time $t(i)$ to the end of the schedule T, wraps around to the first node $i \in \mathcal{N}_\theta^{\text{POOL}}$ following the starting of the schedule, that is, to the first train-service arrival or departure moment to/from the associated terminal $\theta \in \Theta$.

Inter-layer *Car2Block* arcs $\mathcal{A}^{\text{CB}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{CO}}, j \in \mathcal{N}_\theta^{\text{BO}}, \theta \in \Theta\}$ model the car-to-block assignment decisions, linking the corresponding nodes in the car (CO) and block (BO) layers.

4.2.3 The Block Layer

Blocks are built by consolidating empty and loaded railcars brought to a *block-forming* node BO by the Pool2Block ($\mathcal{A}^{\text{PB}}$) and Car2Block ($\mathcal{A}^{\text{CB}}$) arcs, respectively. The just-built blocks are to be attached to the train leaving at the TOUT time moment of the blocking operation, together with blocks delivered earlier to the terminal to be transferred to the same train as part of their journeys (see Section 4.3). Let BT identify a *block transfer and assembly* node. The *BlockBuild* arcs $\mathcal{A}^{\text{BB}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{BO}}, j \in \mathcal{N}_\theta^{\text{BT}}), \theta \in \Theta\}$ then bring the newly built blocks to the assembly node, where they are joined by the transferred blocks brought by the inter-layer arcs $\mathcal{A}^{\text{BT}}$ (see Section 4.3). Inter-layer *BlockAttach* arcs $\mathcal{A}^{\text{BA}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{BT}}, j \in \mathcal{N}_\theta^{\text{TOUT}}), \theta \in \Theta\}$ then move the collected blocks from the assembly node to the TOUT node of the outgoing train service to which they are assigned.

4.3 Moving Between Terminals – The Train Layer

Two sets of arcs define the train-service operations. The set of *TrainMoving* arcs $\mathcal{A}_\sigma^{\text{TM}} = \{a = (i,j), i \in \mathcal{N}_\sigma^{\text{TOUT}}, j \in \mathcal{N}_\sigma^{\text{TIN}}), \sigma \in \Sigma\}$ link consecutive departure (TOUT) and arrival (TIN) nodes of train services moving between terminals. Set $\mathcal{A}_\sigma^{\text{UTM}} \subseteq \mathcal{A}_\sigma^{\text{TM}}, \sigma \in \Sigma_{\text{EXTRA}}$, contains the legs of an extra train service for which a minimum load is required in order for the service to be considered for selection. The *TrainHandling* arcs, collected in the set $\mathcal{A}_\sigma^{\text{TH}} = \{a = (i,j), i \in \mathcal{N}_\sigma^{\text{TIN}}, j \in \mathcal{N}_\sigma^{\text{TOUT}}), \sigma \in \Sigma\}$, connect consecutive arrival (TIN) and departure (TOUT) nodes of

train services, representing the time spent in the terminal to perform block pick-up and drop-off activities (and other maintenance work).

Train services are made up of, and move, blocks. Let $\mathcal{B}_a, a \in \mathcal{A}_\sigma^{\text{TH}}$, be the set of the blocks that stay on the train service $\sigma \in \Sigma$ while terminal activities take place on its Train-Handling arc a. Similarly, let $\mathcal{B}_a, a \in \mathcal{A}_\sigma^{\text{TM}}$, be the set of blocks making up train service σ on its Train-Moving arc a. The load hauled on each of these arcs $a \in \mathcal{A}^{\text{TM}} = \cup_{\sigma \in \Sigma} \mathcal{A}_\sigma^{\text{TM}}$ is limited by capacity u_a. Note that no capacity is enforced on *TrainHandling* arcs as modeling terminal operations is not part of our problem.

A block is moved by a sequence of train services, its journey corresponding to a path of Train-Moving and Train-Handling arcs, *transfers* taking place when the entire block changes trains at intermediary terminals. Inter-layer *BlockTransfer* arcs, gathered in set $\mathcal{A}^{\text{BT}} = \cup_{\theta \in \Theta}\{a = (i,j), i \in \mathcal{N}_\theta^{\text{TIN}}, j \in \mathcal{N}_\theta^{\text{BT}}), \theta \in \Theta\}$, model this activity linking the train and block layers. Let t^{TRANS} be the time required for executing a block transfer. TIN nodes in the train layer are thus linked to all BT nodes of the corresponding terminal such that $t^{\text{TRANS}} \leq t(i) - t(j), i \in \mathcal{N}_\theta^{\text{BT}}, j \in \mathcal{N}_\theta^{\text{TIN}}, \theta \in \Theta$.

4.4 Disembarking From a Train

Blocks, and the railcars and containers making them up, are taken off the last train service at the destination terminal and are dismantled. Railcars carrying containers are then at destination and, once their containers are unloaded, they join the empty railcars which were in the block (if any) in the appropriate inventory pool. The unloaded containers are delivered to their consignees. We follow these activities, completing the time-space network definition, in the reverse order of the embarking activities (see Section 4.2).

4.4.1 The Block Layer

The blocks arriving at destination are taken off the train and dismantled. Inter-layer *Block-Dismantle* arcs $\mathcal{A}^{\text{BD}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), j \in \mathcal{N}_\theta^{\text{TIN}}, j \in \mathcal{N}_\theta^{\text{BD}}), \theta \in \Theta\}$ model these activities, moving the blocks from the train TIN nodes to the corresponding block-at-destination BD node at the destination terminal.

Let $\mathcal{B}_{\theta i}^- = \{b \in \mathcal{B} \mid d_b = \theta(i), \beta_b = t(i), i \in \mathcal{N}_\theta^{\text{POOL}+}, \theta \in \Theta\}$ stand for the blocks dropped at their destination terminal θ at the corresponding train-arrival time. Their dismantlement yields empty railcars, increasing the inventory of the terminal as described next.

4.4.2 The Car Layer

Dismantled blocks yield empty and loaded railcars, of various types, which increase their availability at the terminal. Inter-layer *Empty2Pool* arcs in $\mathcal{A}^{\text{EP}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{BD}}, j \in \mathcal{N}_\theta^{\text{POOL}+}), \theta \in \Theta\}$ model the former activity, moving empty railcars from block-at-destination BD nodes, in the block layer, directly to the corresponding POOL nodes in the car layer.

Loaded railcars must first be unloaded before being added to the inventories corresponding to their types held in the terminal pool. *CarsDest* inter-layer arcs $\mathcal{A}^{\text{CD}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{BD}}, j \in \mathcal{N}_\theta^{\text{CD}}, \theta \in \Theta\}$ bring the loaded railcars from BD nodes to the railcar-unloading CD nodes in the car layer. The *Unloaded2Pool* arcs $\mathcal{A}^{\text{UP}} = \cup_{\theta \in \Theta}\{a = (i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{CD}}, j \in \mathcal{N}_\theta^{\text{POOL}+}, \theta \in \Theta\}$ then bring the now-empty railcars to the pool of the terminal.

4.4.3 The Container Layer

Demand flows exit the system through the container layer. The set of *ContainerDest* inter-layer arcs $\mathcal{A}^{\text{DOUT}} = \cup_{\theta \in \Theta}\{(i,j)|t(i) = t(j), i \in \mathcal{N}_\theta^{\text{CD}}, j \in \mathcal{N}_\theta^{\text{DOUT}}, \theta \in \Theta\}$ groups the arcs bringing unloaded containers from the nodes CD in the car layer, to the $i \in \mathcal{N}^{\text{DOUT}}$ destination nodes

(DOUT) in the container layer. Being in the yard of the destination terminal marks the end of a demand's freight journey. The final delivery to customers is then modeled by the arc linking the DOUT node to the container-sink node CSINK collecting all demand flows sent to the terminal.

The time-space network is completed by the artificial arcs capturing the demand volumes not assigned to services in Σ, due to transport capacity issues. An artificial arc is defined for demand $k \in \mathcal{K}$ from its last feasible DIN node, i.e., the last moment demand k may leave its origin terminal and still get at its destination on time, and the container-sink node CSINK of its destination.

5 The SSND-RM Formulation

The SSND-RM formulation for the intermodal rail blocking and railcar management problem includes the following design and flow decision variables:

$s_\sigma = 1$, if train service $\sigma \in \Sigma_{\text{EXTRA}}$ is selected, 0 otherwise;

$y_b = 1$, if block $b \in \mathcal{B}$ is selected, 0 otherwise;

z_{bk}, number of containers of demand $k \in \mathcal{K}$ on block $b \in \mathcal{B}_k$;

z_k, number of containers of demand $k \in \mathcal{K}$ on the artificial arc associated with this demand;

x_b^γ, number of loaded railcars of type $\gamma \in \Gamma$ assigned to block b;

w_b^γ, number of empty railcars of type $\gamma \in \Gamma$ assigned to block b;

w_θ^γ, number of empty railcars of type $\gamma \in \Gamma$ allocated to terminal $\theta \in \Theta$;

$w_{\theta i}^\gamma$, number of empty railcars of type $\gamma \in \Gamma$ in pool node $i \in \mathcal{N}_\theta^{\text{POOL}}$ of terminal $\theta \in \Theta$;

$\nu_{b\pi}^{\tau,\tau'}$ number of platforms of type $\pi \in \Pi$ loaded with two containers of types τ and $\tau' \in \mathcal{T}$ (no container type τ_{53} on bottom slot of π_{40} platforms);

$\nu_{b\pi}^\tau$, number of platforms of type $\pi \in \Pi$ loaded with a single container of type $\tau \in \mathcal{T}$ (no container type τ_{53} on π_{40} platforms).

The ILP of our problem is:

$$\text{Minimize} \sum_{\sigma \in \Sigma_{\text{EXTRA}}} f_\sigma s_\sigma + \sum_{b \in \mathcal{B}} f_b y_b + \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} c_{bk} z_{bk} + \sum_{k \in \mathcal{K}} c_k z_k + \sum_{b \in \mathcal{B}} \sum_{\gamma \in \Gamma} c_{b\gamma} w_b^\gamma + \sum_{\gamma \in \Gamma} \sum_{\theta \in \Theta} c_{\gamma\theta} w_\theta^\gamma \quad (1)$$

Subject to:

$$\sum_{b \in \mathcal{B}_k} z_{bk} + z_k = \upsilon_k, \qquad\qquad\qquad k \in \mathcal{K}, \qquad (2)$$

$$z_{bk} \leq \upsilon_k y_b, \qquad\qquad\qquad b \in \mathcal{B}_k, \ k \in \mathcal{K}, \ (3)$$

$$\sum_{k \in \mathcal{K}_b | \tau_k = \tau} z_{bk} = \sum_{\pi \in \Pi} \left(\nu_{b\pi}^\tau + 2\nu_{b\pi}^{\tau,\tau} + \sum_{\tau' \in \mathcal{T} | \tau' \neq \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad \tau \in \mathcal{T}, b \in \mathcal{B}, \quad (4)$$

$$\sum_{\gamma \in \Gamma} \eta_\pi^\gamma x_b^\gamma \geq \sum_{\tau \in \mathcal{T}} \left(\nu_{b\pi}^\tau + \nu_{b\pi}^{\tau,\tau} + \frac{1}{2} \sum_{\tau' \in \mathcal{T} | \tau' \neq \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad \pi \in \Pi, b \in \mathcal{B}, \quad (5)$$

$$\sum_{\gamma \in \Gamma_\pi} x_b^\gamma \leq \sum_{\tau \in \mathcal{T}} \left(\nu_{b\pi}^\tau + \nu_{b\pi}^{\tau,\tau} + \frac{1}{2} \sum_{\tau' \in \mathcal{T} | \tau' \neq \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad \pi \in \Pi, b \in \mathcal{B}, \quad (6)$$

$$\sum_{\gamma \in \Gamma} \left\lceil \frac{\eta_{\pi_{40}}^\gamma}{2} \right\rceil x_b^\gamma \geq \nu_{b\pi_{40}}^{\tau_{40}, \tau_{53}}, \qquad\qquad b \in \mathcal{B}, \qquad (7)$$

$$\sum_{\gamma \in \Gamma} \lambda_\gamma (x_b^\gamma + w_b^\gamma) \leq u_b y_b, \qquad\qquad b \in \mathcal{B}, \tag{8}$$

$$w_{\theta i-1}^\gamma + \sum_{b \in \mathcal{B}_{\theta i}^-} (w_b^\gamma + x_b^\gamma) = w_{\theta i}^\gamma + \sum_{b \in \mathcal{B}_{\theta i}^+} (w_b^\gamma + x_b^\gamma), \qquad\qquad i \in \mathcal{N}_\theta^{\text{POOL}}, \gamma \in \Gamma, \theta \in \Theta, \tag{9}$$

$$w_\theta^\gamma = w_{\theta n}^\gamma + \sum_{b \in \mathcal{B}_\theta^+ | \alpha_b > \beta_b} (w_b^\gamma + x_b^\gamma), \qquad\qquad \gamma \in \Gamma, \theta \in \Theta, \tag{10}$$

$$\sum_{\theta \in \Theta} w_\theta^\gamma \leq H^\gamma, \qquad\qquad \gamma \in \Gamma, \tag{11}$$

$$\sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma (x_b^\gamma + w_b^\gamma) \leq u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{INIT}}, \tag{12}$$

$$\sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma (x_b^\gamma + w_b^\gamma) \leq s_\sigma u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{EXTRA}}, \tag{13}$$

$$\frac{1}{u_a} \sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma (x_b^\gamma + w_b^\gamma) \geq U_\sigma s_\sigma, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{UTM}}, \sigma \in \Sigma_{\text{EXTRA}}, \tag{14}$$

$$y_b \in \{0, 1\}, \qquad\qquad b \in \mathcal{B},$$

$$s_\sigma \in \{0, 1\}, \qquad\qquad \sigma \in \Sigma_{\text{EXTRA}},$$

$$x_b^\gamma, w_\theta^\gamma, w_b^\gamma, w_{\theta i}^\gamma, z_{bk}, z_k, \nu_{b\pi}^{\tau \tau'}, \nu_{b\pi}^\tau \in \mathbb{N}, \qquad\qquad k \in \mathcal{K}, b \in \mathcal{B}, \tau \in \mathcal{T}, \tau' \in \mathcal{T}, \pi \in \Pi,$$

$$i \in \mathcal{N}, j \in \mathcal{N}, \gamma \in \Gamma, \theta \in \Theta.$$

The objective function (1) minimizes the total cost of selecting extra trains (when relevant), selecting and building blocks, routing demand flows on blocks and artificial arcs, moving empty railcars, and allocating railcars to terminals initially.

Constraints (2) ensure that all the volume of each demand is transported either on blocks (train services) or on the artificial arcs. The linking constraints (3) enforce containers to be moved on selected blocks only, limiting the volume of any demand moving on a particular block to the total volume of that demand. Constraints (4) define the total container-flow (of a given type) moved by a particular block as the number of platforms (loaded with containers of that type) on the railcars within the block. Notice that, jointly with the railcar-to-block linking constraints (8), Constraints (4) enforce the flow of containers to be loaded on selected blocks only.

Recall that, a railcar is considered loaded as soon as at least one platform carries a container, even if the other platforms are empty. Consequently, Constraints (5) ensure that, the number of platforms, of a given type, making up the loaded railcars of a block is at least equal to the number of platforms of that type used on the block. Similarly, Constraints (6) ensure that the number of loaded railcars composed of platforms of a certain type on a block is not larger than the number of platforms of that type used on the block. Constraints (7) address the particular restrictions regarding the number of 40-on-bottom-53-on-top configurations on 40-ft platforms imposed by the physical rules of loading (see Section 2 for more details).

Constraints (8) link the utilization of blocks by loaded and empty railcars to the selection of the block and its length. Constraints (9) enforce the conservation of empty-railcar flows, of all types, at the pool nodes of the terminals. Constraints (10) count the number of railcars of each type needed at every terminal (initial allocation). The limits on the initial allocations are imposed through Constraints (11). Note that one may impose an initial allocation H_θ^γ to each terminal or let the tactical planning model decide on the best allocation, with or without a H_θ^γ limit.

Constraints (12) and (13) enforce the capacities of the regular and extra train services, respectively, on each of their legs. The latter also link the utilization of extra services to their selection. Finally, constraints (14) enforce the condition that an extra train may be added to

the schedule only if it is to be loaded at more than $U_\sigma\%$ of its capacity on a set of its legs.

6 Solution Approach

Before presenting our solution approach, we discuss key distinctive features of the SSND-RM formulation and outline the practical considerations that guided our work.

The SSND-RM formulation is based on a time-space network whose discretization is governed by the train schedule. Except for the selection of extra trains, the schedule is fixed and given. The set of extra train selection variables is particularly challenging because it introduces an additional layer of design decisions that constrain block selection decisions. The feasible set of blocks $\mathcal{B}$ is large, but is restricted by the train schedule. In addition, the demand and railcar flow variables are integers and are strongly linked through the platform loading constraints.

Given the tactical nature of the problem, an overnight computation time budget is considered acceptable. For deployment of a decision-support system based on the model, reliability and code maintenance are important considerations. For these reasons, we give priority to solution approaches using general-purpose MILP solvers as a black box, and algorithms with solution quality guarantees.

For our application, preliminary experiments using a general-purpose MILP solver based on an a priori enumeration of the set of block paths $\mathcal{B}$ revealed large optimality gaps even after 12 hours of computation. However, it is well known that relatively simple heuristics combined with general-purpose MILP solvers can yield substantial computation speed-ups (e.g., Fischetti and Monaci 2017, Joncour et al. 2023). Inspired by such ideas, we propose a construction heuristic close to the relax-and-fix procedure described in Section 3.6.1 in Pochet and Wolsey (2006). We use the resulting feasible solution as a warm start for a general-purpose solver. Thus, the overall solution approach we propose remains exact.

We partition the decision variables into three disjoint sets. The first set contains the most challenging variables, in our case, the binary extra-train variables. The second set orders groups of decision variables in decreasing order of perceived impact on the structure of the solution, $\mathcal{R} := (\{y_b, x_b^\gamma, b \in \mathcal{B}, \gamma \in \Gamma\}, \{w_b^\gamma, w_\theta^\gamma, w_{\theta i}^\gamma; b \in \mathcal{B}, \theta \in \Theta, i \in \mathcal{N}_\theta^{\text{POOL}}, \gamma \in \Gamma\})$. The third set groups container and platform assignment to block decision variables, $\mathcal{S} := \{z_{bk}, z_k, \nu_{b\pi}^\tau, \nu_{b\pi}^{\tau,\tau'}; b \in \mathcal{B}, k \in \mathcal{K}, \pi \in \Pi, \tau, \tau' \in \mathcal{T}\}$.

Algorithm 1 Compute warm-start solution

Require: Extra train variables Σ_{EXTRA}, ordered sequence of sets of decision variables $\mathcal{R}$, set $\mathcal{S}$, and precision $\varepsilon > 0$,
  1: *argrelax* ← Solve continuous relaxation of SSND-RM ▷ Values of relaxed decision variables
  2: *restrictedModel* ← Fix to zero all variables in Σ_{EXTRA} whose value in *argrelax* is less than ε; Impose integrality on the other variables in Σ_{EXTRA}
  3: *argrestrict* ← Solve *restrictedModel*
  4: *restrictedModel* ← Fix all variables in Σ_{EXTRA} to their values in *argrestrict*
  5: **for all** sets of decision variables $\mathcal{R}[i]$ **do**
  6:   *restrictedModel* ← Fix to zero all variables in $\mathcal{R}[i]$ whose value in *argrestrict* is less than ε; Impose integrality on the other variables in $\mathcal{R}[i]$
  7:   **if** $i = last(\mathcal{R})$ **then**
  8:     *restrictedModel* ← Impose integrality on all variables in $\mathcal{S}$
  9:   **end if**
  10:   *argrestrict* ← Solve *restrictedModel*
  11: **end for**
  12: **return** *argrestrict*

Algorithm 1 provides the details. The first five steps are dedicated to the extra train

variables. Steps 6-10 concern the ordered sequence of decision variables $\mathcal{R}$ where we apply an iterative round-and-fix procedure that only rounds down and fix the variables whose values are close to zero. In the last iteration, we impose integrality on the decision variables $\mathcal{S}$ for which the rounding-and-fixing scheme could cause feasibility issues (Steps 8-10). Finally, we note that the precision ε should be close to zero; we use $1e^{-5}$.

7   Numerical Results

In this section, we report results using the large network of our industrial partner. Their network spans a large share of North America (see illustration in Figure 6 in the Appendix). To increase variability and protect confidentiality, we simulate instances based on a real historical train schedule and demand data. We describe the experimental setup in Section 7.1, followed by an analysis of the results in Section 7.2. We conclude by providing managerial insights based on the results.

7.1   Experimental Setup

In the following we provide details around our experimental setup.

Objective function. The objective function is a specific case of (1). The associated parameter values have been carefully tuned in collaboration with subject matter experts of our industrial partner to ensure that the model generates realistic solutions. We refer to Appendix B for details.

Extra Trains. To evaluate the impact of extra trains we consider a setting where any of the existing train services in the schedule can be duplicated. Moreover, we fix the threshold U_σ of minimum capacity to fill to select an extra train in (14) to 50% for all $\sigma \in \Sigma_{\text{EXTRA}}$.

Block Generation. We restrict the set of potential blocks $\mathcal{B}$ to those that are considered feasible in practice. While generating the paths in a depth-first search, we impose practical constraints given by the train schedule. We note that for all instances it takes 51 seconds (the variance is close to zero) to generate all the blocks (around 47,000). Since this time is constant, it is not included in the computing times reported in Section 7.2.

Demand. To keep real data confidential, we simulate five realistic demand instances based on actual values. There are 165 OD pairs in our network. To maintain the realism with respect to the scheduled train services (network capacity) that are fixed, we vary the characteristics of the demand while keeping overall demand volumes in the network intact. For this purpose, we randomly generate the proportion of 40-ft versus 53-ft containers for each OD pair. In Figure 4, we show that this leads to instances that have different characteristics. We display two box plots, each having one bar per instance (x-axis). The y-axis in the left-hand plot displays demand asymmetry. Based on total weekly volume for each pair of locations, it is computed as the percentage difference in volume (i.e., relative difference in demand between locations A to B, compared to B to A). The plot on the right-hand side, displays the percentage of 40-ft containers for each OD pair. First, we note that few OD pairs are perfectly symmetrical (zero percentage difference) and most are highly asymmetrical. Moreover, the percentage of 40-ft containers vary across ODs and across instances.

Railcar fleet. To assess the impact of different railcar types, we consider seven different railcar fleet scenarios:
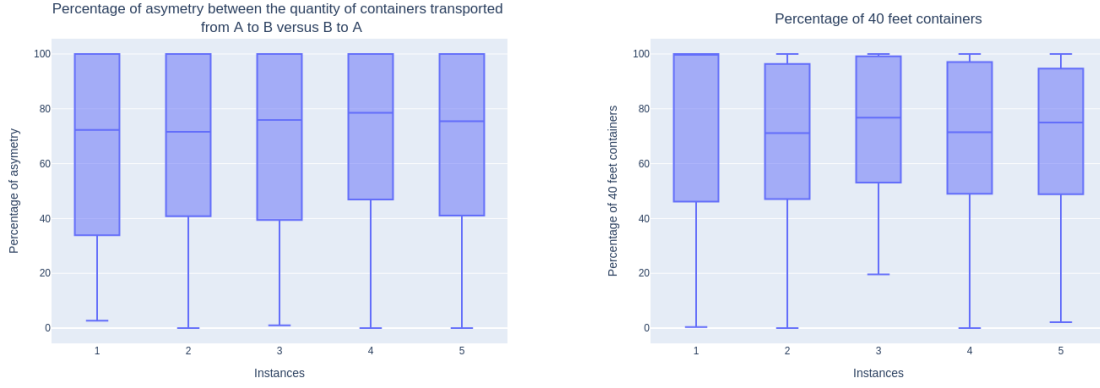
Figure 4: Illustration of how the demands are unbalanced and their types

1. 1-platform 53-ft (1x53)

2. 1-platform 40-ft, and 1-platform 53-ft (1x40,1x53)

3. 3-platforms 53-ft (3x53)

4. 3-platforms 40-ft, and 3-platforms 53-ft (3x40,3x53)

5. 5-platforms 53-ft (5x53)

6. 5-platforms 40-ft, and 5-platforms 53-ft (5x40,5x53)

7. All six types (All)

These scenarios cover different extreme cases: the most flexible cars in terms of loading but least efficient in terms of capacity usage (1x53), the most efficient in terms of capacity usage but most restricted in terms of loading (5x40), and the most flexible scenario from a network planning perspective (All).

Baseline Models.   We assess the value of the proposed model by comparing it to two baselines. First, a model that does not manage the railcar fleet hence assuming an infinite fleet, but that takes into account loading restrictions (called *Unrestricted Fleet Model*). Second, a model that, in addition to ignoring the railcar fleet management, does not take into account loading constraints (called *Unrestricted Loading Model*). The former model resembles the one proposed by Morganti et al. (2020). The latter model simply uses a formula that transforms a number of containers into capacity usage (train length) by using railcar lengths and assuming double stacking is always possible. We detail the integer linear formulations of the baseline models in Appendices C and D.

Solution Approaches and Hardware.   All experiments were performed using an Intel Core i9-10980XE 3.00 GHz processor with 128 GB of RAM. We used ILOG CPLEX 22.1.1, restricting the solver to a single thread, with a time limit of 12 hours and a stopping criteria of 2.5% optimality gap. We compare the total solving times using CPLEX with our warm start approach to the time it takes to solve the SSND-RM formulation directly (without warm start).

Performance metrics.   We use standard performance metrics: computing time in CPU seconds, optimality gap, root gap (optimality gap at the first found feasible solution) and number of instances that reach the time limit. We provide interpretability by computing overall capacity

19

utilization, percentage of railcar slot utilization, percentage of unsatisfied demand, fleet composition statistics, number of extra trains (if applicable), number of platforms, and number of railcars. These metrics are described in more detail next to the results. We solve five instances (demand scenarios) for each railcar scenario and version of the model/algorithm. Unless stated otherwise, we report an average over the five instances for each performance metric.

7.2 Results

The results are structured around the key takeaways. Namely, the warm-start solutions are of high quality, which leads to a significant speed up (Section 7.2.1), the baseline models are faster to solve but importantly underestimate the required capacity (Section 7.2.2), and railcar fleet management favours five-platform and 40-ft railcars whenever possible depending on the mix of container types to transport (Section 7.2.3).

7.2.1 High-quality Warm-start Solutions Yield Significant Computational Speed-ups

To assess computational performance, we solve the model with and without the extra train selection variables. Table 3 reports the computing times for the different railcar fleet scenarios.

Focusing on the left-hand side of the table, we note that it is relatively fast to compute the warm-start solutions. As expected, instances with extra trains are considerably more time consuming to solve than those without extra trains. The resulting warm-start solutions for the former instances are of high quality, which leads to a total average computing time within 3.25 hours. Note that the performance metrics consider instances that reached the time limit and we report the number of such instances in column (#Timed-Out). Hence, the average optimality gap may be higher than the 2.5% stopping criterion without reaching the 12 hour time limit on average (excluding such instances results in a total average time of 1,627 seconds for scenario 3x53, 998 seconds for scenario 3x40, 3x53, and 731 seconds for scenario 5x40, 5x53). Interestingly, 5x40, 5x53 (and not All) is the railcar scenario with the highest number of instances reaching the time limit. This may be due to symmetry issues.

In the case of no extra trains, the quality of the warm-start solutions is already below the 2.5% optimality gap stopping criteria. We note that the averages are inflated by a few large values. The median time to compute a warm-start solution is 37 seconds for the model without extra trains, and 392 seconds for the model with extra trains.

Now, turning to the right-hand side of the table, we note that the warm start outperforms the general-purpose solver alone across all railcar scenarios, with and without extra trains. A total of 39 instances time-out, 20 in the extra train case. All of them concern a railcar scenario with a mix of 40-ft and 53-ft platform railcars.

7.2.2 Baseline Models Underestimate the Required Capacity

As described in Section 7.1, we assess the value of the SSND-RM formulation by comparing solutions with those obtained from two simpler baseline models. We focus on assessing how they differ in the estimation of required capacity, hence we limit the comparison to instances without extra trains. None of those instances reached the time limit. The results are presented in Table 4 and we note that the computing time for SSND-RM is the same as the corresponding results in Table 3. The rightmost sets of columns report results for the unrestricted fleet and unrestricted loading models, respectively.

In addition to computing time, we report the percentage of unsatisfied demand (i.e., demand volumes assigned to artificial arcs compared to total demand volumes in the network), and the relative difference in network capacity usage for each model compared to the SSND-RM. We measure capacity in train length distance (train service capacity multiplied by the distance

Table 3: Average computing times with and without warm start, for instances with and without extra trains

Scenario	Extra Trains - With Warm start					Extra Trains - Without Warm start			
	Warm start Time (s)	Total Time (s)	#Timed-Out	Root Gap	Opt Gap	Time (s)	#Timed-Out	Root Gap	Opt Gap
1x53	672	1,017	0	4.18%	2.40%	3,691	0	87.01%	2.42%
1x40,1x53	3,435	13,313	0	3.44%	2.45%	43,257	5	97.36%	56.34%
3x53	78	9,916	1	6.32%	2.49%	3,757	0	97.29%	2.42%
3x40,3x53	4,707	17,820	2	3.24%	2.61%	43,257	5	97.51%	48.42%
5x53	24	31	0	1.94%	1.94%	565	0	97.46%	1.94%
5x40,5x53	368	26,125	3	3.28%	2.66%	43,257	5	97.57%	82.84%
All	1,393	13,522	0	2.74%	2.38%	43,261	5	97.58%	97.58%

Scenario	No Extra Trains - With Warm start					No Extra Trains - Without Warm start			
	Warm start Time (s)	Total Time (s)	#Timed-Out	Root Gap	Opt Gap	Time (s)	#Timed-Out	Root Gap	Opt Gap
1x53	16	19	0	0.70%	0.70%	308	0	92.71%	1.32%
1x40,1x53	38	68	0	2.26%	2.26%	3,691	4	95.16%	2.24%
3x53	15	19	0	0.79%	0.79%	427	0	95.24%	0.82%
3x40,3x53	1,317	1,479	0	2.07%	2.00%	43,257	5	97.18%	25.88%
5x53	15	22	0	1.45%	1.44%	232	0	96.88%	1.97%
5x40,5x53	159	184	0	1.50%	1.51%	43,257	5	97.56%	82.84%
All	7,258	7,453	0	1.78%	1.78%	43,261	5	97.26%	97.24%

covered by each train leg). The capacity usage is then calculated as the train service used capacity multiplied by distance, divided by total network capacity in the same unit. The figures in the table correspond to the percentage difference compared to the capacity usage of the SSND-RM (which therefore has a usage difference of 0%).

As expected, the SSND-RM model is harder to solve than the baseline models (its average computing times are considerably larger), but it accurately models the repositioning of empty railcars, as well as the loading restrictions specific to each railcar type. Ignoring the repositioning of empty railcars (as in the Unrestricted Fleet model), means that the train capacity required for this purpose is considered available to transport demand. This represents an average of 2-6% of the network capacity. That is, using the unrestricted fleet model leads to an underestimation of the capacity needed by an average of 2-6%, depending on the railcar scenario.

Ignoring intermodal loading constraints, in addition to railcar fleet management, has a larger impact, as can be seen from the solutions to the Unrestricted Loading model. The average underestimation of capacity usage ranges from 17% to 27%, depending on the railcar scenario.

We note that the reduction in unsatisfied demand aligns with the underestimation of required capacity: the larger the underestimation, the more capacity appears available, enabling more demand to be fulfilled. Finally, we note that the usage difference in the *Unrestricted Loading Model* varies as a result of changes in the capacity usage of the SSND-RM model, rather than fluctuations in its own usage. Indeed, as expected, it remains constant across railcar scenarios, as evidenced by the constant computing time and percentage of unsatisfied demand.

Table 4: Comparison to model baselines on instances without extra trains (averages over five instances)

Scenario	SSND-RM (with warm start)			Unrestricted Fleet Model			Unrestricted Loading Model		
	Time (s)	Uns. Dem.	Usage Diff	Time (s)	Uns. Dem.	Usage Diff	Time (s)	Uns. Dem.	Usage Diff
1x53	19.4	5.20%	0%	3.7	5.12%	5%	1.8	1.58%	27%
1x40,1x53	68.3	1.64%	0%	5.7	1.64%	6%	1.8	1.58%	25%
3x53	18.3	2.56%	0%	4.1	2.55%	5%	1.8	1.58%	25%
3x40,3x53	1,478.6	0.63%	0%	27.1	0.62%	4%	1.8	1.58%	20%
5x53	21.8	0.80%	0%	4.1	0.80%	4%	1.8	1.58%	20%
5x40,5x53	184.4	0.55%	0%	8.0	0.55%	4%	1.8	1.58%	17%
All	7,453.4	0.56%	0%	11.9	0.56%	2%	1.8	1.58%	18%

7.2.3 Railcar Fleet Management Favors Five-platform and 40-ft Railcars Whenever the Mix of Container Types Allows

Table 5 reports more detailed results on the selection of extra trains and the composition of the railcar fleet. More precisely, we report the average number of extra trains, the number of platforms and railcars, their slot utilization (the number of slots used compared to the total number of slots available on railcars), and the composition of railcar types in the selected fleet. Since the number of platforms varies across railcar scenarios, we separate the percentage of 53-ft platforms and the percentage of one-, three-, and five-platform railcars.

At our tactical level, the plan is repeated every week. Unsatisfied demand therefore accumulates unless extra resources are used. As expected, extra trains reduce the percentage of unsatisfied demand. Since we impose a minimum capacity usage for such trains, it nevertheless remains non-zero. We note that the number of extra trains varies depending on the railcar scenario.

Recall that railcar fleet management needs to strike the right balance between loading flexibility and capacity usage. The 53-ft one-platform railcars are the most flexible in terms of loading, while they make the least efficient use of capacity. On the other side of the spectrum, the 40-ft five-platform railcars are the most restrictive in terms of loading, but they make the most efficient use of capacity (measured by the number of containers per train foot). Relatively few extra trains are required when the model can choose among the most capacity-efficient railcars (two extra trains on average, compared to 14.2 on average when restricted to using 53-ft one-platform railcars).

Next, turning to the average slot utilization (%slotUsed in Table 5), we note that it is relatively stable across the different railcar scenarios. As expected, it decreases as the percentage of five-platform railcars increases in the fleet. When the model can select railcars of any type, on average, 70% of the platforms are on five-platform cars (68% in the case of extra trains). We note that whenever the model can select between 40-ft and 53-ft railcars, with only one exception, a majority of the platforms are 40-ft (the average percentage of 53-ft platforms varies between 24% and 45%, with one exception at 55%). The exception concerns the extra train instances, because the demand satisfied by the extra trains is mostly 53-ft containers.

Table 5: Results on railcar fleet composition (averages over five instances)

Scenario	Extra Trains									
	Time (s)	Uns. Dem.	#extra	nbPlatforms	nbRailcars	%slotUsed	%53Ftplat.	%1plat.	%3plat.	%5plat.
1x53	1,017	0.43%	14.2	14,752	14,752	91%	100%	100%	0%	0%
1x40,1x53	13,313	0.32%	5	14,726	14,726	90%	27%	100%	0%	0%
3x53	9,916	0.34%	9	15,043	5,014	84%	100%	0%	100%	0%
3x40,3x53	17,820	0.26%	2.2	14,860	4,953	83%	36%	0%	100%	0%
5x53	31	0.27%	3	15,013	3,003	80%	100%	0%	0%	100%
5x40,5x53	26,125	0.23%	2	15,275	3,055	78%	42%	0%	0%	100%
All	13,522	0.21%	2	15,080	3,993	79%	55%	29%	3%	68%

Scenario	No Extra Trains									
	Time (s)	Uns. Dem.	#extra	nbPlatforms	nbRailcars	%slotUsed	%53Ftplat.	%1plat.	%3plat.	%5plat.
1x53	19	5.20%	0	13,229	13,229	92%	100%	100%	0%	0%
1x40,1x53	68	1.64%	0	14,607	14,607	90%	24%	100%	0%	0%
3x53	18	2.56%	0	13,955	4,652	85%	100%	0%	100%	0%
3x40,3x53	1,479	0.63%	0	14,561	4,854	83%	33%	0%	100%	0%
5x53	22	0.80%	0	14,635	2,927	80%	100%	0%	0%	100%
5x40,5x53	184	0.55%	0	14,931	2,998	79%	39%	0%	0%	100%
All	7,453	0.56%	0	14,758	3,664	81%	45%	19%	11%	70%

Through the loading constraints, the composition of the railcar fleet is determined by the mix of container sizes in the demand volumes. To better analyze why, we provide a more granular view of the solutions in Figure 5. To avoid instances that reached the time limit, we use only those without extra trains. The figure displays four scatter plots, one for each railcar scenario that allows selection between 40-ft and 53-ft railcars. Each dot represents a block in the

five instances over which we averaged the statistics in the previous tables. The x-axis displays the ratio of 40-ft platforms in a given block, and the y-axis shows the ratio of 40-ft containers in the demand assigned to that block. The size and the shade of gray of each dot vary: the size increases with increasing density of dots in each plot. Dots of different shades may overlap, each dot's size will vary according to the density of its respective shade. The legend for the shades of gray of the dot is shown on the right and represents the slot utilization, ranging from relatively low (light shade) to high (dark shade).
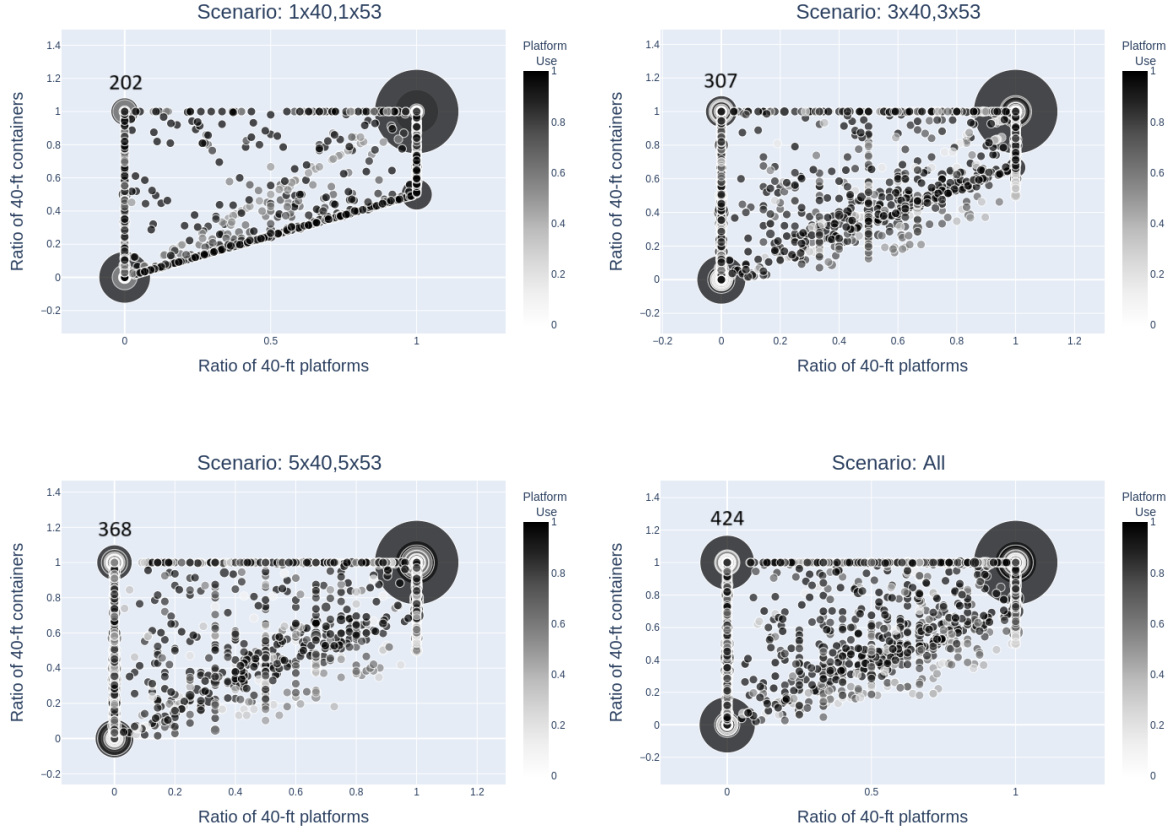
Figure 5: Fleet composition for railcar scenarios with 40-ft and 53-ft platforms without extra trains

Each plot has a large dot at $(1, 1)$, meaning that in each railcar scenario many blocks are made of 40-ft platforms that exclusively transport 40-ft containers. Similarly, each plot has a large dot at $(0, 0)$, meaning that in each scenario many blocks are made of 53-ft platforms that transport exclusively 53-ft containers. The plots differ at $(0, 1)$, which represents blocks made entirely of 53-ft platforms transporting exclusively 40-ft containers. This inefficient use of capacity is explained by unbalanced demand and the need to reposition railcars. The figure displayed above each of the dots at $(0, 1)$ represents the total number of blocks, without taking into account the shade. Observe that, when the model is free to select any railcar type (bottom right), it sacrifices loading capacity efficiency (larger figure) in exchange for better use of the available platforms (larger black $(0, 1)$ dot). Comparing the different plots at this point, we note that the total slot utilization improves when the model can select any railcar type.

The upper-left plot in Figure 5 displays a linear function between $(0, 1)$ and $(1, 0.5)$ that can be explained by the fact that the minimum ratio of 40-ft containers is half the ratio of the 40-ft platforms. Indeed, if only railcars with 40-ft platforms are available, at least half of the containers must be 40-ft, since 53-ft containers can only be loaded on top of 40-ft containers on

40-ft platforms. For all other railcar scenarios, some points are under this linear function, as some platforms can be empty.

7.3 Managerial Insights

Based on the results in the previous section, several general managerial insights emerge.

It is crucial to consider loading constraints for intermodal rail network planning. Ignoring such constraints can lead to a significant underestimation of the required capacity. In practice, this may lead to increased operational costs by requiring the addition of costly capacity on an ad hoc basis, and it may negatively impact customer service through delays.

The results also underline the importance of managing the railcar fleet. That is, accounting for the repositioning of empty railcars is important to adequately estimate network capacity usage. There is another side benefit: empty railcar repositioning represents unproductive moves of equipment. An accurate representation of such capacity can provide valuable information for adjusting the pricing strategy to better leverage this capacity and enhance profitability.

An adequate railcar fleet composition is important for the efficient use of network capacity. If demand in the network at a block level is characterized by a majority of 40-ft containers, then 40-ft railcars should be favored. For networks characterized by a high degree of variation in the mix of container types, the fleet should consist primarily of 53-ft platform railcars to ensure greater flexibility.

From a capacity usage point of view, there is a clear positive effect of using multi-platform railcars. Slot utilization per block is often used as a performance metric, especially to measure terminal productivity (loading at the origin of a block). Our results show that this metric alone may not be adequate. As an extreme example, a high average slot utilization can be achieved at a terminal level using one-platform 53-ft railcars, which may nevertheless result in poor overall network capacity utilization.

8 Conclusion

We introduced a new problem for tactical planning of intermodal railway transport. The problem considers several interrelated decision-making problems capturing three consolidation processes and the management of a heterogeneous railcar fleet. We proposed a relatively simple solution approach that consists of generating warm-start solutions with a construction heuristic inspired by a relax-and-fix procedure. From a practical point of view, it has the advantage of leveraging a general-purpose solver as a black box.

Results based on realistic instances from our industrial partner, the Canadian National Railway Company, showed that we can solve large-scale instances down to an optimality gap of 2.5% in a reasonable amount of time. We analyzed computational performance and the railcar fleet management solutions. Moreover, a comparison with solutions obtained by solving simpler baseline models showed that this new problem has high value. Based on the results, we drew managerial insights that are valuable to any railway company transporting intermodal containers on double-stack railcars.

This work opens up several avenues for future research. We solved a deterministic formulation of the problem, but at the tactical planning level there is typically uncertainty in the demand to be transported. A contextual stochastic formulation (Sadana et al. 2025) could lead to more resilient solutions. However, it poses significant computational challenges, especially since the demand distributions are expected to be decision-dependent (Frejinger and Hewitt 2025). Another important direction for future research is to tackle the corresponding operational network planning problem. In this case, there is less uncertainty about demand, but it is necessary to adapt the tactical plan to account for unforeseen events, such as equipment failures. In turn, this would require managing the railcar fleet at an even more granular level.

Acknowledgments

References

R.K. Ahuja, C.B. Cunha, and G. Şahin. Network Models in Railroad Planning and Scheduling. In *Tutorial in Operations Research INFORMS 2005*, pages 54–101. INFORMS, 2005a. Published online: 14 Oct 2014; http://dx.doi.org/10.1287/educ.1053.0013.

R.K. Ahuja, J. Liu, J.B. Orlin, and L.A. Sharma, D.and Shughart. Solving Real-Life Locomotive-Scheduling Problems. *Transportation Science*, 39(4):503–517, 2005b.

R.K. Ahuja, K.C. Jha, and J. Liu. Solving Real-Life Railroad Blocking Problems. *Interfaces*, 37:404–419, 2007.

J. Andersen, T.G. Crainic, and M. Christiansen. Service Network Design with Management and Coordination of Multiple Fleets. *European Journal of Operational Research*, 193(2):377–389, 2009a.

J. Andersen, T.G. Crainic, and M. Christiansen. Service Network Design with Asset Management: Formulations and Comparative Analyzes. *Transportation Research Part C: Emerging Technologies*, 17(2):197–207, 2009b.

J. Andersen, M. Christiansen, T.G. Crainic, and R. Grønhaug. Branch-and-Price for Service Network Design with Asset Management Constraints. *Transportation Science*, 46(1):33–49, 2011.

A.A. Assad. Modelling of Rail Networks: Toward a Routing/Makeup Model. *Transportation Research Part B: Methodological*, 14:101–114, 1980a.

A.A. Assad. Models for Rail Transportation. *Transportation Research Part A: Policy and Practice*, 14: 205–220, 1980b.

A. Balakrishnan, A. Kuo, and X. Si. Real-Time Decision Support for Crew Assignment in Double-Ended Districts for U.S. Freight Railways. *Transportation Science*, 50(4):1139–1393, 2016.

T. Bektaş and T.G. Crainic. A Brief Overview of Intermodal Transportation. In Taylor, G.D., editor, *Logistics Engineering Handbook*, chapter 28, pages 1–16. Taylor and Francis Group, Boca Raton, FL, USA, 2008.

LD. Bodin, B.L. Golden, A.D. Schuster, and W. Romig. A Model for the Blocking of Trains. *Transportation Research Part B: Methodological*, 14(1):115–120, 1980.

A.M. Bomberault and W.W. White. Scheduling Empty Box Cars. Technical report, IBM New York Scientific Center, Hawthorne, N.Y., 1966.

B. Bouzaïene-Ayari, C. Cheng, S. Das, R. Fiorillo, and W.B. Powell. From Single Commodity to Multiattribute Models for Locomotive Optimization: A Comparison of Optimal Integer Programming and Approximate Dynamic Programming. *Transportation Science*, 50(2):366–389, 2016.

Canadian National Railway Company. Intermodal Network. `https://www.cn.ca/-/media/images/our-services/images/intermodal-map-download-en.jpg`, 2025. Accessed: 2025-08-13.

M. Chouman and T.G. Crainic. Cutting-Plane Matheuristic for Service Network Design with Design-Balanced Requirements. *Transportation Science*, 49(1):99–113, 2015.

M. Chouman and T.G. Crainic. Freight Railroad Service Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 13. Springer, Boston, 2021. 383–426.

J.F. Cordeau, P. Toth, and D. Vigo. A Survey of Optimization Models for Train Routing and Scheduling. *Transportation Science*, 32(4):380–404, 1998.

T.G. Crainic. Rail Tactical Planning: Issues, Models and Tools. In L. Bianco and A. La Bella, editors, *Freight Transport Planning and Logistics*, pages 463–509. Springer-Verlag, Berlin, 1988.

T.G. Crainic. Network Design in Freight Transportation. *European Journal of Operational Research*, 122(2):272–288, 2000.

T.G. Crainic. Long-Haul Freight Transportation. In Hall, R.W., editor, *Handbook of Transportation Science*, pages 451–516. Kluwer Academic Publishers, Norwell, MA, second edition, 2003.

T.G. Crainic. Service Design Models for Rail Intermodal Transportation. In Bertazzi, L., Speranza, M.G., and van Nunen, J.A.E.E., editors, *Lecture Notes in Economics and Mathematical Systems*, volume 619, pages 53–67. Springer-Verlag, Berlin Heidelberg, 2009.

T.G. Crainic. Multi-layer Network Design for Consolidation-based Transportation Planning. In T.G. Crainic, M. Gendreau, and A. Frangioni, editors, *Contributions to Combinatorial Optimization and Applications – A Tribute to Bernard Gendron*, pages 179–205. Springer, 2024.

T.G. Crainic. Service Network Design for Consolidation-based Transportation – Advanced Topics. In S. Parragh and T. Van Woensel, editors, *Research Handbook on Transport Modeling*. Edgar Elgar Publishing, 2025a. forthcoming.

T.G. Crainic. Service Network Design for Consolidation-based Transportation – The Fundamentals. In S. Parragh and T. Van Woensel, editors, *Research Handbook on Transport Modeling*. Edgar Elgar Publishing, 2025b. forthcoming.

T.G. Crainic and M. Hewitt. Service Network Design. In T.G. Crainic, M. Gendreau, and B. Gendron, editors, *Network Design with Applications in Transportation and Logistics*, chapter 12, pages 347–382. Springer, Boston, 2021.

T.G. Crainic and K.H. Kim. Intermodal Transportation. In Barnhart, C. and Laporte, G., editors, *Transportation*, volume 14 of *Handbooks in Operations Research and Management Science*, chapter 8, pages 467–537. North-Holland, Amsterdam, 2007.

T.G. Crainic and G. Laporte. Planning Models for Freight Transportation. *European Journal of Operational Research*, 97(3):409–438, 1997.

T.G. Crainic and W. Rei. 50 Years of Operations Research for Planning Consolidation-based Freight Transportation. *EURO Journal on Transportation and Logistics*, 14:100157, 2025.

T.G. Crainic and J.-M. Rousseau. Multicommodity, Multimode Freight Transportation: A General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research Part B: Methodological*, 20:225–242, 1986.

T.G. Crainic, J.-A. Ferland, and J.-M. Rousseau. A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, 18(2):165–184, 1984.

T.G. Crainic, M. Hewitt, M. Toulouse, and D.M. Vu. Service network design with resource constraints. *Transportation Science*, 50(4):1380–1393, 2014.

T.G. Crainic, M. Hewitt, M. Toulouse, and D.M. Vu. Scheduled Service Network Design with Resource Acquisition and Management. *EURO Journal on Transportation and Logistics*, 7(3):277–309, 2018.

T.G. Crainic, M. Gendreau, and B. Gendron, editors. *Network Design with Applications in Transportation and Logistics*. Springer, Boston, 2021.

T.G. Crainic, B. Gendron, and M.R.A.. Kazemzadeh. A Taxonomy of Multilayer Network Design and a Survey of Transportation and Telecommunication Applications. *European Journal of Operational Research*, 303:1–13, 2022.

P.J. Dejax and T.G. Crainic. A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science*, 21(4):227–247, 1987.

M. Fischetti and M. Monaci. Using a general-purpose mixed-integer linear programming solver for the practical solution of real-time train rescheduling. *European Journal of Operational Research*, 263 (1):258–264, 2017.

M. Florian, G. Bushell, J. Ferland, G. Guertin, and L. Nastansky. The Engine Scheduling Problem in a Railway Network. *INFOR*, 14(2):121–138, 1976.

E. Frejinger and M. Hewitt. Perspectives on optimizing transport systems with supply-dependent demand. *INFOR: Information Systems and Operational Research*, 63(3):405–438, 2025.

M. Gorman. An Application of Genetic and Tabu Searches to the Freight Railroad Operating Plan Problem. *Annals of Operations Research*, 78:51–69, 1998.

A. Haghani. Formulation and Solution of Combined Train Routing and Makeup, and Empty Car Distribution Model. *Transportation Research Part B: Methodological*, 23(6):433–452, 1989.

K.C. Jha, R.K. Ahuja, and G. Şahin. New Approaches for Solving the Block-to-Train Assignment Problem. *Networks*, 51(1):480–62, 2008.

M. Joborn, T.G. Crainic, M. Gendreau, K. Holmberg, and J.T. Lundgren. Economies of Scale in Empty Freight Car Distribution in Scheduled Railways. *Transportation Science*, 38(2):459–464, 2004.

C. Joncour, J. Kritter, S. Michel, and X. Schepler. Generalized relax-and-fix heuristic. *Computers & Operations Research*, 149:106038, 2023.

M. Keaton. Designing Optimal Railroad Operating Plans: Lagrangian Relaxation and Heuristic Approaches. *Transportation Research Part B: Methodological*, 23(6):415–431, 1989.

M.H. Keaton. Designing Optimal Railroad Operating Plans: A Dual Adjustment Method for Implementing Lagrangian Relaxation. *Transportation Science*, 26:263–279, 1992.

T.L. Magnanti and R.T. Wong. Network Design and Transportation Planning: Models and Algorithms. *Transportation Science*, 18(1):1–55, 1984.

S. Mantovani, G. Morganti, N. Umang, T.G. Crainic, E. Frejinger, and E. Larsen. The Load Planning Problem for Double-Stack Intermodal Trains. Technical Report CIRRELT-2016-68, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2016.

S. Mantovani, G. Morganti, N. Umang, T.G. Crainic, E. Frejinger, and E. Larsen. The Load Planning Problem for Double-Stack Intermodal Trains. *European Journal of Operational Research*, 267(1): 107–119, 2017.

P. Miranda, J.-F. Cordeau, and E. Frejinger. A Time-Space Formulation for the Locomotive Routing Problem at the Canadian National Railways. Technical Report CIRRELT-2020-19, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et les transports, Université de Montréal, Montréal, QC, Canada, 2020.

G. Morganti, T.G.. Crainic, E. Frejinger, and N. Ricciardi. Block planning for intermodal rail: Methodology and case study. *Transportation Research Procedia*, 47:19–26, 2020.

E.K. Morlok and R.B. Peterson. A Final Report on a Development of a Geographic Transportation Network Generation and Evaluation Model. In *Proceedings of the Eleventh Annual Meeting*, pages 99–103. Transportation Research Forum, 1970.

J.P. Müller, R. Elbert, and S. Emde. Intermodal Service Network Design with Stochastic Demand and Short-term Schedule Modifications. *Computers & Industrial Engineering*, 159:107514, 2021.

A.M. Newman and C.A. Yano. Scheduling Direct and Indirect Trains and Containers in an Intermodal Setting. *Transportation Science*, 34(3):256–270, 2000.

A. M. Newman, L. K. Nozick, and C. A. Yano. Optimization in the Rail Industry. In P. M. Pardalos and M. G. C. Resende, editors, *Handbook of Applied Optimization*, pages 704–718. Oxford University Press, New York, NY, 2002.

H.N. Newton, C. Barnhart, and P.H. Vance. Constructing Railroad Blocking Plans to Minimize Handling Costs. *Transportation Science*, 32(4):330–345, 1998.

L.K. Nozick and E.K. Morlok. A Model for Medium-Term Operations Planning in an Intermodal Rail-Truck Service. *Transportation Research Part A: Policy and Practice*, 31(2):91–108, 1997.

. Ortiz-Astorquiza, J.-F. Cordeau, and E. Frejinger. The Locomotive Assignment Problem with Distributed Power at the Canadian National Railway Company. *Transportation Science*, 55(2):510–531, 2021.

M.B. Pedersen and T.G. Crainic. Optimization of Intermodal Freight Service Schedules on Train Canals. Publication CIRRELT-2007-51, Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport, Montréal, QC, Canada, 2007.

M.B. Pedersen, T.G. Crainic, and O.B.G. Madsen. Models and Tabu Search Meta-heuristics for Service Network Design with Asset-Balance Requirements. *Transportation Science*, 43(2):158–177, 2009.

F. Piu and M.G. Speranza. The Locomotive Assignment Problem: A Survey on Optimization Models. *International Transactions in Operational Research*, 21(3):327–352, 2014.

F. Piu, V. Première Kumar, M. Bierlaire, and M.G. Speranza. Introducing a Preliminary Consists Selection in the Locomotive Assignment Problem. *Transportation Research Part E: Logistics and Transportation Review*, 82:214–237, 2015.

Y. Pochet and L. A. Wolsey. *Production Planning by Mixed Integer Programming.* Springer Series in Operations Research and Financial Engineering. Springer, New York, NY, 1 edition, 2006.

U. Sadana, A. Chenreddy, E. Delage, A. Forel, E. Frejinger, and T. Vidal. A survey of contextual optimization methods for decision-making under uncertainty. *European Journal of Operational Research*, 320(2):271–289, 2025. doi: https://doi.org/10.1016/j.ejor.2024.03.020.

B. Vaidyanathan, R.K. Ahuja, J. Liu, and L.A. Shughart. Real-Life Locomotive Planning: New Formulations and Computational Results. *Transportation Research Part B: Methodological*, 42(2):147–168, 2008a.

B. Vaidyanathan, R.K. Ahuja, and J.B. Orlin. The Locomotive Routing Problem. *Transportation Science*, 42(4):492–507, 2008b.

D.M. Vu, T.G. Crainic, and M. Toulouse. A Three-Stage Matheuristic for the Capacitated Multicommodity Fixed-Cost Network Design with Design-Balance Constraints. *Journal of Heuristics*, 19:757–795, 2013.

W.W. White. A Program for Empty Freight Car Allocation. Technical Report 360D.29.002, IBM Contributed Program Library, IBM Corporation, Program Information Department, Hawthorne, N.Y., 1968.

W.W. White. Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers. *Networks*, 2(3):211–236, 1972.

W.W. White and A.M. Bomberault. A Network Algorithm for Empty Freight Car Allocation. *IBM Systems Journal*, 8(2):147–171, 1969.

M. Yaghini and R. Akhavan. Multicommodity Network Design Problem in Rail Freight Transportation Planning. *Procedia Social and Behavioral Sciences*, 43:728–739, 2012.

M. Yaghini, M. Mirghavami, and A.Z. Andaryan. A Mathematical Formulation and an LP-based Neighborhood Search Matheuristic Solution Method for the Integrated Train Blocking and Shipment Path Problem. *Networks*, 78:523–542, 2021.

E. Zhu, T.G. Crainic, and M. Gendreau. Scheduled Service Network Design for Freight Rail Transportations. *Operations Research*, 62(2):383–400, 2014.
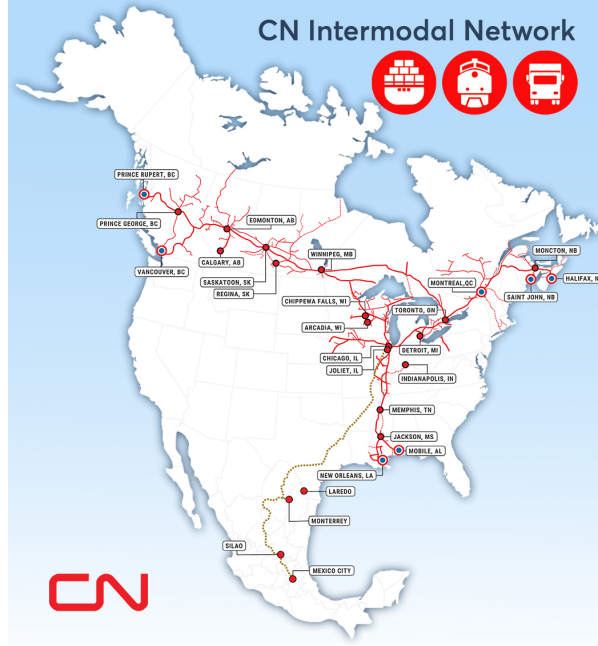
28

A   Intermodal network

Figure 6: Canadian National Intermodal Network (Canadian National Railway Company 2025)

B   Objective Function Parameter Values

The objective function for our application is the following:

$$
\text{Minimize} \sum_{b \in \mathcal{B}} \left(\left(c^{\text{BUILD}} + c^{\text{TRANS}} n_b^{\text{TRANS}}\right) y_b + \left(c^{\text{WAIT}} t_b^{\text{WAIT}} + c^{\text{BORD}} n_b^{\text{BORD}} + c^{\text{KM}} d_b\right) \left(\sum_{k \in \mathcal{K}_b} z_{bk} + \sum_{\gamma \in \Gamma} w_b^{\gamma} \right) \right)
$$

$$
+ \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}} c^{\text{LATE}} t_{bk}^{\text{LATE}} z_{bk} + \sum_{\gamma \in \Gamma} \sum_{\theta \in \Theta} \eta^{\gamma} c^{\text{ALLOC}} w_{\theta}^{\gamma} + \sum_{k \in \mathcal{K}} c^{\text{NDEL}} z_k + \sum_{\sigma \in \Sigma_{\text{EXTRA}}} \left(c^{\text{FIX}} + c^{\text{VAR}} * \sum_{a \in \mathcal{A}_{\sigma}^{\text{TM}}} u_a * d_a \right) s_{\sigma}
$$

(15)

It incorporates several cost components reflecting operational considerations. A construction cost c^{BUILD} is incurred for each block built, while a transfer cost c^{TRANS} is charged for every block transferred between trains (n_b^{TRANS}), penalizing the monetary and temporal burden of such operations. Transfers also generate a waiting cost c^{WAIT} per empty railcar or demand for each minute of transfer time (t_b^{WAIT}). A cost c^{BORD} is incurred for every border crossing (n_b^{BORD}). To discourage circuitous routes, a distance cost c^{KM} is applied per kilometer traveled (d_b), complemented by a lateness cost c^{LATE} per demand for each minute exceeding the shortest-path travel time (t_{bk}^{LATE}). An allocation cost c^{ALLOC} per platform is introduced for assigning railcars at the start of the horizon, scaled by the number of platforms (η^{γ}) of each railcar type $\gamma \in \Gamma$. To ensure high service quality, unmet demands (i.e., flows using artificial arcs) incur a large penalty cost c^{NDEL}. Finally, additional trains incur a fixed cost c^{FIX} and a variable cost dependent on capacity and distance traveled. No explicit loading cost is included, as the loading process implicitly enforces maximum utilization of preferred platforms.

The cost values were established in close collaboration with our industrial partner. They are the following: $c^{\text{BUILD}} = 100$, $c^{\text{TRANS}} = 10020$, $c^{\text{WAIT}} = 1$, $c^{\text{BORD}} = 1000$, $c^{\text{KM}} = 0.75$, $c^{\text{LATE}} = 1$, $c^{\text{ALLOC}} = 200$, $c^{\text{NDEL}} = 100000$, $c^{\text{FIX}} = 700000$, $c^{\text{VAR}} = 0.01$. It is important to note that these do

not represent the real monetary costs that the company pays to route demands through their railroad network.

C   Unrestricted Fleet Model

The model without fleet management can be written as:

$$\text{Minimize} \sum_{\sigma \in \Sigma_{\text{EXTRA}}} f_\sigma s_\sigma + \sum_{b \in \mathcal{B}} f_b y_b + \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} c_{bk} z_{bk} + \sum_{k \in \mathcal{K}} c_k z_k \tag{16}$$

Subject to:

$$\sum_{b \in \mathcal{B}_k} z_{bk} + z_k = \upsilon_k, \qquad\qquad k \in \mathcal{K}, \tag{17}$$

$$z_{bk} \le \upsilon_k y_b, \qquad\qquad b \in \mathcal{B}_k, \ k \in \mathcal{K}, \tag{18}$$

$$\sum_{k \in \mathcal{K}_b | \tau_k = \tau} z_{bk} = \sum_{\pi \in \Pi} \left(\nu_{b\pi}^\tau + 2\nu_{b\pi}^{\tau,\tau} + \sum_{\tau' \in \mathcal{T} | \tau' \ne \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad\qquad \tau \in \mathcal{T}, b \in \mathcal{B}, \tag{19}$$

$$\sum_{\gamma \in \Gamma} \eta_\pi^\gamma x_b^\gamma \ge \sum_{\tau \in \mathcal{T}} \left(\nu_{b\pi}^\tau + \nu_{b\pi}^{\tau,\tau} + \frac{1}{2} \sum_{\tau' \in \mathcal{T} | \tau' \ne \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad\qquad \pi \in \Pi, b \in \mathcal{B}, \tag{20}$$

$$\sum_{\gamma \in \Gamma_\pi} x_b^\gamma \le \sum_{\tau \in \mathcal{T}} \left(\nu_{b\pi}^\tau + \nu_{b\pi}^{\tau,\tau} + \frac{1}{2} \sum_{\tau' \in \mathcal{T} | \tau' \ne \tau} \nu_{b\pi}^{\tau,\tau'} \right), \qquad\qquad \pi \in \Pi, b \in \mathcal{B}, \tag{21}$$

$$\sum_{\gamma \in \Gamma} \left\lceil \frac{\eta_{\pi_{40}}^\gamma}{2} \right\rceil x_b^\gamma \ge \nu_{b\pi_{40}}^{\tau_{40},\tau_{53}}, \qquad\qquad b \in \mathcal{B}, \tag{22}$$

$$\sum_{\gamma \in \Gamma} \lambda_\gamma x_b^\gamma \le u_b y_b, \qquad\qquad b \in \mathcal{B}, \tag{23}$$

$$\sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma x_b^\gamma \le u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{INIT}}, \tag{24}$$

$$\sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma x_b^\gamma \le s_\sigma u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{EXTRA}}, \tag{25}$$

$$\frac{1}{u_a} \sum_{b \in \mathcal{B}_a} \sum_{\gamma \in \Gamma} \lambda_\gamma x_b^\gamma \ge U_\sigma s_\sigma, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{UTM}}, \sigma \in \Sigma_{\text{EXTRA}}, \tag{26}$$

$$y_b \in \{0,1\}, \qquad\qquad b \in \mathcal{B},$$

$$s_\sigma \in \{0,1\}, \qquad\qquad \sigma \in \Sigma_{\text{EXTRA}},$$

$$x_b^\gamma, z_{bk}, z_k, \nu_{b\pi}^{\tau\tau'}, \nu_{b\pi}^\tau \in \mathbb{N}, \qquad\qquad k \in \mathcal{K}, b \in \mathcal{B}, \tau \in \mathcal{T}, \tau' \in \mathcal{T}, \pi \in \Pi,$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad a \in \mathcal{A}, i \in \mathcal{N}, j \in \mathcal{N}, \gamma \in \Gamma, \theta \in \Theta.$$

The objective function (16) minimizes the total cost of selecting extra trains (when relevant), selecting and building blocks, routing demand flows on blocks and artificial arcs.

Constraints (17) to (22) are the same as constraints (2) to (7) in the SSND-RM formulation. Constraints (23) to (26) have the same role as Constraints (8), and (12) to (14) in the SSND-RM formulation, but do not take into account the repositioning of empty railcars when calculating capacity usage.

D   Unrestricted Loading Model

The model without fleet management and loading is:

$$\text{Minimize} \sum_{\sigma \in \Sigma_{\text{EXTRA}}} f_\sigma s_\sigma + \sum_{b \in \mathcal{B}} f_b y_b + \sum_{b \in \mathcal{B}} \sum_{k \in \mathcal{K}_b} c_{bk} z_{bk} + \sum_{k \in \mathcal{K}} c_k z_k \tag{27}$$

Subject to:

$$\sum_{b \in \mathcal{B}_k} z_{bk} + z_k = v_k, \qquad\qquad\qquad k \in \mathcal{K}, \tag{28}$$

$$z_{bk} \leq v_k y_b, \qquad\qquad\qquad b \in \mathcal{B}_k, \ k \in \mathcal{K}, \tag{29}$$

$$\frac{1}{2} \sum_{\tau \in \mathcal{T}} \sum_{k \in \mathcal{K}_b | \tau_k = \tau} \lambda_\tau z_{bk} \leq u_b y_b, \qquad\qquad\qquad b \in \mathcal{B}, \tag{30}$$

$$\frac{1}{2} \sum_{b \in \mathcal{B}_a} \sum_{\tau \in \mathcal{T}} \sum_{k \in \mathcal{K}_b | \tau_k = \tau} \lambda_\tau z_{bk} \leq u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{INIT}}, \tag{31}$$

$$\frac{1}{2} \sum_{b \in \mathcal{B}_a} \sum_{\tau \in \mathcal{T}} \sum_{k \in \mathcal{K}_b | \tau_k = \tau} \lambda_\tau z_{bk} \leq s_\sigma u_a, \qquad\qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \ \sigma \in \Sigma_{\text{EXTRA}}, \tag{32}$$

$$\frac{1}{2u_a} \sum_{b \in \mathcal{B}_a} \sum_{\tau \in \mathcal{T}} \sum_{k \in \mathcal{K}_b | \tau_k = \tau} \lambda_\tau z_{bk} \geq U_\sigma s_\sigma, \qquad a \in \mathcal{A}_\sigma^{\text{TM}}, \sigma \in \Sigma_{\text{EXTRA}}, \tag{33}$$

$$y_b \in \{0,1\}, \qquad\qquad\qquad b \in \mathcal{B},$$

$$s_\sigma \in \{0,1\}, \qquad\qquad\qquad \sigma \in \Sigma_{\text{EXTRA}},$$

$$z_{bk}, z_k \in \mathbb{N}, \qquad\qquad\qquad k \in \mathcal{K}, b \in \mathcal{B}, \tau \in \mathcal{T}, \tau' \in \mathcal{T}, \pi \in \Pi,$$

$$\qquad\qquad\qquad\qquad\qquad\qquad a \in \mathcal{A}, i \in \mathcal{N}, j \in \mathcal{N}, \gamma \in \Gamma, \theta \in \Theta.$$

Objective function (27) minimizes the total cost of selecting extra trains (when relevant), selecting and building blocks and routing demand flows on blocks and artificial arcs.

Constraints (28) and (29) are the same as Constraints (2) and (3) in the SSND-RM formulation. Constraints (30) to (33) have the same role as constraints (8) and (12) to (14) in the SSND-RM formulation, but do not take into account the repositioning of empty railcars and loading of containers when calculating capacity usage. The formula transforms a number of containers into capacity usage (train length) by using railcar lengths and assuming double stacking is always possible.