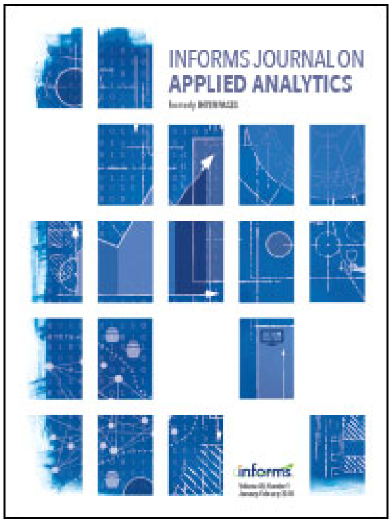


INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Improving Intermodal Train Operations in Indian Railways

Amit Upadhyay

To cite this article:
Amit Upadhyay (2020) Improving Intermodal Train Operations in Indian Railways. INFORMS Journal on Applied Analytics 50(4):213-224. https://doi.org/10.1287/inte.2020.1035

INFORMS JOURNAL ON APPLIED ANALYTICS
Vol. 50, No. 4, July–August 2020, pp. 213–224
ISSN 0092-2102 (print), ISSN 1526-551X (online)
http://pubsonline.informs.org/journal/inte

Improving Intermodal Train Operations in Indian Railways

Amit Upadhyay[a]

[a] Vinod Gupta School of Management, Indian Institute of Technology Kharagpur, 721302 West Bengal, India
Contact: amitupadhyay85@gmail.com, https://orcid.org/0000-0001-9731-4089 (AU)

Received: August 28, 2019
Revised: November 23, 2019
Accepted: February 6, 2020
Published Online in Articles in Advance: May 13, 2020

https://doi.org/10.1287/inte.2020.1035

Abstract. In Indian Railways, container train operators (CTOs) run intermodal trains. The success of a CTO depends on it providing timely delivery of containers at low haulage cost. The CTO must utilize its rolling stock efficiently, select containers, and assign these containers to wagons optimally considering multiple conflicting requirements. I discuss a mathematical programming–based approach we developed for a major CTO for train-load planning, which reduces the average cost of container haulage and increases the timeliness desired by customers. After being used to plan more than 1,000 trains, my model has been estimated to save about 2% in rail haulage cost, which corresponds to an annual saving of more than 300 million Indian rupees for Indian Railways trains. This study has led to a remarkable turnaround in the operations strategy of the operator, with a shift in emphasis from increasing train utilization to maximizing the contribution to profit.

History: This paper was refereed.

Keywords: intermodal trains • decision-support system • container transport • integer programming • Indian Railways

Introduction

In 2006, Indian Railways (IR) introduced competition in intermodal rail transport by allowing private operators, known as container train operators (CTOs), to obtain licenses to run container trains on the IR network. In the same year, the first double-stack container train was introduced to the IR network. Since 2006, the container rail market in India has nearly doubled in volume, mainly due to the liberalization of container rail transport and infrastructure development, resulting in the clearance of more routes for running double-stack trains. This increasing container rail traffic, as evidenced in Figure 1, is expected to grow further after completion of the 3,300-km-long dedicated freight corridors in 2021 (Dedicated Freight Corridors Corporation of India Ltd. 2019).

Today, more than 11 CTOs are competing for the container market in India. CTOs typically own or lease a fleet of rail wagons (aka “flatcars”) and containers. CTOs can also develop, own, operate, and maintain container terminals. CTOs collect containers (or goods) from consigners, store and load the containers onto trains, and collect the charges for these services from their customers. For the rail haulage of containers, CTOs use IR’s locomotives, network infrastructure, and associated services and pay a rail haulage charge (RHC) to IR. RHC is the major cost component that accounts for about 50% of the CTOs’ total operating cost.

To be competitive and to increase profit, the CTOs must focus on maximizing fleet utilization and minimizing RHC. They must make optimal decisions about container routing in the IR network and container-to-wagon assignments at terminals. Starting with this objective, a major CTO (henceforth referred to as CTOY) asked me to carry out an operations research (OR)/management science (MS) study in 2017. I decided to focus on the container train load–planning (CLTP) problem for reasons discussed in the Scope of Study section.

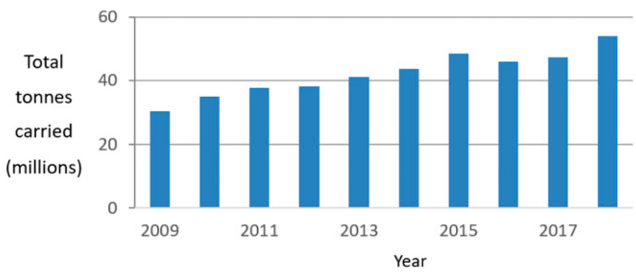
Problem Setting

In general, CTOY’s process for transporting containers from shippers to a port for export is as follows. Upon arrival at a terminal, the export containers are stored in a container yard according to their destination, type, and vessel cutoff time. Typically, about five hours before the expected departure of an outbound train, CTOY’s rail operations team requests a locomotive from IR. Upon arrival of an inbound train, IR personnel conduct a maintenance inspection of the train.

In another process, parallel to the train inspection, the terminal manager chooses a set of suitable containers from the storage yard and generates the container train-loading plan while considering many operational constraints. Then, using gantries and trailers (internal trucks), the yard operations team starts shifting the selected containers from the storage yard to the rail yard, and the rail operations team starts loading the containers onto the wagons according to the train-loading plan.

Upon completion of the train loading, the rail operations team updates the wagon numbers in the final

Figure 1. (Color online) Container Rail Traffic in India Has Grown Steadily over the Last 10 Years



Note. Indian Railways annual reports 2009–2018.

train-load plan, generating a train summary report. CTOY shares the train summary report with IR, destination ports, and customers for further actions.

The four most common types of containers transported by CTOY are 20-foot and 40-foot general containers (referred to as GEN), which are 8.5 feet high, and 20-foot and 40-foot high-cube containers (referred as to HQ), which are 9.5 feet high. For brevity, I refer to a 20-foot container as 20-foot (plural 20s) and to a 40-foot container as 40-foot (plural 40s). Also, I do not discuss consideration of all the other types of containers (reefer, tank, open-top, etc.) having insignificant traffic volume for the purposes of this study.

The containers can be loaded on wagons according to the five loading patterns allowed by IR, as shown in Figure 2. The patterns are as follows:

1. One 40 on top of two 20s having the same height, that is, both 20s should be either 8.5 feet or 9.5 feet high
2. One 40 on top of another 40
3. Only one 40
4. Only two 20s of any height
5. An empty wagon

For safety reasons, IR does not allow loading of four 20s in double stack on a wagon nor does it allow only one 20 to be loaded on a wagon.

In IR, a set of coupled wagons is referred to as a *rake*, that is, a rake is a train without any locomotives. IR does not allow CTOY to couple or decouple the wagons. Therefore, CTOY has to redistribute all the rakes without changing the wagon-composition of the rakes. In this paper, I assume that every train has 45 wagons (the most common train size in IR).

CTOY operates trains in the Northern and Western Railways' network shown in Figure 3, which spans over 20,000 kilometers (km) of track length. Although double stacking of containers provides better train utilization, CTOY often has to run single-stack trains for the following three reasons: (i) double stacking is prohibited on the route, for example, Ludhiana-Delhi; (ii) there are not enough containers for double stacking; or (iii) there are empty rakes to be redistributed in the network. For confidentiality, I have made minor changes in the data, which do not affect the results of this study.

The movement of CTOY's single-stack and double-stack trains in the complex rail network is illustrated with an example in Figure 4(a). To transport 180 20-foot equivalent units (TEUs) from Ludhiana (L) to Mundra (M), CTOY may load the containers in two single-stack trains, 90 TEUs each, toward a hub (H) in Rajasthan because double stacking is allowed only on the route H–M. At hub H, all of these 180 TEUs may be loaded on one of the two inbound trains, and the resulting outbound double-stack train can then be dispatched to M. Alternatively, among other options, CTOY may dispatch the two single-stack trains along the shortest direct route L–M.

Although the total RHC for dispatching the 180 TEUs using the two single-stack and one double-stack trains is less than the total RHC for dispatching the same TEUs via the shortest route, it is quite difficult for the planners to determine the maximum number of TEUs that can be loaded in the double-stack train. Therefore, the interrelationship between container train routing and container train-loading decisions is quite important and complex for CTOY's network.

Previous Planning Approach

Prior to adopting the optimization model, CTOY planned trains using a manual process. That process began with generating a network-wide train-routing plan every day, that is, deciding the number of trains to run and the origin-destination (OD) and intermediate stop of each train. Note that IR allows at most one intermediate stop for each train where the loading and/or unloading of the train can occur. The train-routing process required information mainly on container inventory at all terminals from an enterprise resource planning (ERP) application, target vessel cutoff

Figure 2. (Color online) Patterns Allowed by IR for Loading of Containers on Wagons

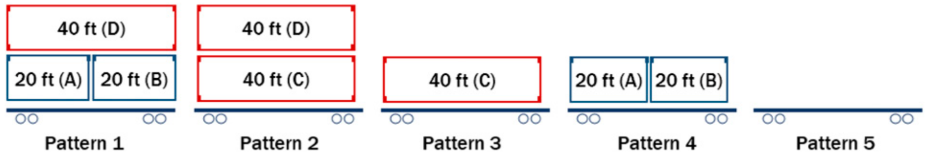


Figure 3. (Color online) CTOY Operates Trains in the Northern and Western Railways' Network that Handle the Heaviest Container Traffic in India

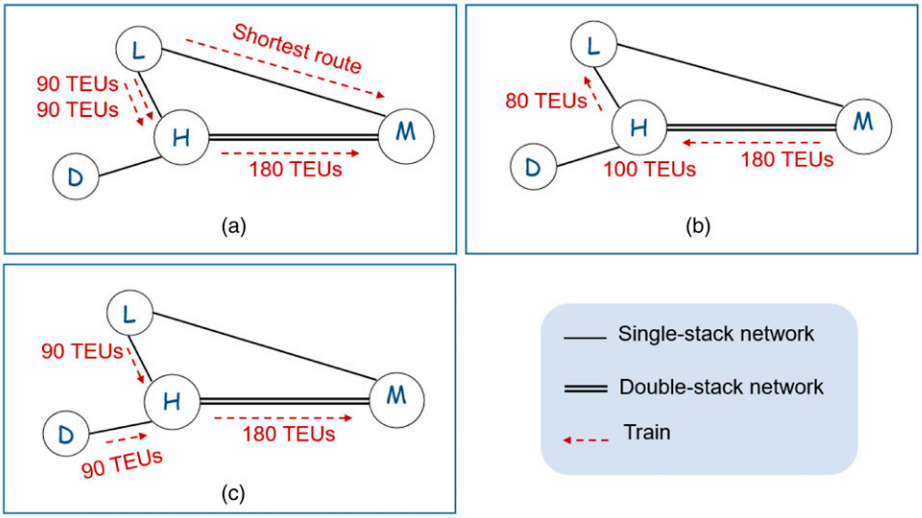
Source. Indian Railways network map (https://erail.in/info/railway-maps-indian-railways/1808).

times from shipping lines, and network congestion and the location of running trains from IR. Then, CTOY planners analyzed container flow imbalances in the network through manual calculations and used static rules of thumb to find the hub-and-spoke and double-stack-train opportunities to minimize RHC.

In this context, I define *rake-deficit* terminals as the terminals having more outbound containers than inbound containers, and *rake-surplus* terminals as those having more inbound than outbound containers. CTOY gave higher priority to the allocation of empty rakes to the trains outbound from rake-deficit terminals and tried to load these trains to their fullest extent through double stacking. On the other hand, the trains outbound from rake-surplus terminals were often moved empty or loaded in single-stack configuration depending on demand factors.

Upon finalizing the train-routing plan, the terminal managers, in consultation with the planner at the head office, prepared the loading plan for each train at their respective terminals. The managers made two major decisions: shortlisting of the set of containers that can be dispatched to the given train (hereafter, this set of containers is referred to as *pendency*), and assignment of these containers to suitable positions on the train. In the mostly manual planning process, the managers used a crude heuristic algorithm to generate a feasible train-loading plan. In practice, the heuristic always produced an inefficient loading plan, and the managers had to find a more suitable

Figure 4. (Color online) An Illustrative Rail Network with Double Stacking Permitted Only on Route H–M



Note. Panels (a) and (c) each show an example of train movements in the export direction, and panel (b) shows an example for the import direction.

reassignment for many containers through manual calculations at the end. The heuristic performed poorly when customers had special requirements or when the train had an intermediate stop. Typically, the entire process of generating, analyzing, and finalizing a train-loading plan took about three man-hours.

In principle, CTOY sought to increase its fleet utilization by reducing empty train movements and by increasing double stacking. In practice, CTOY planners struggled to trade off these objectives while accounting for the complexities of train-load planning. Recognizing the potential benefits of analytical decision support, CTOY asked me to improve their planning decisions.

Scope of Study

Most of the CTOY managers were quite skeptical about the potential for this project, which represented the company's first use of OR/MS. Several operational issues were brought to light during our project-inception meetings. I explained that it was not possible to develop a comprehensive package of analytical tools to plan train loading, train routing, and fleet utilization owing to time and budget constraints.

Among other shortcomings of the existing train load–planning process, I identified some major issues that were not being considered adequately. Frequently, the trains were not loaded to their full capacity despite a large pendency of containers. It was common to see a loading of 176 or 178 TEUs instead of 180 TEUs on the trains originating from rake-deficit terminals. CTOY did not have any way to know whether this inefficient loading was due to some genuine operational constraint, the inefficiency of the heuristic, or a mistake in the manual calculations. Furthermore, there was no mechanism to analyze the cost implications of any change in customer requirements or container characteristics.

Sometimes, the load plan had to be modified manually late in the process owing to some uncertainties. These modifications might also have resulted in lower utilization of trains in the absence of any reoptimization model. Occasionally, some containers were dispatched to a wrong destination as a result of human error during manual adjustment of the train-loading plan. Such mistakes were quite costly, leading to almost double RHC, a long delay in container delivery, and reputation damage for CTOY from a customer service standpoint. Therefore, considering CTOY's desire to reduce RHC through the OR project, I decided to focus on mathematical optimization of the CTLP process.

In a meeting attended by all of the CTOY stakeholders, we discussed multiple objectives and constraints that should be considered in the CTLP optimization model. Everyone agreed about the importance of maximizing train utilization for lower RHC and faster container delivery. However, it was quite challenging to consider other conflicting requirements related to service quality and intraterminal handling of containers.

CTOY had two key requirements related to service quality at the train-loading level. First, containers should be loaded on a first-come-first-served (FCFS) basis to the extent possible, which was expected to minimize delivery delays and customer grievances related to fair dispatching policy. Second, all the containers listed in a shipping bill, referred to as a *lot*, should be loaded onto the same train. Lot-breaking should be avoided because sending all the containers of a lot together in one train helps to manage inspection, storage, and further dispatching of these containers at the destination terminal.

Regarding intraterminal handling of containers, CTOY wanted to minimize total gantry-trailer movements required for loading the containers on wagons. The terminal managers argued that container-handling operations must be considered in the CTLP model because gantry-trailer movements are directly affected by the train-loading plan. Although that is true, I explained that intraterminal operations planning is quite a complex problem in and of itself, which requires a complete consideration of the yard layout, location of containers in the yard, and allocation of gantries and trailers available. Therefore, consideration of intraterminal operations in the CTLP would make the model mathematically intractable. Therefore, the intraterminal operations planning problem is better formulated separately, as discussed by Murty et al. (2005).

Furthermore, I carried out a detailed analysis of major trade-offs involved in the train-loading process, which is discussed further in the Solution Implementation section. For instance, by optimizing train utilization, CTOY can save up to 50,000 Indian rupees (INR) in RHC by loading just one extra 40. This saving of INR 50,000 is at least 50 times more than the cost of shifting the container from the yard to its wagon. Therefore, I gave lower importance to the cost of intraterminal operations than RHC and considered these operations partially in my CTLP model by formulating a novel constraint that reduces the handling of containers for a common case of the train having an intermediate stop.

I also assessed simultaneous optimization of the load plans for multiple trains, which may yield more profit. However, the multitrain CTLP is quite challenging owing to additional complexities of train routing in the hub-and-spoke network. Therefore, I elected to deploy only a single-train CTLP model, which also offers two key practical advantages: a single-train model takes less computation time because of reduced complexity of the problem and generates an implementable loading plan as a result of reduced impact of the uncertainties in the operations. In the last few hours, these uncertainties may arise in (i) the pendency due to new arrivals or inaccessibility of containers in the yard; (ii) container preferences; and (iii) the schedule of vessels, trains, and trailers. Thus, even if I can obtain an optimal solution for multitrain CTLP, the actual load plan may become inefficient or even infeasible owing to the cumulative effect of these uncertainties. Moreover, my single-train CTLP model also partly considers the double stacking of future trains, as explained in the Objective Function section. Upadhyay et al. (2017) argue that a rolling-horizon optimization approach, as proposed by Lai et al. (2008), requires significantly more computation time for the multitrain CTLP and does not guarantee a better solution than the single-train CTLP, especially when the uncertainties are significant.

Only a few studies relevant to CTLP appear in the literature. I refer readers to Upadhyay et al. (2017) for a comprehensive review of literature on CTLP and to Boysen et al. (2013) and Steenken et al. (2004) for broader reviews on operations planning problems at container terminals. Upadhyay et al. (2017) presented an optimization model for CTLP considering the complexities and operational constraints of a real-life application in IR. This study extends the research in Upadhyay et al. (2017) by modifying the model to consider CTOY's requirements and recent developments in IR. In this paper, I focus on implementation issues and business insights, and I refer the readers to Upadhyay et al. (2017) for more discussion on the theoretical aspects of CTLP.

The CTLP model presented here has evolved after being employed for the loading of hundreds of trains. My key contributions, not considered by Upadhyay et al. (2017), are as follows. I two key loading constraints imposed by IR. First, empty wagons and double-stack wagons cannot coexist in a train because such double stacking increases the vertical center of mass of the train unduly. Second, an empty 20 can be paired with only another empty 20 on any wagon, which maintains weight balance on the wagon in the horizontal direction. If the weight on any wagon is heavily biased on one side, it may raise the wheels of the wagon on the opposite side, which may derail the wagon when the train is moving. As an approximation to simplify the loading rules, IR has prohibited the pairing of a loaded 20 with an empty 20.

In addition, I consider multiple ODs of containers and develop a novel constraint to reduce container-handling efforts at an intermediate hub. I discuss important trade-offs in the implementation of the new constraints and customer requirements related to urgent delivery and lot-breaking, which are explained in the Solution Implementation section.

Optimization Model

I define the CTLP problem as maximizing CTOY's profit by selecting the containers from a given set of containers and assigning them to the wagons of a given train while satisfying operational constraints. I model the CTLP problem as a binary integer program (see the appendix). To facilitate an easy comparison, I use the same notations as used by Upadhyay et al. (2017). The program outputs the maximum-profit (minimum RHC) container train-loading plan.

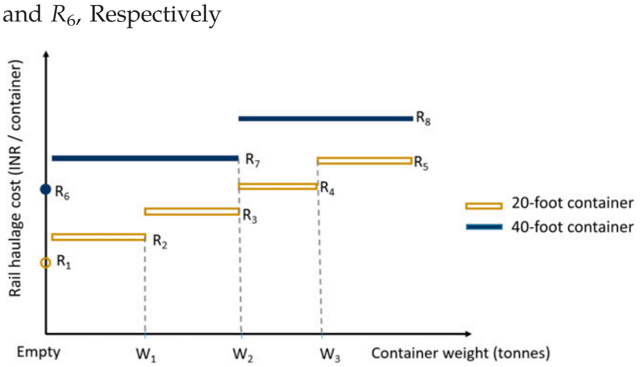
In case of a large pendency, CTOY wants to dispatch the containers on a FCFS basis to the extent possible. However, following the FCFS rule can adversely affect the more important objective, namely, minimizing RHC. Therefore, I incorporated the following two provisions in my model that can help improve customer satisfaction: (i) the planner can indicate a cutoff time, typically based on container booking time, to shortlist the oldest containers at the preprocessing stage, which reduces the number of the input containers for the model; and (ii) the planner can label urgent containers as compulsory, which triggers adding hard constraints to the model.

Objective Function

RHC increases with increasing haulage distance, weight, and length of the container, as shown in Figure 5. From an IR point of view, the fixed cost of rail haulage (associated with locomotive, crew, and infrastructure) is high and does not depend on the number of containers loaded on the train. Therefore, to encourage double stacking of containers in the network shown in Figure 3, IR charges less RHC for the containers loaded in the upper stack than for those in the lower stack. I use P_i^L and P_i^U to denote the profits from assigning the container i to lower- and upper-stack positions on the given train, respectively. I calculate P_i^L and P_i^U by considering only the revenue and cost components that are pertinent to the rail haulage of the container. Excluding irrelevant cost components does not affect the optimality of the load plan. If I set $P_i^L = 1$ for all 20s and $P_i^U = P_i^L = 2$ for all 40s, then the objective function (Equation (1) in the Appendix) maximizes train utilization.

In the CTLP model, I also consider the utilization of future trains indirectly by incorporating a factor α ($0 < \alpha \leq 1$) in the objective function. Note in Figure 2 that only 40s can be loaded in the upper stack. Therefore, in case of a relative shortage of 40s compared with 20s, I set $\alpha < 1$ (e.g., $\alpha = 0.5$) to penalize the assignment of 40s to the lower-stack positions. This penalty minimizes the number of 40s used in the current train without reducing the train utilization and saves the 40s for loading in the upper stack of subsequent trains. Note that α should always be greater than zero because an available 40 should be loaded if there is any empty wagon in the train. As explained in Figure 6, α is not required (i.e., $\alpha = 1$) when 40s are abundant or when empty slots are available on the train.

Figure 5. (Color online) For a Given Haulage Distance, RHC for an Empty 20 and an Empty 40 are R_1 and R_6, Respectively



Note. RHC increases with the container's weight according to weight ranges: from R_2 to R_5 for weight ranges of 20s, and from R_7 to R_8 for 40s.

Constraints

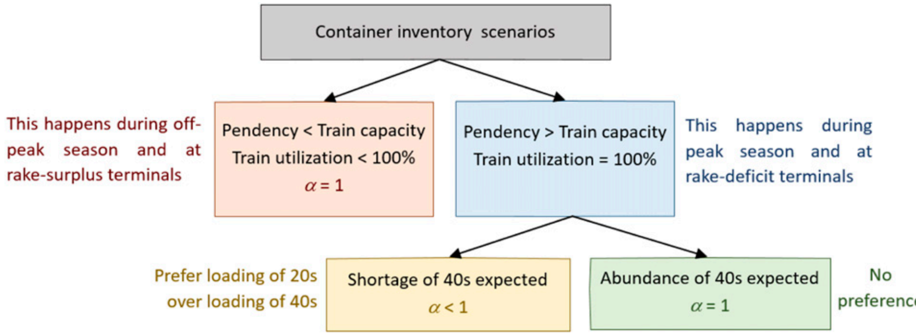
I classify the constraints into four main categories: pattern feasibility, safety, service quality, and lower-stack loading constraints. The pattern feasibility constraints (Equations (A.2)–(A.8) in the formulation presented in the appendix) are necessary for generating a feasible loading plan. Constraints (A.2) ensure that exactly one loading pattern is assigned to each wagon. Constraints (A.3)–(A.5) specify the requirements of the loading patterns. Constraints (A.3) assign one 20 each to the positions A and B (shown in Figure 2) if loading pattern 1 or 4 is used. Constraints (A.4) assign a 40 to position C if pattern 2 or 3 is used, and Constraints (A.5) assign a 40 to position D if pattern 1 or 2 is used. Constraints (A.6) ensure the two 20s assigned to a wagon in pattern 1 are of the same height. Constraints (A.7) and (A.8) restrict the assignment of each noncompulsory container to at most one wagon.

The safety constraints (A.9)–(A.14) consider IR's safety guidelines for train loading, which CTOY has to respect. Constraints (A.9) restrict the total weight of the containers loaded on each wagon not to exceed the wagon's payload limit. Constraints (A.10) stipulate that the total weight of the lower-stack container(s) on each wagon should not be less than the weight of the upper-stack container. Constraints (A.11)–(A.13) ensure that a container can be loaded in the upper stack only if all the lower-stack positions on the train have been occupied. Constraints (A.14) ensure that an empty 20 can be paired with only another empty 20 on any wagon.

The service quality constraints (A.15)–(A.17) consider special requests from the customers. Constraints (A.15) forbid the lot-breaking for shipping bills, where I_{20}^s and I_{40}^s denote the sets of 20s and 40s belonging to shipping bill $s \in S$, respectively. Constraints (A.16) and (A.17) ensure that the compulsory containers, denoted by I_{20}^C and I_{40}^C, are assigned to the train.

For a double-stack train having an intermediate stop, I reduce container-handling efforts at the intermediate stop through a unique constraint (A.18), referred to as the lower-stack loading constraint. At the origin of the train, I load the containers heading for the last stop of the train in the lower-stack positions so that there is no handling of these containers at the intermediate stop. Here, I give preference to the desired sets of 20s and 40s, denoted by I_{20}^{LS} and I_{40}^{LS},

Figure 6. (Color online) Usage of Factor α Under Different Container Inventory Scenarios



for loading in the lower stack at any cost. Of course, if the sets I_{20}^{LS} and I_{40}^{LS} are small, then the model has the flexibility to optimize the loading of the other containers in the lower stack, as explained in the next section.

Solution Implementation

Because CTOY has no internal analytics team, my key challenge was to address the concerns of all the stakeholders and develop their trust in the optimization model. I held several meetings with the planners, terminal managers, and top management. However, they could not realize the complete implications of their conflicting requirements until the first trial implementation of the CTLP model. I conducted a training program on using the model and explained the key trade-offs involved in the load-planning process. I developed an acceptable solution approach by considering CTOY's preferences for the key trade-offs, mainly between RHC, lower-stack ODs, lot-breaking, compulsory containers, computation time, and ease of use.

CTOY found the provision for loading desired ODs in the lower stack (Constraint (A.18)) useful for reducing container handling at the intermediate stop. For example, in Figure 4(c), the single-stack train originating from Ludhiana (L) is loaded such that all the 40s (90 TEUs) loaded in the single-stack train from Dadri (D) can be feasibly loaded on the upper stack of the train from Ludhiana. Thanks to the model, both of the terminal managers can plan their trains such that no Ludhiana container needs to be handled at the hub.

Similarly, Constraint (A.18) is quite useful for the import trains. For example, in Figure 4(b), assume that a total of 80 TEUs destined for Ludhiana are available at Mundra, and CTOY wants to dispatch the rake of the double-stack train (M–H) further as the single-stack train (H–L). In this case, the model will load the 80 TEUs in the lower stack and will try to load up to 100 TEUs of non-Ludhiana containers in the remaining positions on the train. As a result, only the non-Ludhiana containers are unloaded at the hub and the train can quickly depart for Ludhiana. If there are additional containers for Ludhiana at the hub, then these containers can occupy the 10 TEUs of remaining space in this train.

Sometimes, Constraint (A.18) may lead to an inefficient load plan because it prefers the loading of I_{40}^{LS} and I_{20}^{LS} containers in the lower stack at any cost. For example, in Figure 4(b), if the 80 TEUs (M–L) are heavy and loaded in the lower stack of the wagons, then the model may be able to load only 176 TEUs in the train as a result of a shortage of the relevant lightweight 40s. However, if I relax Constraint (A.18), then the model may be able to load the full 180 TEUs in the same train. Conversely, I also get an inefficient loading plan if there are empty 20s in I_{20}^{LS}, along with a shortage of empty 40s for loading in the upper stack. In this case, if I relax Constraint (A.18), then the model may still load the full 180 TEUs by replacing the empty 20s with some loaded containers.

The planners found the provision for compulsory containers in CTLP counterintuitive. Initially, they tried to mark too many containers as compulsory, expecting that the model will load more containers. However, this often resulted in either no solution or an inefficient solution. I explained two reasons for getting the inefficient solution: (i) CTLP runtime typically exceeds the desired time limit (15 minutes) when $|I_{40}^{C}| + |I_{20}^{C}| > 20$; and (ii) the compulsory containers adversely affect the performance measures (mainly RHC). Because CTLP is inherently designed to maximize profit, one should not mark containers as compulsory with the errant expectation of increasing train utilization. A container should be marked compulsory only when moving that container is urgent because CTLP will try to load this container at any cost. For example, there may be no solution or an inefficient solution when the user marks as compulsory an odd number of 20s, or an odd number of HQ 20s, or when compulsory 20s do not have any space for loading onto the train ($I_{20}^{C} \not\subset I_{20}^{LS}$ and $2|I_{40}^{LS}| + |I_{20}^{LS}| > 2|K|$).

Similarly, lot-breaking constraints (A.15) can also lead to an infeasible or inefficient solution. Again, the two main reasons are (i) CTLP does not allow lot-breaking for I_{20}^s and I_{40}^s at any cost; and (ii) CTLP runtime exceeds the desired time limit in case of big lots, whereas the runtime is typically less than 15 minutes for the common case of small lots (i.e., $|I_{40}^s| + |I_{20}^s| < 10\ \forall\, s \in S$). Big lots, having many heavy or empty containers, may lead to higher RHC as well as an undesirable lower-stack loading (i.e., extra handling at the intermediate hub). Therefore, the lot-breaking constraint should be used sparingly.

Train Load–Planning Procedure

The CTLP application, hosted on a Microsoft® Windows server (4.2 GHz, 16 core, 64 gigabytes RAM), relies on the CPLEX optimization engine (version 12.6) and some preprocessing and postprocessing steps to generate the loading plan. The overall flow of the CTLP model–based process is shown in Figure 7.

Most of the input parameters are obtained from CTOY's ERP or set to its default values. Therefore, all of the inputs can be entered in but a few minutes. For example, if a wagon is found to be unfit for loading, its payload capacity is set to zero. If there is any preference for loading in the lower stack, the user can mention the desired OD pairs to shortlist I_{40}^{LS} and I_{20}^{LS}. Similarly, in case of a large pendency, as a rule of thumb, it is sufficient to specify the cutoff time such that the total number of TEUs is more than 2.5 times the train's capacity, that is, $2|I_{40}| + |I_{20}| > 10|K|$. Occasionally, in case of a large pendency for lower-stack ODs as well, the user may specify another cutoff time to reduce the number of containers in I_{40}^{LS} and I_{20}^{LS}.

CTLP performs a data-validation process and may generate error or warning flags if it finds incongruent values of any parameters that may lead to an inefficient or infeasible solution. For example, if any container is marked compulsory, CTLP flags a warning with the number of compulsory containers and a potential increase in RHC. Furthermore, by default, marking just one container from I_{40}^s or I_{20}^s as compulsory will also make all the containers in the associated lot compulsory owing to the lot-breaking constraint.

CTLP outputs an optimal load plan, as shown in Figure 8, along with the key metrics such as the total RHC, tonnage, and TEUs loaded. Many columns, such as tare weight and port of discharge of container, which are insignificant for the purpose of this study, have been removed from Figure 8.

CTLP also flags a warning for any expected inefficiency or issue in the output loading plan. For example, CTLP allows dispatching of an odd number of 20s for any OD at an intermediate hub, but sometimes there are no other relevant 20s with which to pair the odd 20 in the next connecting train. Refer to the trains in Figure 4(b); this problem arose when CTLP loaded 79 TEUs of M–L and 11 TEUs of M–D in a train on route M–H. Then, at the hub, the 10 TEUs of M–D were dispatched further in the next train, but the leftover 20 could not be loaded for a few days because of the unavailability of a suitable pair.

Ideally, the terminal manager should finalize the load plan well before train loading starts, which will enable better planning of the intraterminal operations. However, an important part of my training program focused on explaining how the operations team can minimize its container-handling efforts at the postprocessing stage. The wagon number allotted in the CTLP-generated load plan is not sacrosanct for the terminal operations because the total RHC, TEUs, lot-breaking, priority, lower-stack ODs, and safety criteria remain unaffected if I swap all the containers of one wagon with those of any other identical wagon. This CTLP property of multiple optimal solutions is quite valuable for long trains and can be exploited to reduce the container-handling efforts. For example, if two 40s, assigned to the first wagon in the load plan, are at a location in the yard that is close to the tail of the train, the operations team is free to load both 40s on any of the last wagons. Moreover, the team need not receive approval for making such changes.

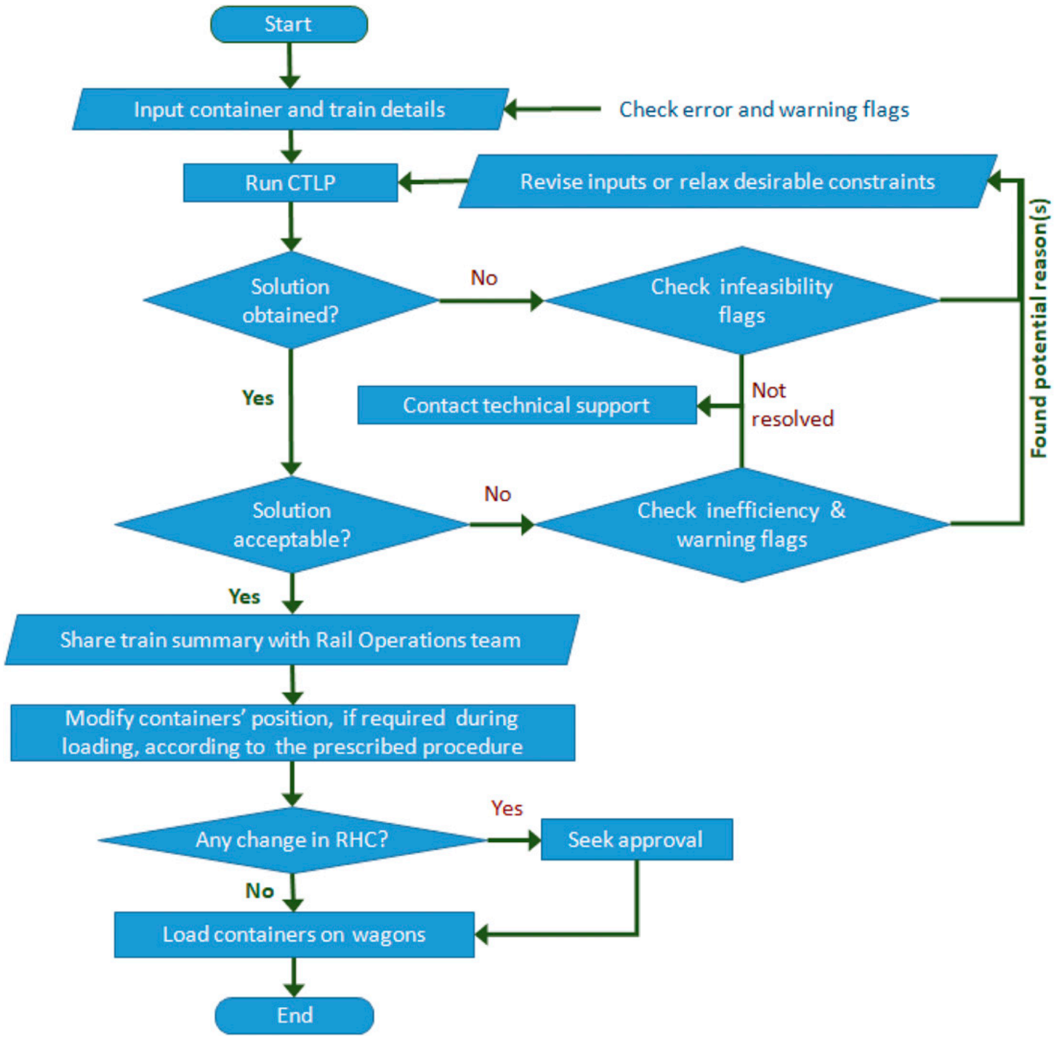
The same multiple optima property is also useful in dealing with any uncertainty in the last hour that requires changing the assignment of any container. The operations team should try to load the new container by replacing the one having identical characteristics, mainly OD and weight range. For example, a 23-tonne M–L 40 may be replaced by a 22-tonne M–L 40, if doing so is feasible, without affecting the RHC.

The CTLP addressed the concerns of all the stakeholders adequately. Therefore, an intangible but important benefit of this OR project was the credibility and acceptance of the load plans generated by CTLP in CTOY teams that emerged as the decision makers interacted with the model.

A planner can evaluate a single scenario (inputs entered, model executed, and output analyzed) typically within 15 minutes. The optimal load plan may be analyzed for a set of scenarios, if required. The CTLP runtime was quite acceptable for CTOY because this process used to consume about two hours before the model was introduced. Moreover, the planners were delighted to get an optimal load plan within a few minutes in case of a large pendency. Because the runtime of the heuristic used earlier increased with increasing pendency, the planners found it counterintuitive that CTLP runtime decreases with increasing pendency, that is, when $|I| > 300$.

Having understood the benefits from optimization, mainly a saving of up to INR 50,000 in RHC by being able to load just one extra 40, CTOY preferred to wait for an optimal solution over getting a quick heuristic solution. However, realizing the need for shorter

Figure 7. (Color online) CTLP Model–Based Process



CTLP runtime, I am trying to develop more efficient algorithms that will enable faster (i) evaluation of the important trade-offs (RHC, compulsory containers, lot-breaking, and lower-stack ODs) under multiple scenarios, (ii) reoptimization of the load plan to consider the uncertainties, and (iii) generation of load plans for longer (over 1.5 km) trains expected on the dedicated freight corridors.

Results

I compared the optimal and actual loading plans for more than 300 trains selected by CTOY. I observed the savings in RHC to vary from 0% to 9.5% and estimated the average savings to be about 2%. The 2% average savings in RHC implies an annual savings of more than INR 300 million for the trains in IR. The savings from CTLP were so high that CTOY discontinued using the manual planning approach in favor of the CTLP-based model immediately after the first trial implementation of the model in 2017.

For brevity, I illustrate key numerical cases based on the network illustrated in Figure 4, which shows a part of CTOY's network. I consider three terminals—L, H, and M—and compare the actual train-load plan generated by the old method and the optimal plan generated by CTLP for the five numerical cases in Table 1. For each case, the train-routing plan for the CTLP is the same as the actual plan. However, to be conservative about the input set of containers (I) for the CTLP, I restrict I to include only the containers that were actually loaded on the trains using the old method. For example, consider the first instance in Table 1, in which I assume that the set, I, includes only the 268 TEUs to be transported from M to L. In other words, for each instance in Table 1, I assume the smallest possible I so that there is no doubt about the potential availability of any container. Therefore, the savings mentioned in Table 1 are the minimum guaranteed savings. However, in practice, a rake-deficit terminal, such as Mundra, often has a larger pendency of containers that are

Figure 8. (Color online) Excerpt of a Train-Loading Plan Generated by CTLP

Container ID	Customer ID	Shipping Line	Origin	Dest	Size	Type	Gross Wt tonne	Wagon No.	Lower/ Upper	Wagon Wt tonne
PZU3617187	1288488	APL	Mundra	Jaipur	20	GEN	17.5	1	Lower	60.28
ONU0609163	1288516	MAI	Mundra	Jaipur	20	GEN	18.4	1	Lower	
YKU5462968	1288537	NYK	Mundra	Jaipur	40	GEN	24.38	1	Upper	
AXU3277477	1288516	MAI	Mundra	Jaipur	40	GEN	29.72	2	Lower	51.88
SKU9786241	1288516	MAI	Mundra	Jaipur	40	HQ	22.16	2	Upper	
RKU8117743	1288516	MAI	Mundra	Jaipur	20	GEN	16.17	44	Lower	57.59
OLU0290170	1288485	OOCL	Mundra	Jaipur	20	GEN	14.1	44	Lower	
EMU7379648	1288488	APL	Mundra	Jaipur	40	GEN	27.32	44	Upper	
AEU6851176	1288516	MAI	Mundra	Jaipur	40	GEN	28.39	45	Lower	50.04
YKU5696713	1288537	NYK	Mundra	Jaipur	40	GEN	21.65	45	Upper	

ready for dispatch, and a larger I would generally produce a more profitable load plan from CTLP. Nevertheless, significant savings were achieved by optimizing the load plan of the given trains through reassignment of the containers I to the wagons.

Each instance in Table 1 consisted of one double-stack train and zero to three single-stack trains loaded on the same day. Each of the trains had 45 wagons. The total number of TEUs loaded on each train is shown in parentheses. The TEUs of candidate 40s and 20s for each OD are presented in the "40s OD" and "20s OD" columns, respectively. The "Profit" column shows the cumulative profit for all of the trains loaded according to the old method, which is compared against the cumulative profit corresponding to the optimal loading plan to calculate the saving obtained from the CTLP in the last two columns.

In instance 1, four trains (three rakes) carrying a total of 268 TEUs were dispatched on the same day: one double-stack train on route M–H, two single-stack trains on route H–L for further dispatching of the containers on the double-stack train, and one single-stack direct train on route M–L. Because Mundra is rake-deficit, most of the trains were fully loaded. For such instances, CTLP first optimized the load plan of the double-stack train, and then the remaining ML-containers were assigned to the single-stack train. The result for instance 1 showed that the optimal plan increased the profit considerably by INR 119,330. Additionally, CTLP was also able to identify more suitable containers from the pendency to increase the utilization of the double-stack train from 178 to 180 TEUs. Note that this loading of an extra two TEUs might have increased the profit further by about INR 41,000, which was in addition to INR 119,330.

Instances 2 and 3 each had two OD pairs for containers in the export direction: LM and HM. Because both terminals L and H are rake-surplus, many trains were not fully loaded. For such instances, CTLP first optimized the assignment of 90 TEUs to the single-stack direct train-LM. The remaining LM-containers were assigned to the other single-stack train-LH. Then, CTLP optimized the load plan of the double-stack train-HM. We again observed a significant increase in profits for both instances.

The last two instances (4 and 5) each had only one OD pair for containers, MH, and only one double-stack train-MH. For such instances, because all the containers were to be assigned to the same train, the savings from CTLP came only from the optimal reassignment of the containers to the lower and upper stacks of the train. Nevertheless, significant savings, INR 87,423 and INR 31,540, were observed for the instances. Furthermore, I found that it was possible to load four extra TEUs in instance 5, which could have increased profit further.

After comparing many instances, like 1 and 5 in Table 1, one of the most critical findings for CTOY was to realize that many double-stack trains originating from rake-deficit terminals were not fully loaded owing to the inefficiency of their cumbersome manual planning process. Overall, after using CTLP to plan more than 1,000 trains, CTOY estimated the model to save about 2% in RHC by maximizing utilization of scarce wagons and by optimizing container-to-wagon assignments based on the marginal contribution of these assignments to profit.

The CTOY team was also delighted by my consideration of the lower-stack OD constraints, which have been proven to significantly reduce the handling of lower-stack containers at intermediate hubs. CTLP also eliminated inadvertent human errors, which during the manual planning process caused high extra cost and customer dissatisfaction.

Table 1. Comparison of Actual and Optimal Load Plans for Some Real Trains in the Network Shown in Figure 4

Instance no.	Containers loaded by the old method (= input *I*)		Actual loading plan using the old method			Optimal loading plan using CTLP model			Minimum increase in profit (savings)	
	40s OD (TEUs)	20s OD (TEUs)	Double-stack train (TEUs)	Single-stack train (TEUs)	Profit (INR)	Double-stack train (TEUs)	Single-stack train (TEUs)	Profit (INR)	INR	%
1	ML (102)	ML (166)	MH (178)	ML (90) HL (90) HL (86)	1,851,997	MH (178 + 2)	ML (90) HL (90) HL (90)	1,971,327	119,330	6.44
2	LM (90) HM (66)	LM (28) HM (44)	HM (138)	LM (90) LH (28)	1,490,769	HM (138)	LM (90) LH (28)	1,603,481	112,712	7.56
3	LM (70) HM (52)	LM (38) HM (54)	HM (124)	LM (90) LH (18)	1,550,989	HM (124)	LM (90) LH (18)	1,632,445	81,456	5.25
4	MH (116)	MH (64)	MH (180)	NA	1,277,459	MH (180)	NA	1,364,882	87,423	6.84
5	MH (96)	MH (80)	MH (176)	NA	1,226,552	MH (176 + 4)	NA	1,258,092	31,540	2.57

Note. In instances 1 and 5, the model found opportunities of loading an extra 2 TEUs and 4 TEUs, respectively, but these savings due to increased utilization of the double-stack trains are not included the calculation of the minimum savings mentioned in the last column.

I designed standard operating practices for the CTLP-based planning process and trained CTOY team members to avoid common pitfalls, deal with application errors, and consider special cases. I also recommended close monitoring of the inventory of twistlocks (a device essential for securing a container in the upper stack) at terminals. A safety stock of the twistlocks is essential at all double-stack terminals because a lack of just one twistlock can prevent double stacking, causing a loss of more than 20 times the purchase cost of the twistlock.

Thanks to CTLP, the planners can examine more scenarios and perform sensitivity analyses. Now, the planners can understand which parameters (40s, wagons, lot-breaking, etc.) are actually constraining operational flexibility and raising operational costs as well as when the parameters appear to be constraining but actually are not. The CTLP model helped in analyzing whether the requests for lot-breaking, urgency, and lower-stack loading can be fulfilled and at what cost. CTOY found this information quite valuable for pricing new containers and considering special requests from customers.

In summary, the CTLP model has empowered CTOY to optimize the loading of container trains across all terminals to ensure more profitable, faster, and customer-friendly train-load planning. The model has radically improved CTOY's train-planning process and has now become its primary decision-support tool. The model has brought about a remarkable turnaround in CTOY's operations strategy, a shift in emphasis from just increasing train utilization to maximizing the contribution to profit.

Acknowledgments

The author thanks CTOY for providing support and partial funding for this study.

Appendix

Sets

I Set of candidate containers; index $i \in I$

I_{20} , I_{40} Sets of 20s and 40s, respectively ($I = I_{20} \cup I_{40}$)

I_{20}^C , I_{40}^C Sets of compulsory 20s and 40s that must be loaded on the train ($I_{20}^C \subseteq I_{20}$, $I_{40}^C \subseteq I_{40}$)

I_{20}^E Set of empty 20s, $I_{20}^E \subseteq I_{20}$

I_{20}^{LS} , I_{40}^{LS} Sets of 20s and 40s to be given the highest preference for loading in the lower stack ($I_{20}^{LS} \subseteq I_{20}$, $I_{40}^{LS} \subseteq I_{40}$)

J Set of allowed loading patterns; index $j \in J \equiv \{1, 2, 3, 4, 5\}$

K Ordered set of wagons in the train; index $k \in K$ also refers to the wagon's numbered position counted from the front

S Set of shipping bills for which lot-breaking is not allowed; index $s \in S$

I_{20}^s , I_{40}^s Sets of 20s and 40s listed in a shipping bill. $s \in S$ $I_{20}^s \subseteq I_{20}$; $I_{40}^s \subseteq I_{40}$; $|I_{40}^s| + |I_{20}^s|$ is the number of containers listed in s

Parameters

G_k	Maximum payload limit for wagon $k \in K$
H_i	Height of container $i \in I$
m	Refers to a loading position shown in Figure 2: $m \in \{A, B\}$ for 20s and $m \in \{C, D\}$ for 40s
P_i^L, P_i^U	Profit parameters corresponding to the assignment of container i to the lower- and upper-stack positions in the train, respectively
W_i	Gross weight of container i
α	Parameter used for addressing a relative shortage of 40s for loading, $0 < \alpha \le 1$

Binary Decision Variables

$$x_k^j = \begin{cases} 1, \text{ if wagon } k \in K \text{ is assigned pattern } j \in J \\ 0, \text{ otherwise} \end{cases}$$

$$y_{ik}^m = \begin{cases} 1, \text{ if a 20-foot container } i \in I_{20} \text{ is assigned position } m \in \{A, B\} \text{ on wagon } k \in K \\ 0, \text{ otherwise} \end{cases}$$

$$z_{ik}^m = \begin{cases} 1, \text{ if a 40-foot container } i \in I_{40} \text{ is assigned position } m \in \{C, D\} \text{ on wagon } k \in K \\ 0, \text{ otherwise} \end{cases}$$

$$b_s = \begin{cases} 1, \text{ if all the containers in shipping bill } s \in S \text{ are assigned to the train} \\ 0, \text{ otherwise} \end{cases}$$

$$n = \begin{cases} 0, \text{ if the train load-planning solution leaves no wagon in the train empty} \\ 1, \text{ otherwise} \end{cases}$$

Formulation

$$\max \sum_{k \in K} \left[\sum_{i \in I_{20}} P_i^L \left(y_{ik}^A + y_{ik}^B \right) + \sum_{i \in I_{40}} \left(\alpha P_i^L z_{ik}^C + P_i^U z_{ik}^D \right) \right], \tag{A.1}$$

subject to

$$\sum_{j \in J} x_k^j = 1, \ \forall k \in K, \tag{A.2}$$

$$\sum_{i \in I_{20}} y_{ik}^m - x_k^1 - x_k^4 = 0, \ \forall k \in K, \ m \in \{A, B\}, \tag{A.3}$$

$$\sum_{i \in I_{40}} z_{ik}^C - x_k^2 - x_k^3 = 0, \ \forall k \in K, \tag{A.4}$$

$$\sum_{i \in I_{40}} z_{ik}^D - x_k^1 - x_k^2 = 0, \ \forall k \in K, \tag{A.5}$$

$$\left| \sum_{i \in I_{20}} H_i \left(y_{ik}^A - y_{ik}^B \right) \right| \le 1 - x_k^1, \ \forall k \in K, \tag{A.6}$$

$$\sum_{k \in K} \left(y_{ik}^A + y_{ik}^B \right) \le 1, \ \forall i \in I_{20} \backslash I_{20}^C, \tag{A.7}$$

$$\sum_{k \in K} \left(z_{ik}^C + z_{ik}^D \right) \le 1, \ \forall i \in I_{40} \backslash I_{40}^C, \tag{A.8}$$

$$\sum_{i \in I_{20}} W_i \left(y_{ik}^A + y_{ik}^B \right) + \sum_{i \in I_{40}} W_i \left(z_{ik}^C + z_{ik}^D \right) \le G_k, \ \forall k \in K, \tag{A.9}$$

$$\sum_{i \in I_{20}} W_i \left(y_{ik}^A + y_{ik}^B \right) + \sum_{i \in I_{40}} W_i z_{ik}^C \ge \sum_{i \in I_{40}} W_i z_{ik}^D, \ \forall k \in K, \tag{A.10}$$

$$\sum_{k \in K} \left(x_k^1 + x_k^2 \right) \le |K| (1 - n) ; //, \tag{A.11}$$

$$\sum_{k \in K} x_k^5 \le |K| \, n \, , \tag{A.12}$$

$$\sum_{k \in K} x_k^5 \ge n, \tag{A.13}$$

$$\sum_{i \in I_{20}^E} y_{ik}^A = \sum_{i \in I_{20}^E} y_{ik}^B, \ \forall k \in K, \tag{A.14}$$

$$\sum_{k \in K} \left(\sum_{i \in I_{20}^s} \left(y_{ik}^A + y_{ik}^B \right) + \sum_{i \in I_{40}^s} \left(z_{ik}^C + z_{ik}^D \right) \right) = b_s \left(\left| I_{40}^s \right| + \left| I_{20}^s \right| \right), \ \forall s \in S, \tag{A.15}$$

$$\sum_{k \in K} \left(y_{ik}^A + y_{ik}^B \right) = 1, \ \forall i \in I_{20}^C, \tag{A.16}$$

$$\sum_{k \in K} \left(z_{ik}^C + z_{ik}^D \right) = 1, \ \forall i \in I_{40}^C, \tag{A.17}$$

$$\sum_{k \in K} \left(\sum_{i \in I_{20}^{LS}} \left(y_{ik}^A + y_{ik}^B \right) + \sum_{i \in I_{40}^{LS}} \left(z_{ik}^C + z_{ik}^D \right) \right) \ge \min \left(2|K|, 2\left| I_{40}^{LS} \right| + \left| I_{20}^{LS} \right| - 1 \right), \tag{A.18}$$

$$x_k^j \in \{0, 1\}, \forall k \in K, j \in J, \tag{A.19}$$

$$y_{ik}^m \in \{0, 1\}, \forall k \in K, \ m \in \{A, B\}, i \in I_{20}, \tag{A.20}$$

$$z_{ik}^m \in \{0, 1\}, \forall k \in K, \ m \in \{C, D\}, i \in I_{40}, \tag{A.21}$$

$$b_s \in \{0, 1\}, \forall s \in S, \tag{A.22}$$

$$n \in \{0, 1\}. \tag{A.23}$$

References

Boysen N, Fliedner M, Jaehn F, Pesch E (2013) A survey on container processing in railway yards. *Transportation Sci.* 47(3):312–329.

Dedicated Freight Corridors Corporation of India Ltd. (DFCCIL) (2019) Salient features, DFCCIL. Accessed January 1, 2020, https://dfccil.com/Home/DynemicPages?MenuId=75.

Lai YC, Ouyang Y, Barkan CP (2008) A rolling horizon model to optimize aerodynamic efficiency of intermodal freight trains with uncertainty. *Transportation Sci.* 42(4):466–477.

Murty KG, Wan YW, Liu J, Tseng MM, Leung E, Lai KK, Chiu HW (2005) Hongkong International Terminals gains elastic capacity using a data-intensive decision-support system. *Interfaces* 35(1):61–75.

Steenken D, Voss S, Stahlbock R (2004) Container terminal operation and operations research—A classification and literature review. *OR Spectrum* 26(1):3–49.

Upadhyay A, Gu W, Bolia N (2017) Optimal loading of double-stack container trains. *Transportation Res. Part E: Logist. Transportation Rev.* 107:1–22.

Verification Letter

The paper includes anonymous sources. For verification purposes, the editor-in-chief has received a list of these sources.

Amit Upadhyay is an assistant professor at the Vinod Gupta School of Management, Indian Institute of Technology (IIT) Kharagpur, India. After receiving his PhD in operations research from IIT Delhi, India, he did postdoctoral work at the National University of Singapore. He has keen research and consultancy interest in optimization, business analytics, logistics, and transportation planning. His analytics models for transportation and supply chain planning are being used in the industry.