Transportation Research Part E 107 (2017) 1–22

Contents lists available at ScienceDirect

Transportation Research Part E

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

Optimal loading of double-stack container trains

Amit Upadhyay [a,*], Weihua Gu [a], Nomesh Bolia [b]

[a] Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong
[b] Department of Mechanical Engineering, Indian Institute of Technology Delhi, New Delhi, India

ARTICLE INFO

Article history:
Received 27 May 2016
Received in revised form 28 June 2017
Accepted 21 August 2017
Available online 8 September 2017

Keywords:
Container trains
Decision support system
Intermodal transport
Heuristics
Indian Railways

ABSTRACT

We develop a new mathematical model for optimizing the loading of double-stack container trains. We analyze the practical importance of multiple objectives reported in the literature and formulate two new objectives: maximizing profit and minimizing tardiness. The model accounts for containers of different types, weights, and heights, and their feasible loading combinations on a wagon satisfying real operational constraints. The model is solved optimally by CPLEX after exploiting the problem specific properties. A decision support system based on this optimization model has been deployed by a major train operator in India. Numerical cases show that our model can reduce the container haulage cost by about 3%.

1. Introduction

1.1. Background

Intermodal container transport is an efficient mode of freight transport, which is growing globally at 6.1% per year (Container Terminal Foresight, 2014). In India, container trains took up 33% of the total land-based container transport in 2013 (RITES, 2014). Due to this huge demand, double-stack loading of containers on trains has become an important strategy to achieve more cost-efficient rail haulage. Containers on these trains are loaded two stacks high (one above the other) on the wagons as shown in Fig. 1. Compared to the conventional single-stack loading, the use of double-stack trains enables: (i) faster processing of containers at terminals; (ii) less congestion in the rail network due to more containers per wagon; and (iii) more efficient container trains due to less requirement of locomotives, rakes, crew, and fuel consumption per container. The collective economic and environmental gains obtained from the double-stack train operation can be huge. For example, the number of trains required for container movement in North Western and Western Railways in India has decreased by 20–48% due to the double-stacking.

Thanks to the above sizable benefits, many countries are currently planning to build new railways or renovate their existing infrastructures to provide enough vertical clearance for operating double-stack trains. These countries include India, the US, Canada, China and Australia. For example, India is building dedicated freight corridors that have enough vertical clearance under the overhead catenary to operate double-stack trains (DFCCIL, 2016). Thus, double-stack container transport is expected to grow steadily in these countries for the next several years.

* Corresponding author at: Assistant Professor at Vinod Gupta School of Management, Indian Institute of Technology, Kharagpur, India.
E-mail addresses: amitupadhyay85@gmail.com (A. Upadhyay), weihua.gu@polyu.edu.hk (W. Gu), nomesh@mech.iitd.ac.in, nomesh@mech.iitd.ac.in (N. Bolia).

http://dx.doi.org/10.1016/j.tre.2017.08.010
1366-5545/

Fig. 1. A double-stack container train in India.

Container train operator (CTO) and port (or terminal) operator are the two major entities in the container rail transport ecosystem. A CTO is responsible for the container to train assignment, train routing and fleet management in the rail network. The port operator's business is limited to the container stevedoring, storage and ancillary services. Fig. 2 shows the roles of the CTO and port operator in the entire process of container handling at a typical large seaport. Upon arrival, the containers are sorted and stacked in a container yard according to their types and destinations. Before a train is available for loading, the CTO chooses a set of containers from the container inventory at the port (which is termed "pendency") and instructs the port operator to load them on the train. The CTO's selection of containers is based mainly on the objective of maximizing profit, while other factors like timeliness of delivery are also taken into consideration. This set of containers is retrieved from the container yard by the port operator using cranes (e.g. RTGs and RMGs[1]) and internal trucks (trailers). Using another set of cranes on the rail-side, these containers are loaded on the rail wagons while satisfying various technical criteria such as wagon payload limit and safety requirements.

The haulage cost charged by the railway company to the CTOs depends mainly on the containers' type, weight, distance and the train route. To encourage double-stack loading, the railways offers discount on the haulage cost of the containers loaded in the upper-stack of a train; e.g. see Rates Circular (2014) for the detailed cost structure followed by Indian Railways. Considering the haulage cost structure, CTOs seek to optimize the container selection and container-to-wagon assignment for double-stack trains with an objective of minimizing the container rail haulage. This problem, termed as the double-stack container train loading problem (DSLP), has not been adequately studied in the literature.

In the next section, we review the literature related to the DSLP. We define the research problem and its scope in Section 1.3.

1.2. Literature review

The DSLP is complicated because each wagon can have multiple loading pattern options (e.g., one 40-ft container above two 20-fts, or one 40-ft above another 40-ft) given the varied container sizes and heights. However, only a few studies were found to be related to this problem, despite an increase in the research on intermodal freight transport (Bontekoning et al., 2004; Boysen et al., 2013). Among these studies, some relied on Monte Carlo simulation (Jahren and Pacanovsky, 1993; Pacanovsky et al., 1995) and thus were unable to furnish optimal double-stack loading plans. Lai et al. (2008a,b) developed mathematical programs that focused on minimizing aerodynamic resistance of the intermodal trains, but ignored the more important objective of profit maximization and double-stack operational constraints. Other analytical works for double-stack train planning include Lang et al. (2011) and Chih et al. (1990),[2] where heuristic approaches were employed to solve their models. Of note is that they did not consider multiple double-stack loading patterns, container heights, and necessary operational constraints. Hence, to our best knowledge, there is no mathematical model in the literature that can properly account for the realistic complexities and unique operating features of the DSLP, which is necessary for a successful real-world application of optimal double-stack container train loading.

To furnish a comprehensive review of the literature on the DSLP, we further extend our review to include the studies on single-stack train loading that are relevant to the DSLP. Note that we are not intended to provide a comprehensive review of the container rail transport planning literature, since there already exist a number of reviews on that topic; for example, Steenken et al. (2004), Günther and Kim (2005), Kim and Günther (2006), Caris et al. (2008), Meisel (2009), Bierwirth and Meisel (2010), and Boysen et al. (2013).

[1] RTG refers to Rubber Tyred Gantry, and RMG refers to Rail Mounted Gantry.

[2] Chih et al. (1990) focused on network-wide wagon and container distribution at an aggregate level. Their model is not suitable for the DSLP.

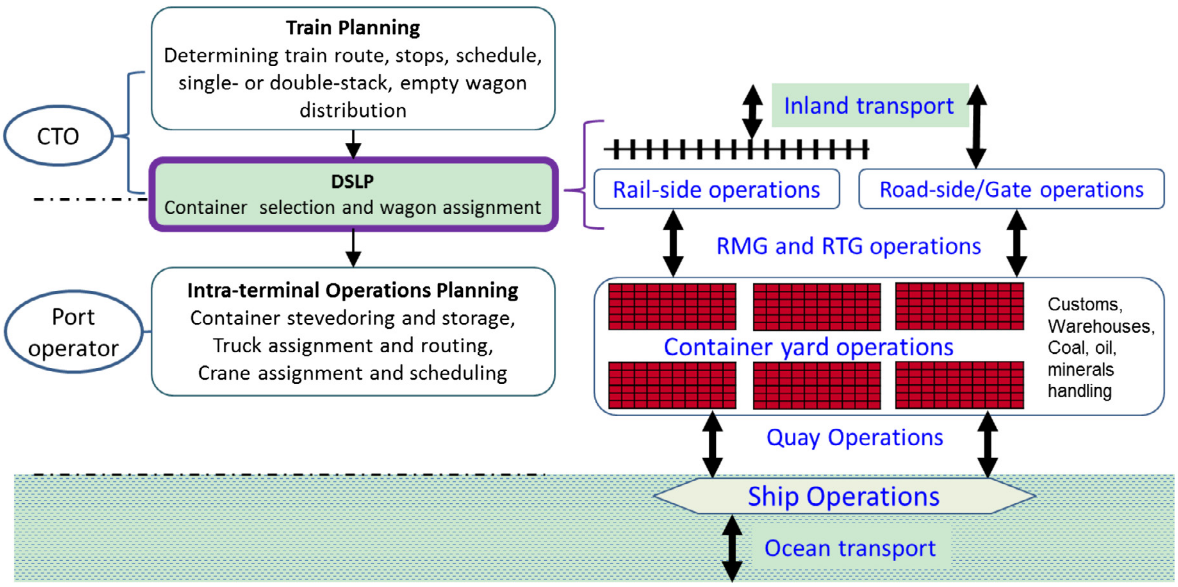

Fig. 2. Container stevedoring and storage at a seaport, and the focus of the DSLP.

The single-stack train studies we have reviewed include Ambrosino and Siri (2015), Corry and Kozan (2006, 2008), Feo and González-Velarde (1995), Bruns and Knust (2012), Bruns et al. (2014), Bostel and Dejax (1998), and Lutter and Werners (2015). These studies formulated binary linear programming and mixed integer programming models for different objectives and sets of constraints; and the models were solved by CPLEX or heuristic algorithms (e.g. simulated annealing, GRASP, tabu-search). In short, these single-stacking works cannot be directly extended to solve the DSLP because the DSLP involves a distinct set of operational constraints such as safety and feasibility of different double-stack loading patterns. These constraints are either absent or unbinding for single-stack train loading.

However, the review of the single-stacking studies can unveil what objectives should be included in the DSLP, since the CTOs have the similar interests for both single-stack and double-stack train operations. To this end, in Table 1, we summarize 11 objectives used in the optimization models of both the single- and double-stacking studies reviewed here. We further surveyed the management of four major CTOs in India, including two multinationals, on their opinions regarding the relative importance of the objectives in real-world operations. All the four companies claimed that haulage cost and timeliness of container delivery are the primary and secondary business goals for planning of container train loading in Indian Railways. Based upon the responses of the practitioners, we mark the relative importance of the 11 objectives in Table 1 in terms of their impacts on the total rail haulage cost and the timeliness of delivery; see columns 5 and 6 in the table.

The first three objectives in Table 1 all target at maximizing the utilization of the train. Among them, maximizing *slot utilization* is the most commonly used objective in the literature, where the slot utilization is defined as the ratio of the total TEUs (twenty-feet equivalent units) loaded and the total loading capacity of the train (measured in TEUs). Clearly, these objectives[3] are most closely aligned with the primary business goal, i.e. reducing the haulage cost. For example, in Indian Railways, the loading of one additional 40-ft container on a 1200-km distance train will result in a cost saving of about INR (Indian Rupee) 37,000 for the CTO. This cost is at least 25 times more than the total cost of crane and truck movements required for moving the same container from the container-yard to the train. Maximizing train utilization is also aligned with the second business goal: timely delivery. Note that the rail haulage of a container may take over 24 h while the intra-terminal movement from the yard to the train takes only a few minutes.

Objective 4 in the table minimizes the aerodynamic resistance[4] of a train, which is equivalent to minimizing the air-gap between containers and/or trailers (Lai et al., 2008a,b). Reducing the aerodynamic resistance will result in less fuel consumption, and thus save the train's fuel cost. However, this objective is more important for a train that carries a diverse mix of different containers, trailers, and swap-bodies. The cost saving from minimizing the aerodynamic resistance is small for trains carrying only standard-sized containers, as admitted in the cited works. Objective 5 aims to reduce the wear on the train's breaking system by shifting the horizontal center of mass towards the locomotive. This will reduce the maintenance cost of the breaking system, but the reduction is marginal as compared to the saving from improving the train's utilization.

[3] Precisely, the third objective, minimizing train length, is not suitable for the container train loading problem discussed in this paper. This is because the decision of the train's length is a part of a more complicated train-sizing problem that depends, among other factors, on the network-wide distributions of empty wagons and the demand. For more details of the train sizing problem formulated at the network scale, please refer to Powell and Carvalho (1998) and Chih et al. (1990).

[4] For a fuller discussion on the aerodynamic resistance of freight trains, please refer to Lai and Barkan (2005) and Li et al. (2014).

Table 1
Comparison of objectives of optimization models for the container train loading problem.

S.N.	Optimization objective	Literature	Target	Impact on haulage		Focus
				Cost	Timeliness	
1	Max slot utilization	Ambrosino and Siri (2015), Bruns and Knust (2012), Bruns et al. (2014), Jahren and Pacanovsky (1993) and Pacanovsky et al. (1995)	Train Utilization	High	High	Train operations
2	Max total TEUs loaded	Lang et al. (2011)				
3	Min train length or number of wagons	Corry and Kozan (2008) and Feo and González-Velarde (1995)				
4	Min total adjusted air-gap length (aerodynamic resistance)	Lai et al. (2008a,b)	Aerodynamic efficiency	Moderate	Low	
5	Min horizontal center of mass of train	Corry and Kozan (2006)	Maintenance	Low	Low	
6	Min vertical center of gravity of train	Lang et al. (2011)	Safety	NA	NA	
7	Min weight difference of 20 ft containers loaded on the same wagon	Lang et al. (2011)	Safety	NA	NA	
8	Min container double-handling inside terminal	Ambrosino and Siri (2015), Corry and Kozan (2006) and Corry and Kozan (2008)	Crane moves	Low	Low	Intra-terminal operations
9	Min total distance travelled by cranes (gantries)	Bostel and Dejax (1998) and Corry and Kozan (2006)	Crane moves	Low	Low	
10	Min trucks movements (truck-km) between yard and train	Ambrosino and Siri (2015), Bruns and Knust (2012), Bruns et al. (2014) and Corry and Kozan (2008)	Truck moves	Low	Low	
11	Min setup time of changing wagons' pin configurations	Bruns and Knust (2012), Bruns et al. (2014) and Corry and Kozan (2008)	Labor cost	Low	Low	

Objectives 6 and 7 were proposed to maximize the stability of the train so as to reduce the risk of wagon derailment. They are not aligned with either the haulage cost reduction or improvement of the timeliness of container delivery. They are better modeled as constraints than objectives because railway guidelines always ensure that the safety of the train is never compromised by any subjective preferences of the CTO and port operator. For example, Indian Railways stipulates a threshold for the maximum vertical center of gravity of every wagon. Further, it prohibits the double-stacking of containers in case of heavy winds.

Unlike objectives 1–7, the last four objectives reflect the concerns of the port operator regarding the efficiency of intra-terminal container handling. Port operators are interested in minimizing crane movements (objectives 8 and 9), truck movements (objective 10), and the setup time for changing the pin-configurations of the wagons[5] (objective 11). These objectives are given a lower importance mainly due to the relatively much lower cost of intra-terminal container movements. Recall that the cost saving for loading an extra container to a train is about 25 times the cost for its intra-terminal movement and loading operations. Thus these objectives are better formulated in a separate intra-terminal operations planning problem since it requires a full consideration of the terminal layout, height of stacks in the container yard, availability and assignment of internal-trucks and cranes; see for example (Murty et al., 2005) and Wang and Meng (2012).[6] This intra-terminal container operations planning problem accepts the results of train and vessel planning problems as the inputs.

1.3. Research scope and overview of our work

Based upon the literature survey presented above and the practical requirements of the CTOs, we define the DSLP as minimizing the rail haulage cost (or equivalently, maximizing the CTO's profit since the revenue associated with the given containers is fixed), while ensuring timely delivery of the containers, by selecting the containers from a candidate set and assigning them to a set of wagons of a train in the feasible loading patterns.

We hereby clarify that the network-wide train planning, including train routing, stops, schedules, and the associated empty wagon distribution are higher-level decisions, and are assumed to be given for the DSLP. These network-wide rail operations cannot be incorporated into the DSLP model because otherwise the model would be too complicated and mathematically intractable. For the same reason, we exclude the intra-terminal operations planning (e.g. optimizing truck and crane movements in the terminal) from the DSLP too, as already explained in the above section. Therefore, the scope of the DSLP is limited to the container selection and container-to-wagon assignment, as shown in Fig. 2.[7]

We further limit the DSLP to model a single train, albeit our proposed formulation can be easily extended to model multiple trains that depart from the same terminal (see Section 4.3). This is because practitioners prefer to optimize the loading

[5] Manual change of the pin-configuration, if required, takes hardly one minute per wagon to complete.
[6] This is also consistent with the status quo of practice; i.e., the port operators do not interfere with the container selection process for vessels and trains.
[7] In the area of railway operations research, it is common to break down a multi-stage, mathematically intractable planning problem into several subproblems and solve them in a sequential manner; see Ahuja et al. (2005).

plan of one train at a time mainly due to the multifarious uncertainties in real-life railway and intra-terminal operations. These uncertainties include changes in: (i) the candidate set of containers available for loading due to new container arrivals and changes in the containers' accessibility in the yard; (ii) container priorities and urgency; (iii) trains' routes, stops, and schedules due to unexpected weather, network congestion, etc.; and (iv) the arrival and departure times of ships, trains and trucks. Thus, an "optimal" multi-train loading plan which does not consider these uncertainties properly will become suboptimal or even infeasible as these uncertainties accumulate over time. In reality, often one train is loaded at a time,[8] and the train's loading plan is finalized only shortly (less than one hour) before the train loading actually starts. Lai et al. (2008b) recommended iterative use of the multi-train optimization model on a rolling horizon for addressing the uncertainties. However, this rolling-horizon optimization approach does not guarantee a better loading plan compared to our proposed solution approach, especially when the uncertainties are large. Further, it is not embraced by the practitioners because the DSLP model, which is more complex than the one proposed by Lai et al. (2008a,b), may take more than an hour to provide an optimal multi-train solution. Thus, in this paper, we develop a deterministic optimization model for a single train so that the above operational uncertainties can be ignored. Further, when planning for the present train, our DSLP model penalizes certain loading patterns to avoid potential adverse effects on the loading of future double-stack trains (see Section 2.3).

In light of the literature review and the research scope defined above, we propose two new objectives for the DSLP: (i) *maximizing profit*, which is a more general objective than *maximizing train utilization*; and (ii) *maximizing total tardiness of selected containers*, which prioritizes the loading of the oldest containers. *Maximizing profit* is chosen as the objective instead of the equivalent *minimizing the haulage cost* for a single train because the latter may result in an empty train (which has the lowest haulage cost). The second objective (maximizing total tardiness) aims at improving customers' satisfaction and minimizing customer grievances regarding the violation of the First-Come-First-Served (FCFS) order. This objective is not always needed. For example, at a rake-surplus terminal, all the containers will be served without delay. In any case, the profit maximization objective always dominates the second one, as is required by the CTOs; see Section 2.3.

Objective 4 in Table 1, minimizing the aerodynamic resistance, is considered as a constraint in our DSLP model based on the following proposition:

Proposition 1. *For any double-stack container train, a near-minimal aerodynamic resistance can be achieved by simply grouping the double-stack loaded wagons in the front of the train (i.e. near the locomotive), and the empty wagons, if any, towards the tail.*

The objective 5, minimizing the horizontal center-of-gravity (HCG), can be considered as the third objective in our DSLP model. However, due to its lower importance and potential conflict with the other objectives, we consider this optional objective through a fast heuristic approach at the post-processing level. The heuristic is based on the following proposition:

Proposition 2. *The minimum HCG of the train can be nearly attained by arranging the wagon loads from the locomotive to the tail of train in a non-increasing order of their total weight.*

By virtue of these simple properties, the objectives 4 and 5 can be addressed to great extents in the DSLP without compromising the two main objectives and without increasing the computation time. These properties were overlooked by previous studies (Corry and Kozan, 2006).

We incorporate practical constraints into the DSLP such as the limits for wagons' payload, vertical center-of-gravity (VCG) and the weight difference between the two 20-ft containers loaded on a wagon, etc. Note too that most previous works have overlooked at least one of these practical constraints.

We propose an exact solution approach to solve the DSLP problem via lexicographic optimization using CPLEX. Exact solution approaches are especially preferred over heuristics given the huge cost savings from loading just one additional container. The approach is able to find an optimal loading plan in practically acceptable computation times, as manifested by our extensive numerical cases. A decision support system based on our DSLP model and the solution approach has been implemented by a major CTO in India. To our best knowledge, this is the first ever real-life implementation of an optimization model for double-stack container train loading.

The rest of the paper is structured as follows. The multi-objective DSLP is formulated in Section 2. The solution approach is presented in Section 3. The model and the solution approach are validated using both real and hypothetical cases in Section 4. Section 5 introduces the decision support system and discusses key practical issues. Insights and future research opportunities are summarized in Section 6.

2. Problem formulation

We first introduce the container types and the feasible loading patterns for the DSLP in Section 2.1, and describe the practical loading constraints in Section 2.2. Then, we furnish the mathematical formulation in Section 2.3.

[8] For example, in case of two trains to be loaded, the loading of one-train at a time expedites the dispatching of the first train (and its containers) and vacates the rail track for a new incoming train. On the other hand, loading of two or more trains simultaneously may result in reduced utilization of the trains, containers, and rail tracks at the terminal.

2.1. Container types and loading patterns

In the real world, four types of containers are commonly allowed for the double-stack loading: 20-ft and 40-ft ISO General (GEN) containers with a height of 8′6″ (2.591 m), and 20-ft and 40-ft High-Cube (HQ) containers with a height of 9′6″ (2.896 m). In Indian Railways, the GEN containers account for about 97% of the double-stack containers and the HQ containers account for the rest. For brevity, we only consider these four types of containers in the DSLP model.

Fig. 3 illustrates the six practically feasible loading patterns for a wagon. Patterns 1–3 are double-stack: (1) one 40-ft container loaded on top of two 20-ft containers; (2) one 40-ft container on top of another 40-ft container; and (3) two 20-ft containers on top of another two 20-ft containers. Note that the pattern with two 20-fts on top of a 40-ft is not feasible because there are no slots (similar to the corner castings) in the middle of the 40-ft container, which are needed to securely lock the 20-ft containers on the top. Further, even pattern 3 is quite rare in the practice (and prohibited in India) because it is much more unstable than patterns 1 and 2. For simplicity, this pattern is also not considered in the presented DSLP model.[9] In addition, patterns 4 and 5 are single-stack, and pattern 6 is an empty wagon. We also ignore the single-stack pattern with only one 20-ft container on a wagon because it is extremely rare (and prohibited in India) due to safety reasons. Finally, the two 20-ft containers in the lower stack in pattern 1 must have the same height; i.e., they have to be both GEN or both HQ containers, which is not necessary in pattern 4.

For the modeling purpose, we label the positions of the 20-ft containers on a wagon by A and B, and the positions of the 40-fts by E and F, as shown in Fig. 3. We denote m as the index of a container's position: $m \in \{A, B, E, F\}$, and j as the index of a loading pattern: $j \in J = \{1, 2, 4, 5, 6\}$.

2.2. Double-stack loading constraints

We consider the following key constraints for the double-stack loading:

(1) Safety constraints. The railways stipulate that the total weight of the containers loaded in the upper stack of a wagon cannot exceed the total weight of those in the lower stack. Further, the VCG of a loaded wagon cannot exceed a given threshold.
In the horizontal direction, if the load on a wagon is heavily biased towards one end of the wagon, the chance of lifting the wheels on the opposite end of the wagon increases (Lopez-Gomez, 1987), which may cause wagon derailment when the train is moving. Thus, the railways require that the weight difference between the 20s loaded on the front and back of every wagon (for patterns 1 and 4) does not exceed a maximum allowable limit.[10]
(2) Minimizing the aerodynamic resistance. We specify the following constraints to achieve this objective:
   (i) The double-stack wagons should appear contiguously on the front side of the train; and they are followed by single-stack wagons, if any.
   (ii) Empty wagons, if any, should be appended contiguously at the end of the train. However, empty wagons cannot coexist with double-stack wagons in the same train.
Fig. 4 illustrates a case where a single-stack wagon is placed between two double-stack ones, which creates an air gap between the two double-stack wagons. In this case, swapping the loading patterns of this single-stack wagon and a rear-side double-stack wagon can reduce the aerodynamic resistance. For the same reason, empty wagons should also be arranged contiguously. In case a double-stack wagon appears together with an empty wagon in a rake, the upper-stack container of the double-stack wagon can be loaded on the empty wagon. This would reduce both the aerodynamic resistance and the VCG.
(3) Minimizing the HCG of the train. To achieve this objective, we can specify that the wagons are loaded in a non-increasing order of their total weight (including the containers), i.e. heavier sets of containers nearer to the locomotive. In most situations, this arrangement of the wagons is consistent with what is required above for minimizing the aerodynamic resistance (e.g., a double-stack wagon is usually heavier than a single-stack one). However, sometimes the two constraints may still conflict. Thus, in the solution approach we choose to relax this constraint, and consider this objective at the post-processing stage through a heuristic approach; see Section 3.4.
(4) Group transport of containers. Some customers require that all their containers, booked through a common shipping bill, be sent together in one train. It helps the consignees to manage the logistics at the destination terminal even if the delivery of the containers is delayed.

2.3. Model formulation

The DSLP is formulated as a binary mathematical program. A complete list of notation is presented in Appendix A. The binary decision variables are defined as follows:

[9] With modest modifications, our DSLP model and solution approach can be applied to cases where pattern 3 is also included.
[10] Note that the railway authority has detailed guidelines for uniform and stable loading of the cargo in every container. Hence, we assume each container to have a uniform density of mass for the modeling purpose.

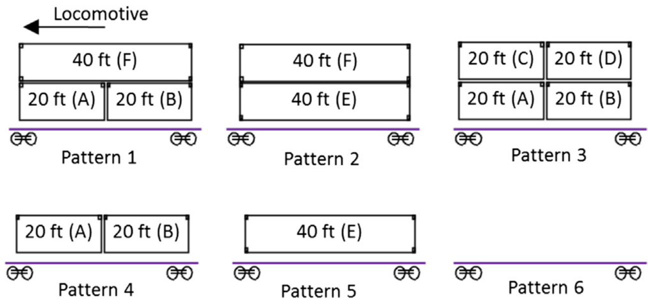

Fig. 3. Double-stack and single-stack container loading patterns.

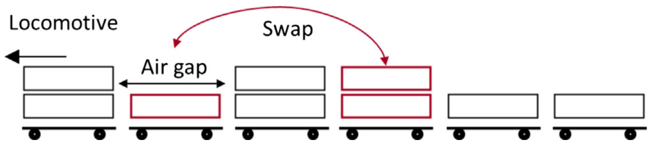

Fig. 4. An example for reducing the aerodynamic resistance.

x_k^j	$= 1$ if wagon $k \in K$ is loaded in pattern $j \in J$, else $x_k^j = 0$; where K is the set of wagons of the train, and k is the wagon index counted from the locomotive.
y_{ik}^m	$= 1$ if a 20-ft container $i \in I_{20}$ is assigned position $m \in \{A, B\}$ on wagon $k \in K$, else $y_{ik}^m = 0$;
z_{ik}^m	$= 1$ if a 40-ft container $i \in I_{40}$ is assigned position $m \in \{E, F\}$ on wagon $k \in K$, else $z_{ik}^m = 0$; where I_{40} and I_{20} are the sets of all the 40-ft and 20-ft containers available for loading on the present train, respectively.
b_s	$= 1$ if all the containers booked through a shipping bill $s \in S$ are assigned to the train, else $b_s = 0$; where S is the set of all the shipping bills with the request of the group transport of containers.

We further define the following intermediate binary variables for the assignment of 20- and 40-ft containers to wagon k:

$$y_{ik} = y_{ik}^A + y_{ik}^B, \ \forall i \in I_{20}, \ k \in K \tag{1}$$

$$z_{ik} = z_{ik}^E + z_{ik}^F, \ \forall i \in I_{40}, \ k \in K \tag{2}$$

These binary variables also ensure that a container can occupy at most one position on a wagon.

The primary objective function, viz. profit maximization, is formulated as follows:

$$\max \pi \equiv \sum_{k \in K} \left[\sum_{i \in I_{20}} P_i^L (y_{ik}^A + y_{ik}^B) + \sum_{i \in I_{40}} \left(\alpha P_i^L z_{ik}^E + P_i^U z_{ik}^F \right) \right] \tag{3}$$

where P_i^L and P_i^U are the profits for assigning contain i to the lower- and upper-stack positions of a wagon, respectively.[11] The parameter α is employed just to penalize the 40-ft containers loaded in the lower stack. Note in the loading patterns 1 and 2 (see Fig. 3) that the upper-stack container must be a 40-ft. Thus, when the 40s are in shortage, we employ a small α to discourage the loading of 40s in the lower stacks (i.e. position E in patterns 2 and 5) so that the 40s are not overused in the current train and the double-stack loading of future trains is not compromised. The α can be any value satisfying the following condition:

$$\begin{cases} 0 < \alpha < \frac{2P_{min}^{20}}{P_{max}^{40}}, & \text{if } \frac{2\bar{N}_{40}^F}{|I_{20}| + 2|I_{40}|} \leqslant \lambda \\ \alpha = 1, & \text{otherwise} \end{cases} \tag{4}$$

where $P_{min}^{20} = \min\{P_i^L | i \in I_{20}\}$, $P_{max}^{40} = \max\{P_i^L | i \in I_{40}\}$, and $\bar{N}_{40}^F$ is the number of 40s that are suitable for loading in the upper stack. Here, λ is a user-defined parameter that indicates the user's perception of the scarcity of 40s. A higher λ means the CTO expects a greater shortage of 40s in the future. Typically λ takes a value between 0.6 and 1. Note that α has to be greater than

[11] These profit parameters are calculated without considering the revenue and cost components that are unrelated to the container train loading plan. Omission of these "irrelevant" components, such as container handling and stuffing costs, maintenance cost, and overheads, will not affect the optimality of the DSLP solution. Also note that if we set P_i^L and P_i^U equal to 1 for 20s and 2 for 40s, then the objective function (3) reduces to the maximization of slot utilization, which is the commonly used objective in the literature.

zero, because an available 40-ft should always be assigned to any empty wagon that may exist in the present train, even if the CTO expects a shortage of 40s in the future. Also, note that α should be set to 1 if there is no shortage of 40s or if the loading pattern 3 is allowed.

The secondary objective, maximizing the total tardiness of selected containers at the time when the train is formed, is formulated as follows:

$$\max \tau \equiv \sum_{k\in K}\left(\sum_{i\in I_{20}\setminus I_{20}^C} T_i y_{ik} + \sum_{i\in I_{40}\setminus I_{40}^C} T_i z_{ik}\right) \tag{5}$$

where T_i is the tardiness of container i, which is measured as the difference between the time of train formation and a reference time of container (e.g. In India, the reference time is the time of transshipment permit for the import containers). The I_{20}^C and I_{40}^C are sets of compulsory 20- and 40-ft containers, respectively, which must be assigned to the present train. Note that the CTOs often omit this objective during off-peak demand seasons.

The DSLP model is formulated as follows.

$$\max \pi$$

$$\text{and} \ \max \tau \ (\text{optional})$$

subject to

$$\sum_{j\in J} x_k^j = 1, \ \forall k \in K \tag{6}$$

$$2\sum_{j\in\{1,2\}} x_k^j + \sum_{j\in\{4,5\}} x_k^j \geqslant 2\sum_{j\in\{1,2\}} x_{k+1}^j + \sum_{j\in\{4,5\}} x_{k+1}^j, \ \forall k \in \{1,\ldots,|K|-1\} \tag{7}$$

$$\sum_{i\in I_{20}} y_{ik}^m - x_k^1 - x_k^4 = 0, \ \forall k \in K, m \in \{A,B\} \tag{8}$$

$$\sum_{i\in I_{40}} z_{ik}^E - x_k^2 - x_k^5 = 0, \ \forall k \in K \tag{9}$$

$$\sum_{i\in I_{40}} z_{ik}^F - x_k^1 - x_k^2 = 0, \ \forall k \in K \tag{10}$$

$$\sum_{k\in K} y_{ik} \leqslant 1, \ \forall i \in I_{20} \tag{11}$$

$$\sum_{k\in K} z_{ik} \leqslant 1, \ \forall i \in I_{40} \tag{12}$$

$$\sum_{i\in I_{20}} W_i y_{ik} + \sum_{i\in I_{40}} W_i z_{ik} \leqslant G_k, \ \forall k \in K \tag{13}$$

$$\sum_{i\in I_{20}} W_i y_{ik} + \sum_{i\in I_{40}} W_i z_{ik}^E \geqslant \sum_{i\in I_{40}} W_i z_{ik}^F, \ \forall k \in K \tag{14}$$

$$\left|\sum_{i\in I_{20}} W_i (y_{ik}^A - y_{ik}^B)\right| \leqslant \Delta_k, \ \forall k \in K \tag{15}$$

$$\sum_{i\in I_{20}} H_i (y_{ik}^A - y_{ik}^B) \geqslant x_k^1 - 1, \ \forall k \in K \tag{16}$$

$$\sum_{i\in I_{20}} H_i (y_{ik}^A - y_{ik}^B) \leqslant 1 - x_k^1, \ \forall k \in K \tag{17}$$

$$CG_k \leqslant R_k, \ \forall k \in K \tag{18}$$

$$\sum_{k\in K}\left(\sum_{i\in I_{20}^s} y_{ik} + \sum_{i\in I_{40}^s} z_{ik}\right) = b_s\left(|I_{40}^s| + |I_{20}^s|\right), \ \forall s \in S \tag{19}$$

$$\sum_{k \in K} y_{ik} = 1, \ \forall i \in I_{20}^{C} \tag{20}$$

$$\sum_{k \in K} z_{ik} = 1, \ \forall i \in I_{40}^{C} \tag{21}$$

$$x_k^j, y_{ik}^j, z_{ik}^j, b_s \in \{0, 1\} \tag{22}$$

Constraints (6) stipulate that a wagon can assume exactly one of the loading patterns. Constraints (7) ensure that the double-stack patterns are assigned to contiguous wagons towards the front and empty wagons, if any, are contiguous towards the tail of the train; see the 2nd point in Section 2.2. Constraints (8)-(10) ensure that each wagon is properly loaded according to the assigned loading pattern; for example, constraints (8) stipulate that two 20s must occupy positions A and B respectively if loading pattern 1 or 4 is applied. Constraints (11) and (12) specify that each container can be assigned to at most one position on at most one wagon. Constraints (13) ensure that the total weight of the containers loaded on wagon k does not exceed its payload limit G_k, where W_i denotes the gross weight of container $i \in I$.

Constraints (14) ensure that the weight of the upper-stack container does not exceed the total weight of the lower-stack container(s) on the same wagon. Constraints (15) require that the weight difference between the two 20s loaded on each wagon in patterns 1 and 4 does not exceed a specified limit, Δ_k. Constraints (16) and (17) specify that the two 20s loaded on a wagon in pattern 1 should be of the same type (either GEN or HQ), where H_i is the height of container i. Recall that the height difference between the two container types is one foot.

Constraints (18) specify that the VCG of wagon k, CG_k, must not exceed a specified CG threshold denoted by R_k, where CG_k is calculated by the following equation, assuming that the CG of each container is at its geometrical center:

$$CG_k = \frac{\bar{R}_k \bar{W}_k + \left(\bar{H}_k + \frac{H_k^L}{2}\right) W_k^L + \left(\bar{H}_k + H_k^L + H^o + \frac{H_k^U}{2}\right) W_k^U}{\bar{W}_k + W_k^L + W_k^U}, \ \forall k \in K \tag{23}$$

where $\bar{W}_k$ and $\bar{R}_k$ denote the tare weight and the CG height, respectively, of the empty wagon k; $\bar{H}_k$ the height of the wagon platform measured from the rail surface; H_k^L and W_k^L the height and (total) weight of the lower-stack container(s); H_k^U and W_k^U the height and weight of the upper-stack container; and H^o the height of the interbox twist-lock used for locking the upper-stack container on the top of the lower-stack container(s). The weight of the twist-lock is small and thus ignored here. The H_k^L, W_k^L, H_k^U and W_k^U are calculated by the following equations:

$$W_k^L = \sum_{i \in I_{20}} W_i y_{ik} + \sum_{i \in I_{40}} W_i z_{ik}^E, \ \forall k \in K \tag{24}$$

$$W_k^U = \sum_{i \in I_{40}} W_i z_{ik}^F, \ \forall k \in K \tag{25}$$

$$H_k^L = \frac{1}{2} \sum_{i \in I_{20}} H_i y_{ik} + \sum_{i \in I_{40}} H_i z_{ik}^E, \ \forall k \in K \tag{26}$$

$$H_k^U = \sum_{i \in I_{40}} H_i z_{ik}^F, \ \forall k \in K \tag{27}$$

Constraints (19) guarantee that all the containers of a particular shipping bill $s \in S$ are sent together in one train. Here I_{20}^s and I_{40}^s denote the sets of 20-ft and 40-ft containers belonging to s, respectively; and $|I_{40}^s| + |I_{20}^s|$ is the total number of containers booked in s. Finally, constraints (20) and (21) ensure that the sets of compulsory containers, denoted by I_{20}^C and I_{40}^C, are assigned to the train.

3. Solution approach

We have a priori preference information from the CTOs that we should never compromise the primary objective (3) for the sake of the secondary objective (5). Further, since the train loading plan is required within 10–15 min,[12] we use the *lexicographic optimization* approach that is widely used for similar multi-objective problems.

We first discuss how to convert the nonlinear DSLP program formulated in Section 2 (which is difficult to solve) to a linear one (Section 3.1), and pre-processing steps needed to further simplify the program (Section 3.2). We then employ the lexicographic optimization method to find an optimal DSLP solution (Section 3.3). Finally a heuristic algorithm is developed to post-process the solution for minimizing the HCG of the train (Section 3.4).

[12] Generating a Pareto frontier for the two objectives is neither required by the CTOs nor possible in the short time available.

3.1. Problem linearization

Note that the DSLP model becomes linear if we can linearize or remove the balanced-loading constraints (15) and the VCG constraints (18).

First, constraints (15) can be replaced by the following linear constraints:

$$\sum_{i \in I_{20}} W_i (y_{ik}^A - y_{ik}^B) \leqslant \Delta_k, \ \forall k \in K \tag{28}$$

$$\sum_{i \in I_{20}} W_i (y_{ik}^A - y_{ik}^B) \geqslant -\Delta_k, \ \forall k \in K \tag{29}$$

However, this can be simplified further. Note that by swapping the two 20s loaded in positions A and B of any wagon in an optimal loading plan, we create *multiple optima*. The number of multiple solutions can be safely reduced to half by stipulating that the weight of the container loaded in position A is never less than that in position B of the same wagon. Hence, (29) can be further replaced by the following constraint:

$$\sum_{i \in I_{20}} W_i y_{ik}^A \geqslant \sum_{i \in I_{20}} W_i y_{ik}^B, \ \forall k \in K \tag{30}$$

These symmetry-breaking constraints (30) will reduce the size of the search space by eliminating symmetric solutions. Practically, it also helps in reducing the HCG of the train.

Regarding the VCG constraints (18), first, we argue that these constraints are redundant for Indian Railways, and then discuss the general case. For example, for a single-stack wagon k, CG_k is less than $\bar{H}_k + \frac{\hat{H}}{2}$, where $\hat{H}$ is the largest possible container height. Real data show that for the flat wagons commonly used in India, $\bar{H}_k = 1.009$ m and $R_k = 3.139$ m (RDSO Drawing, 2013). Considering that $\hat{H} = H_{HQ} = 2.896$ m (for HQ containers), it is easy to verify that $CG_k < \bar{H}_k + \frac{\hat{H}}{2} < R_k$. Further, we derive a general condition to assess the redundancy of the VCG constraints for double-stack wagons, which can be modified for single-stack wagons as well. We start by presenting the following lemma.

Lemma 1. *In presence of the payload limit constraint (13) and the safety constraint (14), the VCG constraint (18) is redundant for wagon k if the following condition is satisfied:*

$$\frac{G_k \left(\bar{H}_k + \hat{H} + \frac{H^o}{2} \right) + \bar{W}_k \bar{R}_k}{G_k + \bar{W}_k} \leqslant R_k \tag{31}$$

Proof. For a wagon k loaded in a double-stack pattern, it is clear from the VCG formula (23) that for the VCG to reach its maximum, the heights of both the upper- and lower-stack containers should be maximum (i.e. $H_k^L = H_k^U = \hat{H}$) and W_k^U should also attain its maximal value (i.e., $W_k^U = W_k^L = \frac{G_k}{2}$ due to the constraints (13) and (14)). Plugging the above conditions into the definition of VCG (23), we have that the maximum CG_k is equal to the left-hand-side of (31). □

The VCG constraints can be dropped for all the wagons that satisfy (31). Specifically for the flat wagons used by Indian Railways, we have $G_k = 61$ tonnes, $\bar{W}_k = 19.1$ tonnes, $\bar{R}_k = 0.551$ m, $H^O = 0.03$ m, $\hat{H} = 2.896$ m, $\bar{H}_k = 1.009$ m, and $R_k = 3.139$ m (RDSO Drawing, 2013). Calculation shows that the condition (31) is always satisfied, and thus constraints (18) can be dropped.

Finally, in case there are still wagons that do not satisfy (31), the VCG constraints (18) may be binding. For those wagons, (18) can be replaced with four linear constraints presented in (32)-(35). These constraints are based upon a finer classification of double-stack loading patterns, illustrated in Fig. 5, that considers different heights of the lower-stack containers. Note that the different heights of the upper-stack container is not an issue here.

$$M_1(1 - x_k^{1a} - x_k^{2a}) + R_k \left(\bar{W}_k + \sum_{i \in I_{20}} W_i y_{ik} + \sum_{i \in I_{40}} W_i z_{ik} \right) \geqslant \bar{R}_k \bar{W}_k + \left(\bar{H}_k + \frac{H_{GEN}}{2} \right) \left(\sum_{i \in I_{20}} W_i y_{ik} + \sum_{i \in I_{40}} W_i z_{ik}^E \right) + \sum_{i \in I_{40}} \left(\bar{H}_k + H_{GEN} + H^o + \frac{H_i}{2} \right) W_i z_{ik}^F, \ \forall k \in K \tag{32}$$

$$M_2(1 - x_k^{1b} - x_k^{2b}) + R_k \left(\bar{W}_k + \sum_{i \in I_{20}} W_i y_{ik} + \sum_{i \in I_{40}} W_i z_{ik} \right) \geqslant \bar{R}_k \bar{W}_k + \left(\bar{H}_k + \frac{H_{HQ}}{2} \right) \left(\sum_{i \in I_{20}} W_i y_{ik} + \sum_{i \in I_{40}} W_i z_{ik}^E \right) + \sum_{i \in I_{40}} \left(\bar{H}_k + H_{HQ} + H^o + \frac{H_i}{2} \right) W_i z_{ik}^F, \ \forall k \in K \tag{33}$$

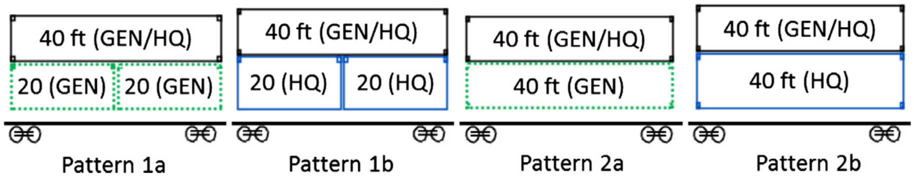

Fig. 5. Double-stack loading patterns considering variable heights of lower-stack containers.

$$x_k^{1a} + x_k^{1b} = x_k^1 \tag{34}$$

$$x_k^{2a} + x_k^{2b} = x_k^2 \tag{35}$$

where $x_k^{1a}, x_k^{1b}, x_k^{2a}, x_k^{2b}$ are the new binary loading pattern variables (which will replace the variables x_k^1, x_k^2) indicating whether wagon k is loaded in double-stack pattern 1a–2b; and M_1 and M_2 are big constants whose minimum values are given by (36) and (37), respectively. We recommend these values because larger values will increase the computation time due to the resulting weaker LP relaxation.

$$M_1^{min} = (\bar{R}_k - R_k)\bar{W}_k + \left(\bar{H}_k + \frac{3H_{GEN}}{4} + \frac{H^o}{2} + \frac{H_{HQ}}{4} - R_k\right)G_k \tag{36}$$

$$M_2^{min} = (\bar{R}_k - R_k)\bar{W}_k + \left(\bar{H}_k + H_{HQ} + \frac{H^o}{2} - R_k\right)G_k \tag{37}$$

With the above linearization, now the program can be solved by CPLEX. However, the computation time for obtaining an optimal solution for any real-life instance is still too long. Typically, there can be about one thousand candidate containers, and the maximum train size is about 30 wagons in Europe and 45 wagons in India, which means the DSLP model can have about 100,000 binary variables ($5|K| + 2|K||I_{20}| + 2|K||I_{40}|$). Numerical experiments show that CPLEX can take about 1–6 h to find an optimal solution, even if constraints (18) are relaxed. To further reduce the computation time, we next explore valid inequalities and pre-processing steps.

3.2. Pre-processing and valid inequalities

First, the double-stack loading is needed only if the total number of TEUs in the input set of containers (I) is greater than the maximum number of TEUs that can be loaded in the single stack train, i.e. $2|K|$. Similarly, an empty wagon exists only if the total number of TEUs is less than $2|K|$. These necessary conditions can be used for preprocessing the decision variables as follows.

$$\text{If } |I_{20}| + 2|I_{40}| \leqslant 2|K|, \begin{cases} z_{ik}^F = 0, \ \forall k \in K, i \in I_{40} \\ x_k^j = 0, \ \forall k \in K, j \in \{1,2\} \end{cases} \tag{38}$$

$$\text{If } |I_{20}| + 2|I_{40}| \geqslant 2|K|, x_k^6 = 0, \ \forall k \in K \tag{39}$$

Second, following constraints (13) and (14), the weight of a 40-ft container loaded in the upper stack of wagon k cannot exceed $\frac{G_k}{2}$; i.e.,

$$z_{ik}^F = 0, \ \forall k \in K, i \in I_{40} | W_i \geqslant \frac{G_k}{2} \tag{40}$$

Moreover, the payload constraints (13) further infer that there is an upper limit for the total weight of the 20-ft containers loaded on a wagon in pattern 1:

$$\sum_{i \in I_{20}} \sum_{m \in \{A,B\}} W_i y_{ik}^m \leqslant G_k - W_{40}^{min} x_k^1, \ \forall k \in K \tag{41}$$

where $W_{40}^{min} = \min\{W_i | i \in I_{40}\}$. Adding the above valid inequalities to the DSLP will further reduce the computation cost.

For the same reason, the safety constraints (14) are replaced with (42) and (43) as follows:

$$\sum_{i \in I_{20}} \sum_{m \in \{A,B\}} W_i y_{ik}^m \geqslant M_3(x_k^1 - 1) + \sum_{i \in I_{40}} W_i z_{ik}^F, \ \forall k \in K \tag{42}$$

$$\sum_{i\in I_{40}} W_i z_{ik}^E \geqslant M_3(x_k^2 - 1) + \sum_{i\in I_{40}} W_i z_{ik}^F,\ \forall k \in K \tag{43}$$

where M_3 is a big constant, whose minimum value is $\frac{C_k}{2}$.

With the above pre-processing and valid inequalities added, the computation time of the DSLP is reduced to about 10 min, which is acceptable for real-world implementation.

3.3. Lexicographic optimization

In general, given a multi-objective program with a sequence of objective functions: $\{f_i, i = 1, \ldots, n\}$, a lexicographically optimal solution is obtained by first optimizing f_1, next optimizing f_2 without compromising the optimal value of objective function f_1, and so on; i.e., for each $i = 2, \ldots, n$, we optimize f_i under the constraining condition that the optimal values obtained in previous steps for $f_1, f_2, \ldots, f_{i-1}$ are not compromised. For more details on the theory of lexicographic optimization, please refer to Rentmeesters et al. (1996).

For the DSLP, we solve a similar sequence of single-objective optimization problems, where the objectives are selected, one at a time, in the order of their importance. In the first stage, we optimize only objective (3) to obtain the optimal train-loading plan. If objective (5) is also desired by the CTO, then we optimize it in the second stage without compromising the optimal profit.[13] The details are as follows:

Step 1. The DSLP with only the primary objective (3) is solved using CPLEX. The resulting optimal loading plan is denoted by δ_1, and the corresponding maximum profit by π_0. Note that in the case of a large candidate container set, multiple optima may exist with the same profit value π_0.

Step 2. We choose a solution from the potential multiple optima of Step 1 by maximizing objective (5). To this end, we fix the values of the loading pattern variables x_k^j $(\forall k \in K, j \in J)$ to be equal to their values in solution δ_1, so that the double-stack loading of the next train is not unduly compromised and the number of decision variables at this step is also reduced. With the following constraint added to the original DSLP model, and a warm-start using solution δ_1, the DSLP is solved with objective (5) only. The new constraint ensures that the optimal profit value is not compromised.

$$\sum_{k\in K}\left[\sum_{i\in I_{20}} P_i^L(y_{ik}^A + y_{ik}^B) + \sum_{i\in I_{40}}\left(\alpha P_i^L z_{ik}^E + P_i^U z_{ik}^F\right)\right] \geqslant \pi_0 \tag{44}$$

The resulting optimal loading plan, denoted by δ_2, ensures the global optimality of the profit. It is also Pareto optimal as none of the two objectives can be improved further without degrading the other one.

3.4. Post-processing for minimizing the HCG of the train

In case, a CTO also desires to minimize the HCG of the train, it can be added as the third step of the above lexicographic method. However, due to the computation time constraints, we propose the following heuristic algorithm to minimize the HCG at the post-processing level. This heuristic begins with the optimal solution obtained in the previous section and takes hardly a second to run.

In the given DSLP solution (δ_2), we denote the set of containers that are assigned to wagon k by V_k, and the set of wagons loaded in single- and double-stack loading patterns by K_1 and K_2 respectively. The heuristic algorithm reassigns the container sets V_k for all the$k \in K_i$ $(i = 1, 2)$ to all the wagons in K_i, such that the containers' positions (i.e., A, B, E, or F) within each set remain unchanged. Only the container-set-to-wagon assignments are changed within K_1 and within K_2. The HCG can be minimized by reassigning these sets of containers in a non-increasing order of their total weight; i.e., a heavier set of containers is assigned to a wagon nearer to the locomotive if the payload limit constraints (13) are not violated. Note that this approach will not compromise objectives (3) and (5) because they do not depend on the container-set-to-wagon assignments. The reassignment will also not change whether a wagon is single- or double-stack loaded, so constraints (7), which ensure the near-minimal aerodynamic resistance, are not violated. One could easily verify that other constraints of the DSLP are not violated too. The details of the greedy heuristic are presented below.

[13] We have also experimented with the alternative weighted sum method, i.e. converting the DSLP to an equivalent single-objective optimization problem by assigning suitable weights to both the objectives. However, empirically, we found it to take more computation time than the lexicographic method.

Algorithm 1: Greedy heuristic for minimizing the HCG

Begin
Sort the sets of containers ($V_k, \ \forall k \in K_2$) in the descending order of their total weight
Set wagon_status = empty $\forall k \in K$
For ($j = 1$ to $|K_2|$){// for the sets of double-stack containers
  For ($k = 1$ to $|K_2|$) {// for the sets of double-stack wagons
    if wagon_status(k) = empty
      if $G_k \geqslant$ total weight of V_j[1]
        Assign V_j to wagon k;
        wagon_status = loaded;
        BREAK
      end-if
    end-if
      }
}
Sort the sets of containers ($V_k, \ \forall k \in K_1$) in the descending order of their total weight
For ($k = |K_2| + 1$ to $|K_2| + |K_1|$){// for the sets of single-stack containers
  Assign V_k to wagon k;
  wagon_status = loaded;
}
Print the revised container loading plan
End

[1] For a more general case, where VCG constraints (18) may be binding, another if-condition can be used here to ensure that constraints (18) are not violated.

4. Model applications

The DSLP model and solution approach described above have been implemented in the real practice for a major CTO in India (henceforth referred by CTOX). To validate the applicability and effectiveness of our model and solution approach, we first apply it to a number of historical cases of CTOX in which the train loading plans were manually determined. Comparisons between the actual, manually planned results and the optimal results generated by the DSLP are furnished in Section 4.1. More cases generated using real data with varied key parameters are examined in Section 4.2. A special case of simultaneously planning multiple trains using our DSLP model and solution approach is discussed in Section 4.3. The effect of the value of α on the performance of the DSLP is explored in Section 4.3.1.

4.1. Validation of the DSLP with real cases

All the numerical cases discussed in this section are based upon a simple rail network shown in Fig. 6, which represents part of the real rail network in India. The rail network in Fig. 6 has three terminals: P, Q, and R. Only the link between Q and R (plotted as the double lines) allows double-stack trains. For a given demand set, CTOX decides the routing of the empty, single-stack and double-stack trains, which is an input for the DSLP. Fig. 6 illustrates an example routing plan for transporting 270 TEUs from R to P. This plan consists of a single-stack train carrying 90 TEUs dispatched directly from R to P, and a double-stack train carrying the remaining 180 TEUs from R to Q, plus two single-stack trains to transport those 180 TEUS from Q to P.

To be conservative, we limit the set of candidate containers of the DSLP, I, to only those containers that were actually loaded onto the trains in a given routing plan. Take the routing plan in Fig. 6 as an example, we assume the set of candidate containers, I, includes only the 270 TEUs to be transported from R to P. Namely, for each case studied in this section, we assume the smallest possible I. Note that in reality, a terminal often holds a much larger pendency of containers that are

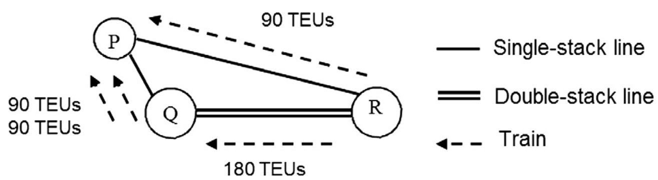

Fig. 6. An illustrative small network with single- and double-stack rail lines.

ready to be loaded on a train, and a larger I would usually lead to a higher profit obtained from the DSLP. Still, considerable profit increases are achieved by optimally reassigning the containers I to the wagons of the given trains, as we shall see shortly.

In this section, we present the results for 10 select real instances that involve 23 actual trains in total. Each of the 10 instances has a fixed routing plan, which includes one double-stack train and 0-to-5 single-stack trains dispatched on the same day. Each train has 45 wagons ($|K| = 45$) having the same DSLP parameters (although different wagon designs), with $G_k = 61$ tonnes, and $\Delta_k = 20$ tonnes $\forall k$. Thus, a maximum of 180 TEUs can be loaded on a double-stack train. The VCG constraints are not binding because inequalities (31) are always satisfied, as explained in Section 3.1. We further assume $|I_{20}^C| = |I_{40}^C| = |S| = 0$. While the container-specific parameters T_i, W_i, H_i are too voluminous to mention in the paper, the details of parameters P_i^L and P_i^U are available on the website of the CTOs[14] and in the Rates Circular (2014). We choose $\alpha = 1$ to ignore any potential shortage of the 40s because all the candidate containers will be assigned to the same trains.

Other key parameters and the results of the 10 instances are summarized in Table 2, where the numbers of candidate 40s and 20s for each OD pair (as specified in the parentheses) are presented in columns 2 and 3, respectively. The total tonnage (in tonne) of the containers is mentioned in column 4. The train routing plan is furnished in columns 5 (for the double-stack train) and 6 (for the single-stack trains if any). The number of TEUs transported by each train is given in the parentheses. Columns 7, 8, and 9 show the total profit, total tardiness (τ, as the secondary objective), and the HCG of the actual loading plan, respectively. Note that the total profit is summed up for all the trains in each instance, while the total tardiness and the HCG are calculated for the double-stack train only. They are compared against columns 10, 14, and 15, respectively, for the total profit, tardiness, and HCG of the optimal DSLP loading plan. In addition, columns 11 and 12 show the profit increase in Indian Rupees and the percentage increase, respectively, if our DSLP model was used instead of the manual planning. Column 13 displays the computation time (via CPLEX 12.6.2 installed on an Intel® Core™ i7 4790 CPU, 16 GB RAM, 3.60-GHz system) for each instance.

Details of the 10 instances are discussed as follows. In instance 1, 6 trains were used to transport 448 TEUs in total on the same day: one double-stack train from R to Q plus two connecting single-stack trains from Q to P, and three single-stack trains directly from R to P. In instance 2, 5 trains (one less direct train from R to P as compared to instance 1) are used to transport 356 TEUs. Since terminal R is *rake-deficit* in both instances, almost all the trains are loaded to their full capacity. For these instances, the DSLP selects the containers to be loaded in the double-stack train, and the remaining containers are assigned to the single-stack trains. The results of both instances show that the optimal double-stack loading plan increases the profit considerably by INR 113,204 (1.67%) for instance 1, and INR 107,336 (2.23%) for instance 2. Note that 1% increase in the total profit means an annual increase of hundreds of million INR for a big CTO in India.

The total tardiness shown in columns 8 and 14 are for illustrative purpose only. In fact, the total tardiness for all the containers should be equal between the actual and optimal loading plans because all the containers are assigned in both the plans. However, it is noteworthy for instance 2 that the total tardiness of the double-stack train actually decreases (from 1579 to 1026 container-days) after optimizing the loading plan. This is because the primary objective of profit maximization, having the absolute dominance, may conflict with the tardiness maximization objective. Finally, comparison of the HCG values in columns 9 and 15 (in the unit of wagons because all the wagons have almost the same length) manifests the effectiveness of the heuristic algorithm for minimizing the HCG of the train.

Instances 3 and 4 each consists of two OD pairs: PR and QR. In the real situation for these instances, both terminals P and Q are *rake-surplus*; i.e., they have more inbound container train traffic than outbound. Hence most trains of the two instances are not fully loaded. It also implies that the secondary objective can be dropped because there are always enough trains to transport all the containers held at the terminal. Therefore, the DSLP is solved only for the primary objective for these two instances. For each instance, the DSLP is first applied to select 90 TEUs at terminal P to load the single-stack train from P to R.[15] The remaining PR containers are assigned to another single-stack train from P to Q. Then the DSLP is applied again to optimize the loading plan of the double-stack train from Q to R. We again observe significant profit increases (over 3%) for both instances. The computation times shown in column 13 (for solving the double-stack train problem only) are less than half a minute. And the HCG's are again reduced by the heuristic at the post-processing stage.

The last six instances each has a single OD pair, which is served by a double-stack train. For these instances, as all the candidate containers will be loaded on the same train, the benefit of the DSLP would only come from the optimal reassignment of the containers to the upper- and lower-stacks of the wagons. Still, significant profit increases (about INR 60,000) are observed for all the six instances. Again, the total tardiness is irrelevant and the HCG is reduced for these instances.

Overall, the above instances show that the DSLP yields a sizable profit growth of about 3% as compared to the manual planning results. Similar profit increases were observed for numerous other instances, which are not shown here in the interest of brevity. To understand how conservative our estimate for the DSLP benefit are, note that many trains originating from a rake-deficit terminal R in Table 2 are not fully loaded. This is due to the potential inefficiency and human errors in the loading plans generated through the cumbersome manual planning process. Thus, still greater profit can be expected from the

[14] http://www.actoindia.org/members/.

[15] Note that the proposed DSLP model and solution approach can also be used for optimizing the loading plan of a single-stack train, which usually takes just a few seconds via CPLEX.

Table 2
Comparison of actual and optimal loading plans.

S.N.	Set of candidate containers			Train plan		Actual loading plan			Optimal loading plan					
	# 40s (OD)	# 20s (OD)	Total tonnage	Double-stack train (TEUs)	Single-stack train (TEUs)	Total profit (INR)	Double-stack train		Total profit (INR)	Profit increase		Runtime (s)	Double-stack train	
							τ (container-days)	HCG		(INR)	%		τ (container-days)	HCG
1	135 (RP)	178 (RP)	6212	RQ (178)	RP (90) RP (90) RP (90) QP (90) QP (88)	6,789,874	1745	22.91	6,903,078	113,204	1.67	59	1826	20.14
2	116 (RP)	124 (RP)	4582	RQ (176)	RP (90) RP (90) QP (90) QP (86)	4,803,213	1579	22.77	4,910,549	107,336	2.23	61	1026	19.83
3	50 (PR) 32 (QR)	18 (PR) 54 (QR)	3991	QR (146)	PR (90) PQ (28)	3,131,538	NA	22.66	3,228,651	97,113	3.10	29	NA	17.82
4	51 (PR) 13 (QR)	80 (QR)	3224	QR (118)	PR (90) PQ (12)	2,630,496	NA	21.82	2,714,952	84,456	3.21	22	NA	19.07
5	88 (RQ)	4 (RQ)	1950	RQ (180)	NA	1,254,917	NA	21.11	1,296,340	41,423	3.30	39	NA	19.11
6	74 (RQ)	28 (RQ)	2123	i. RQ (176)	NA	1,224,591	NA	23.24	1,280,307	55,716	4.55	80	NA	19.78
7	74 (RQ)	30 (RQ)	1982	i. RQ (178)	NA	1,185,254	NA	22.97	1,262,794	77,540	6.54	106	NA	18.22
8	84 (RQ)	4 (RQ)	1666	i. RQ (172)	NA	1,153,103	NA	21.56	1,230,643	77,540	6.72	134	NA	18.06
9	88 (RQ)	0 (RQ)	1589	i. RQ (176)	NA	1,114,890	NA	22.52	1,170,606	55,716	5.00	45	NA	18.17
10	59 (RQ)	60 (RQ)	2255	i. RQ (178)	NA	1,379,050	NA	23.19	1,445,293	66,243	4.80	16	NA	19.92

Table 3
Randomly generated large instances.

	Inputs					Optimal loading plan			
S.N.	Total TEUs	# 20s	# 40s	G_k(tonnes)	Δ_k(tonnes)	Tonnage loaded	Total TEUs	Total profit (INR)	Runtime (s)
1–12	>1400	–	–	61–78	10–20	–	360	–	<60
13	720	408	156	78	20	4809.4	360	2,265,198	15
14					17	4901.8	360	2,265,198	9
15					14	4865.4	360	2,265,198	15
16					10	4837.8	360	2,265,198	37
17	720	408	156	68	20	4574.8	360	2,126,669	19
18					17	4674.3	360	2,161,946	17
19					14	4863.3	360	2,217,440	38
20					10	4710.4	360	2,155,938	49
21	720	408	156	61	20	4574.8	360	2,119,592	20
22					17	4666.4	360	2,197,504	17
23					14	4863.3	360	2,315,024	27
24					10	4224.1	360	2,144,528	32
25	360	148	106	78	20	4603.1	360	2,156,444	10
26					17	4603.1	360	2,156,444	19
27					14	4603.1	360	2,156,444	10
28					10	4603.1	360	2,156,444	30
29	360	148	106	68	20	4603.1	360	2,156,444	26
30					17	4603.1	360	2,156,444	20
31					14	4603.1	360	2,156,444	13
32					10	4603.1	360	2,156,444	99
33	360	148	106	61	20	4603.1	360	2,156,444	90
34					17	4603.1	360	2,156,444	303
35					14	4603.1	360	2,156,444	501
36					10	4603.1	360	2,156,444	619
37–48	<200	–	–	61–78	10–20	–	<200	–	<60

DSLP, ensuring optimal train utilization, due to the loading of a few extra containers. Recall that loading one additional 40-ft container on a train would result in a profit increase of about INR 37,000.

Additionally, the DSLP reduces the HCG of a double-stack train by 10% or more, and ensures that the oldest containers are loaded first if the profit is not compromised. The aerodynamic resistance of the train is also reduced to near minimum, thanks to constraints (7). Finally, the CPLEX computation time is within 5 min for all the 10 instances, which is quite satisfactory for real-world implementation. We next analyze the performance of the DSLP w.r.t. key operating parameters including G_k, Δ_k, and $|I|$.

4.2. Parametric case studies

Table 3 presents the key parameters and results for 48 larger-size numerical experiments. We randomly generate the candidate container sets of these instances based upon the real data set of CTOX. Each instance involves the demand of a single OD pair to be served by one double-stack train only. The total TEUs, total numbers of 20s and 40s in the set of candidate containers are given in columns 2–4 respectively. The wagon payload limit (G_k) and the weight difference limit (Δ_k) are furnished in columns 5 and 6, respectively. We assume $|K| = 90$ (i.e., a train's capacity is 360 TEUs) since the maximum train size is not likely to exceed this value. The same as in Section 4.1, we assume $|I^C_{20}| = |I^C_{40}| = |S| = 0$, and $\alpha = 1$. For brevity, only the primary objective is optimized for these instances.[16] The optimal results, including the total tonnage and TEUs loaded, the total profit, and the CPLEX runtime are furnished in columns 7–10, respectively.

Instances 1–12 are summarized in the first row of Table 3. These are (likely rake-deficit) instances with very large sets of candidate containers (about four times the train's capacity), but with various numbers of 20s and 40s, G_k and Δ_k. Also see the last row of the table, which summarizes the key parameters and results of 12 rake-surplus instance. Note for all these 24 instances the computation time is always less than a minute. Thus, the very large and very small sizes of I seem not to be a concern for the DSLP computation time.

The remaining instances (numbered 13–36) in Table 3 span over all the combinations of $|I_{20}| + 2|I_{40}| \in \{720, 360\}$, $G_k \in \{78, 68, 61\}$ tonnes, and $\Delta_k \in \{20, 17, 14, 10\}$ tonnes. Comparison of these instances reveals that the DSLP runtime tends to (but not always) increase as the payload limit G_k or the balanced-load limit Δ_k diminishes. For most of these instances, the runtime is less than 100 s. However, for instances 34–36 with a candidate container set of 360 TEUs, the runtimes are much longer. This is possibly because when the candidate set size almost equals the train capacity (i.e. no extra containers for

[16] A note regarding the computation time for optimizing the secondary objective: in our extensive numerical experiments (not shown in this paper), optimizing the total tardiness always takes less than a minute using CPLEX.

selection), it is more difficult to find an optimal solution that satisfies the tight limits of G_k and Δ_k. Nevertheless, the runtime for the worst case in Table 3 is only about 10 min, which is still acceptable for real-world implementation.

Finally, note that all the trains in Table 3 are loaded to their full capacity. This manifests the effectiveness of the DSLP. Recall that in real practice many trains were not fully loaded due to human errors in manual planning (see Section 4.1).

4.3. Simultaneous optimization of loading plans for multiple trains

As explained in Section 1.3, practitioners desire to jointly optimize the loading plans of multiple trains only if those trains are not dispatched at nearly the same time on the same route. However, in case that multiple double-stack trains for a same route are being loaded at a terminal, we can simply apply our DSLP model to a "virtual train" that consists of the wagons of all the trains that are being planned simultaneously. For example, two 45-wagon trains can be treated as a virtual train of 90 wagons and optimized using the DSLP. The optimally-loaded 90-wagon "train" can be split into two real trains, with the first 45 wagons being the first train, and the remaining wagons being the second. This way ensures that the total profit and the container-tardiness of the multiple trains are optimized, and a near-optimal aerodynamic-resistance and HCG are obtained for each train.

The joint optimization of the planning for multiple trains can bring more profit for the CTO. The computation time is not an issue if the virtual train is not too long (e.g. when two 45-wagon trains are planned simultaneously). However, the joint optimization of more trains would likely require more efficient solution algorithms. In this paper, considering the computation time constraints, the uncertainties, and the importance of loading of one train at a time (explained in Section 1.3), our DSLP model has taken the double-stacking of future trains into account through a simple parameter α (see Section 2.3). Next we examine how this parameter affects the DSLP solution.

4.3.1. Effect of α on the DSLP solution

We test the effect of the value of α on the DSLP solution through eight randomly generated instances, as summarized in Table 4. In each instance, the total number of TEUs of the candidate container set is always greater than the double-stack train's capacity, which is 180 TEUs for 45 wagons, because otherwise α should always be equal to 1. The total TEUs of 40s and 20s, and the total tonnage of the candidate container set are presented in columns 2, 3, and 4 of the table 4, respectively. Other detailed input parameters are omitted. To simplify the analysis, we assume that assigning a 40-ft container to a lower-stack position yields the same profit as assigning two 20s instead. Thus any value of $\alpha < 1$ would make the assignment of two 20s to the lower stack more preferable than the assignment of one 40-ft.

The optimal loading plans for three different values of α (1, 0.9, and 0.2) are presented in columns 5–16. The output of each loading plan includes the total TEUs of 40s and 20s loaded on the double-stack train, the total tonnage loaded, and the computation time via CPLEX.

For instances 1–3 and 5–7, the train is loaded to its full capacity regardless of α, but using an α less than 1 (no matter what the exact value is) does reduce the number of 40s loaded on the train to the minimum possible. For example, for instance 1 with $\alpha = 1$, the DSLP assigns 170 TEUs of 40s to the train, while with $\alpha = 0.9$ and 0.2, only 160 TEUs of 40s are assigned. This number cannot be reduced further because only 20 TEUs of 20s are available in the candidate container set, and thus at least 160 TEUs of 40s should be loaded to make the train full. For instances 4 and 8, on the other hand, all the 40s are loaded regardless of the α because there are not enough 40s to even fill up the upper stacks (i.e., fewer than 90 TEUs of 40s). In these two instances the train's slot utilization is less than 100%, and the α is irrelevant. These instances validated the effectiveness of using α to save the 40s for loading upper-stack of future trains without compromising the utilization of the present train (recall the saving from even one extra container loading is huge). This is especially useful when the CTO anticipates a shortage of 40s in the future.

Finally, note in all the instances presented in Table 4 that assigning $\alpha = 0.9$ and 0.2 yield almost the same loading plan in very similar runtimes. However, extreme values of α that are very close to zero or one should be avoided, since they may affect the numerical accuracy of the solution and potentially lead to an unintended solution.

5. A decision support system

Built upon the DSLP model, an interactive decision support system (DSS) has been developed to assist the assignment clerks in generating efficient double-stack train loading plans. This DSS, named the "Double-Stack Loading Planner", has been successfully implemented by the CTOX. Some other Indian CTOs have also expressed their interest in this application.

The overall architecture of the DSS is shown in Fig. 7. The DSS is linked to the CTO's management information system to access the container data from every terminal. The DSS itself is hosted on Microsoft® Windows and CPLEX® (version 12.5) is used as the optimization solver. The application is connected with a centralized Oracle 11 g database. It can also work with Microsoft® Excel in a standalone mode. The DSS provides efficient train loading plans with key details such as the total profit, haulage cost, total tonnage, total TEUs loaded, and slot utilization. These data can be used to generate the train summary report, which is required by the Indian Railways.

Table 4
The effect of α.

	Inputs			Optimal loading plan ($\alpha = 1$)				Optimal loading plan ($\alpha = 0.9$)				Optimal loading plan ($\alpha = 0.2$)			
S. N.	Total TEUs of 40s	Total TEUs of 20s	Total Tonnage	Total TEUs of 40s	Total TEUs of 20s	Total Tonnage	Runtime (s)	Total TEUs of 40s	Total TEUs of 20s	Total Tonnage	Runtime (s)	Total TEUs of 40s	Total TEUs of 20s	Total Tonnage	Runtime (s)
1	340	20	3557.53	170	10	2163.9	102	160	20	2289.57	91	160	20	2311.54	97
2	290	70	3943.15	120	60	2530.79	88	110	70	2633.89	88	110	70	2595.58	92
3	220	140	4510.72	126	54	2166.19	11	90	90	2521.86	8	90	90	2490.18	9
4	70	290	5824.38	70	90	2278.17	8	*NA*				*NA*			
5	250	20	2709.53	166	14	2003.12	110	160	20	2130.18	116	160	20	2047.22	125
6	200	70	3140.57	120	60	2318.66	91	110	70	2433.68	98	110	70	2399.46	90
7	130	140	3750.73	124	56	2354.61	10	90	90	2391.70	10	90	90	2431.12	11
8	70	200	4204.20	70	90	2212.69	6	*NA*				*NA*			

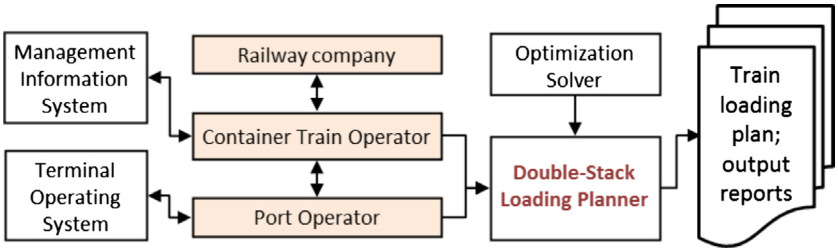

Fig. 7. The architecture of the DSS.

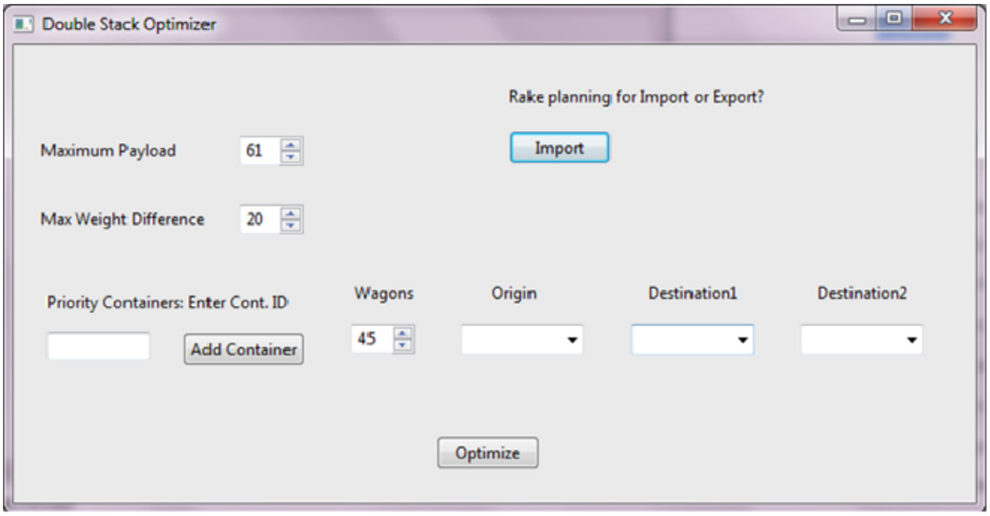

Fig. 8. The main window of the Double-Stack Loading Planner.

Fig. 8 shows the main window of the simple and user-friendly graphical user interface (GUI) of the DSS. Most of the DSLP parameters are extracted from the connected database or set to their default values. The remaining inputs take only one minute to enter. An ordinary clerk can easily operate this application after some basic training.

In addition to the input parameters of the DSLP described in Section 2, the DSS also retrieves the following key information to assist the users in making proper decisions: (i) the network's congestion status, the trains' current positions, expected arrival and departure times (retrieved from the railway company); (ii) the trains' routing plan and the criteria for shortlisting containers at the terminals (from the CTO); and (iii) containers' positions in the yard and their accessibility (from the terminal operator). Depending on the above information, we make the following recommendations for using the DSS.

(i) *Choosing between forming a single- and a double-stack train.* The DSLP does not decide whether a single- or a double-stack train should be formed. However, the following rules of thumb are recommended for this issue:
  (a) If a terminal is rake-deficit, i.e., there are more outbound containers than inbound containers, then a double-stack train should be formed and loaded to its fullest capacity. For a rake-surplus terminal, this decision mainly depends on the empty rake distribution among other factors.
  (b) For a congested route, double-stack trains should be formed and loaded to their fullest capacity, which will reduce the rail traffic.
  (c) For small terminals with light container traffic, the decision depends on whether delaying the delivery of some containers can enable a double-stack train at a later time and reduce the container haulage cost.

(ii) *Selection of candidate containers.* In addition to designating the sets of compulsory containers (I^C_{40}, I^C_{20}) and group-containers (I^s_{40}, I^s_{20}), the following thumb-rules are recommended for determining the candidate container set:
  (a) In case of a large container pendency, it is good enough to specify a cut-off reference time (T_i) to shortlist the older containers as I. The cut-off time should be selected so that $|I|$ is about 2–3 times more than the train capacity.

(b) In most of the time, the selected candidate containers should have the same destination. This is consistent with the practice in India because most of the freight trains are through-trains without any intermediate stop.[17] The remaining few trains have only one intermediate stop where some of the containers carried by the train will be unloaded. To minimize the container handling and train delay at the intermediate stop for these trains, one can arrange to: (1) place the containers destined for the intermediate stop in the upper stacks; and/or (2) place the containers destined for the final stop in the lower stacks. This special requirement can be entertained in the preprocessing stage of the DSLP by specifying relevant sets of containers in the constraints (8)-(10).

(c) Occasionally, a few candidate containers stored in the yard might be inaccessible when the train is being loaded. In such last minute changes, the train loading supervisor can replace them with some accessible containers.

In addition, the train loading planner should make sure that a sufficient inventory of the twist-locks is maintained at all times, because a shortage of even one twist-lock may cause one less container to be loaded on the train. The resulting profit loss can be more than ten times the purchase cost of the twist-lock. Finally, the train loading plan generated by the DSS should be shared with the terminal manager well before the train loading starts for better intra-terminal operations planning and timely container retrieval.

6. Conclusions

Our work has contributed to both the academia and the industry. We develop a new deterministic model, DSLP, for optimizing the loading plan of double-stack container trains with the main objective of maximizing the CTO's profit. The DSLP is, to our best knowledge, the first container train loading optimization model that accounts for many realistic features, including multiple double-stack loading patterns, different heights, lengths, and weights of the containers, and safety limits for the wagons' payload and load-balance. The model is also the first to analyze and jointly consider a number of practically important objectives. These objectives include: (i) maximizing the CTO's profit, which is a more general (and practically preferred) objective than the commonly used 'maximize train utilization'; (ii) maximizing selection of the oldest containers for improving the customers' satisfaction; (iii) minimizing the aerodynamic resistance for saving the fuel; and (iv) minimizing the horizontal center-of-gravity of the train for reducing the maintenance cost.

Despite the above complexity, we manage to develop an efficient CPLEX based solution approach by exploiting key mathematical properties of the DSLP. (Recall that the practitioners strongly prefer an exact optimal solution due to the significant potential gains over any suboptimal solution.) With this efficient solution approach and a user-friendly decision-support-system built upon the DSLP, our work has been implemented to assist the daily load planning of double-stack container trains for a major CTO in India. Note that this is, to our knowledge, the first real-life implementation of an optimization model for double-stack container train loading. Thus, it represents a notable progress towards bridging the gap between research and practice.

The real-world application has shown that the DSLP can increase the profit generated by an individual train by about 2–3% as compared to the previous manual planning results. In addition, implementation of the DSS expedites the train planning process and rules out any inadvertent human errors that may cause high extra cost, hassles, and customer dissatisfaction. Thus, this will improve the overall efficiency of container train operations.

In the future, we will seek to extend the DSLP model for the joint optimization of multiple trains. Although it is desirable, it is quite challenging to model and solve this new problem due to the additional uncertainties and complexities associated with the routing and scheduling of multiple trains and containers. Further, the present DSLP model ignores the intra-terminal planning of truck and crane movements, mainly because the potential cost savings are small as compared to the potential savings from the DSLP. However, a terminal operator is interested in minimizing the intra-terminal container handling cost without increasing the cost of train operations. We are seeking to develop a realistic model for minimizing the intra-terminal container handling cost for given DSLP loading plans, which account for the terminal layout, position of containers in the yard, and the availability of trucks and cranes.

Acknowledgements

This study is partially funded by a start-up grant from the Hong Kong Polytechnic University and partially by the CTOX. We thank the anonymous referees for their constructive comments. We also thank Adhyan Hirwani and Dev Priyam Mishra for their support in coding the GUI and deploying the application for the CTOX.

[17] Even for the US case studied by Lai et al. (2008a), 80% of the container trains have no intermediate stop.

Appendix A. List of notation

Parameters

G_k	Payload limit (maximum permissible payload) for wagon $k \in K$
H_i	Height of container $i \in I$
$\bar{H}_k$	Height of the wagon platform measured from the rail surface, $k \in K$
H_k^L, H_k^U	Heights of the lower- and upper-stack container(s) loaded on wagon $k \in K$
H^o	Height of interbox twist-lock
$\hat{H}$	Highest possible height of a container (9.5 feet for the Indian case)
I	Set of candidate containers; index $i \in I$
I_{20}, I_{40}	Sets of 20-ft and 40-ft containers, respectively ($I = I_{20} \cup I_{40}$)
I_{20}^C, I_{40}^C	Sets of 20-ft and 40-ft compulsory containers, respectively, which must be loaded in the train ($I_{20}^C \subseteq I_{20}, I_{40}^C \subseteq I_{40}$)
I_{20}^s, I_{40}^s	Sets of 20-ft and 40-ft containers belonging to a shipping bill $s \in S$ ($I_{20}^s \subseteq I_{20}, I_{40}^s \subseteq I_{40}$)
J	Set of allowed container loading patterns; index $j \in J \equiv \{1, 2, 4, 5, 6\}$
K	Set of rail wagons for the train; index $k \in K$ refers to the position of a wagon counted from the locomotive
K_1, K_2	Sets of wagons loaded in single- and double-stack loading patterns, respectively
m	Index to refer to a position of a container loaded on a wagon, $m \in \{A, B, C, D\}$
M_1, M_2	Big constants
$\bar{N}_{40}^F$	Number of total 40-ft containers suitable for loading in the upper-stack positions of the train
P_i^L, P_i^U	Profit parameters for the assignment of container i to the lower- and upper-stack position of the train, respectively
R_k	Maximum limit on the vertical center-of-gravity (CG) of a loaded wagon k
$\bar{R}_k$	CG height of an empty wagon k
S	Set of shipping bills with a special request to load either all or none of the containers on the train; index $s \in S$
T_i	Time of booking of container i
V_k	Set of containers that are assigned to wagon k
W_i	Gross weight of container i
$\bar{W}_k$	Tare weight of wagon k
W_k^L, W_k^U	(Total) weight of the lower- and upper-stack container(s) on wagon k, respectively
W_{40}^{min}	Weight of the lightest 40-ft container in I_{40}
Δ_k	Maximum limit for the weight difference between the two 20s loaded on the same wagon k
α	Parameter to address the relative shortage of 40-ft containers for future double-stack trains
π	Total container haulage profit of the train
τ	Total container haulage tardiness of the train

Binary decision variables

x_k^j	if wagon $k \in K$ is loaded in pattern $j \in J$, $x_k^j = 1$; else $x_k^j = 0$
y_{ik}^j	if a 20-ft container $i \in I_{20}$ is loaded in position $m \in \{A, B\}$ of wagon $k \in K$, $y_{ik}^j = 1$; else $y_{ik}^j = 0$
z_{ik}^j	if a 40-ft container $i \in I_{40}$ is loaded in position $m \in \{E, F\}$ of wagon $k \in K$, $z_{ik}^j = 1$; else $z_{ik}^j = 0$
b_s	if all the containers in $s \in S$ are assigned to the train, $b_s = 1$; else $b_s = 0$

References

Ahuja, R.K., Sahin, G., Cunha, C.B., 2005. Network models in railroad planning and scheduling. In: Tutorials Operat. Res.: Emerging Theory, Methods, Appl.. INFORMS, pp. 54–105.

Ambrosino, D., Siri, S., 2015. Comparison of solution approaches for the train load planning problem in seaport terminals. Transp. Res. Part E: Logistics Transp. Rev. 79, 65–82.

Bierwirth, C., Meisel, F., 2010. A survey of berth allocation and quay crane scheduling problems in container terminals. Eur. J. Oper. Res. 202 (3), 615–627.

Bontekoning, Y.M., Macharis, C., Trip, J.J., 2004. Is a new applied transportation research field emerging?––a review of intermodal rail–truck freight transport literature. Transp. Res. Part A: Policy Practice 38 (1), 1–34.

Bostel, N., Dejax, P., 1998. Models and algorithms for container allocation problems on trains in a rapid transshipment shunting yard. Transp. Sci. 32 (4), 370–379.

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

Bruns, F., Knust, S., 2012. Optimized load planning of trains in intermodal transportation. OR Spectrum 34 (3), 511–533.

Bruns, F., Goerigk, M., Knust, S., Schöbel, A., 2014. Robust load planning of trains in intermodal transportation. OR Spectrum 36 (3), 631–668.

Caris, A., Macharis, C., Janssens, G.K., 2008. Planning problems in intermodal freight transport: accomplishments and prospects. Transp. Plan. Technol. 31 (3), 277–302.

Chih, K.C.K., Bodden, M.P., Hornung, M.A., Kornhauser, A.L., 1990. Routing and Inventory Logistics System (RAILS): a heuristic model for optimally managing intermodal double-stack trains. J. Transp. Res. Forum 31 (1).
Container Terminal Foresight 2020, 2014. available on: <http://www.ctf2020.info/> (accessed on 10-Sep-2014).
Corry, P., Kozan, E., 2006. An assignment model for dynamic load planning of intermodal trains. Comput. Oper. Res. 33 (1), 1–17.
Corry, P., Kozan, E., 2008. Optimised loading patterns for intermodal trains. OR Spectrum 30 (4), 721–750.
DFCCIL, 2016. Dedicated Freight Corridors Corporation of India Ltd. <http://www.dfccil.org> (accessed on 30.01.2016).
Feo, T.A., González-Velarde, J.L., 1995. The intermodal trailer assignment problem. Transp. Sci. 29 (4), 330–341.
Günther, H.O., Kim, K.H., 2005. Container Terminals and Automated Transport Systems. Springer, Heidelberg.
Jahren, C.T., Pacanovsky, D.L., 1993. A decision support system for loading containers onto double-stack railcars. No. TNW93-13, University of Washington, Seattle.
Kim, K., Günther, H.O., 2006. Container Terminals and Cargo Systems. Springer, Berlin.
Lai, Y.C., Barkan, C., 2005. Options for improving the energy efficiency of intermodal freight trains. Transp. Res. Rec.: J. Transp. Res. Board 1916, 47–55.
Lai, Y.C., Barkan, C.P.L., Onal, H., 2008a. Optimizing the aerodynamic efficiency of intermodal freight trains. Transp. Res. Part E 44 (5), 820–834.
Lai, Y.C., Ouyang, Y., Barkan, C.P.L., 2008b. A rolling horizon model to optimize aerodynamic efficiency of intermodal freight trains with uncertainty. Transp. Sci. 42 (4), 466–477.
Lang, M., Przybyla, J., Zhou, X., 2011. Loading containers on double-stack cars: Multi-objective optimization models and solution algorithms for improved safety and reduced maintenance cost. Unpublished, available on <http://www.civil.utah.edu/~zhou/container_scheduling.pdf>.
Li, C., Kost, M., Sheridan, D., Burton, J., Thompson, M.C., 2014. Impact of gap/load spacing on the aerodynamic performance of intermodal freight trains. In: 19th Australasian Fluid Mechanics Conference. Melbourne, Australia, 8-Dec-2014, pp. 2–5.
Lopez-Gomez, A., 1987. Computer-Aided Simulation in Railway Dynamics. CRC Press, p. 47.
Lutter, P., Werners, B., 2015. Order acceptance for motorail transportation with uncertain parameters. OR Spectrum 37 (2), 431–456.
Meisel, F., 2009. Seaside Operations Planning in Container Terminals. Physica-Verlag, Berlin, Heidelberg.
Murty, K.G., Liu, J., Wan, Y.W., Linn, R., 2005. A decision support system for operations in a container terminal. Decis. Support Syst. 39 (3), 309–332.
Pacanovsky, D.L., Jahren, C.T., Palmer, R.N., Newman, R.R., Howland, D., 1995. A decision support system to load containers to double-stack rail cars. Civil Eng. Syst. 11 (4), 247–261.
Powell, W.B., Carvalho, T.A., 1998. Real-time optimization of containers and flatcars for intermodal operations. Transp. Sci. 32 (2), 110–126.
Rates Circular, 2014. Haulage Charge for movement of containers in privately owned wagons. Indian Railway, Circular No. 2007 ITC-1/302/1 Pt.H, 20-Nov-2014. <www.indianrailways.gov.in/railwayboard/uploads/directorate/traffic_comm/Freight_Rate_2K13/RC_03_13.pdf>.
RDSO Drawing, 2013. Maintenance manual for rail wagons. BLC container wagon drawing number: CONTR – 9406 – S/2, Research Designs & Standards Organization, Lucknow, India.
Rentmeesters, M.J., Tsai, W.K., Lin, K.J., 1996. A theory of lexicographic multi-criteria optimization. In: Engineering of Complex Computer Systems, Proceedings, Second IEEE International Conference, October 1996, pp. 76–79.
RITES, 2014. Total Transport System Study on Traffic Flows and Modal Costs. RITES Ltd., Gurgaon, India.
Steenken, D., Voss, S., Stahlbock, R., 2004. Container terminal operation and operations research - a classification and literature review. OR Spectrum 26 (1), 3–49.
Wang, S., Meng, Q., 2012. Robust schedule design for liner shipping services. Transp. Res. Part E: Logistics Transp. Rev. 48 (6), 1093–1106.