# A Rolling Horizon Framework for Intermodal Loading Assignment to Improve Fuel Efficiency

**Yung-Cheng Lai, Yanfeng Ouyang, Christopher P.L. Barkan,**
**Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign**

**Summary:** Aerodynamic efficiency of intermodal freight trains can be significantly improved by minimizing the adjusted gap lengths between adjacent loads. This paper first develops a static model to optimize load placement on a sequence of intermodal trains that have scheduled departure times. This model applies when full information on all trains and loads is available. Then, a dynamic model is developed to account for the realistic situation where information on future trains and incoming loads may not be completely available. This paper seeks ways to balance between: (i) the advantage from optimizing multiple trains together; and (ii) the risk of making suboptimal decisions due to incomplete future information. We propose a rolling horizon scheme to address this challenge, where exponentially decreasing weights are assigned to the objective functions of future trains. Numerical results based on empirical data show significant aerodynamic efficiency benefits from these optimization models.

**Index Terms:** intermodal transportation, energy efficiency

## NOTATION AND UNITS

| | |
|---|---|
| $i$ | index referring to the type and size of the load |
| $C$ | subset of i for containers |
| $J_i$ | number of loads of a specific type and size $i$ |
| $t$ I | index for outgoing trains ($t$ = 1, 2…$T$). $k$ position of each unit in the train |
| $A_k$ | adjustment factor of the $k^{th}$ gap shown in $U_{tk}$ length of $k^{th}$ unit of train t |
| $\delta_k$ | the type of $k^{th}$ unit, where $\delta_{tk} = 1$ when the unit is a well-car unit, and $\delta_{tk} = 0$ otherwise |
| $L_i$ | length of the $i^{th}$ type load |
| $Q_{tkp}$ | length limit of position $p$ in the $k^{th}$ unit of train $t$ |
| $w_{ij}$ | weight of the $j^{th}$ load of type $i$ |
| $C_{tk}$ | weight limit of the $k^{th}$ unit of train t |
| $R_{itpk}$ | a four dimensional matrix for loading capabilities of each slot, where $R_{itpk} = 1$ if the $i^{th}$ type of load can be assigned to position $p$ in unit $k$ of the train $t$, and it is 0 otherwise |
| $\Phi$ | represents an arbitrarily specified large number |
| $\alpha_t$ | an smooth decreasing weight |

## INTRODUCTION

Intermodal (IM) freight is among the largest segments of US railroad freight transportation, and has experienced rapid growth over last 15 years *[1, 2]*. However, intermodal trains are generally the least fuel efficient due to the physical constraints imposed by the combination of loads and the railcar design *[3, 4]*. This is particularly ironic given that intermodal trains are typically the fastest freight trains in operation.

At intermodal terminals, containers or trailers are assigned to available well, spine or flat cars *[5]*. Although computer software *[6]* is often used by terminal managers to assist in this task, it is still a largely manual process. Railroads already provide incentives to terminal managers for maximizing slot utilization, but they do not take into account the size of the slot compared to the size of the load. A perfect slot utilization indicates maximal use of the feasible slots available, it is not intended to, nor does it ensure, that intermodal cars are loaded to maximize the energy efficiency of intermodal trains. For example, two trains may have identical slot utilization, but different loading patterns and consequent train resistances *[7]*. Consequently, there is a gap between slot utilization and energy efficiency, and work needs to be done to merge automatic loading assignment models into the terminal software used by major North American railroads.

Lai et al *[7]* developed an optimal loading assignment model in which the objective function was maximizing aerodynamic efficiency of the outgoing intermodal freight train given any particular static combination of loads and railcar types. Because terminal managers usually load only one train at a time, each optimization analysis considers only the current outgoing train with available loads. If advance information on outgoing trains and loads is available, better loading plan solutions may be possible by optimizing for multiple trains simultaneously. The larger pool of loads and railcars will enable better matching.

This research is particularly timely in light of recent increases in fuel prices, their impact on industry operating costs, and the need to conserve energy and reduce green house gas emissions. Class I railroads spent more than $6.2 billion on fuel in 2005 making it their second largest operating expense *[8]*. As of 2005, fuel costs had increased by 2.67 times since 1998, and this trend continues, making fuel efficiency more important than ever *[9]*. This calls for investigation of options to improve aerodynamic efficiency, which have important economic and environmental implications for rail freight transportation *[10, 11]*.

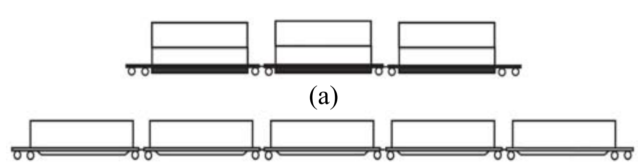
In this paper, we first extend the model to optimize the aerodynamic efficiency at the multiple train system level. The benefit of optimizing more trains and loads will be evaluated assuming static information of trains and loads. We will then allow dynamic load information by developing a rolling horizon scheme for continuous terminal operations. This will be followed by an empirical case study and some discussion.

## METHODOLOGY

### Loading Assignment at Intermodal Terminals

At intermodal terminals, containers and trailers of a variety of lengths are assigned to available well, spine or flat cars by terminal managers *[5, 12]*. In this study, we focus on intermodal services of the BNSF Railway between Chicago and Los Angeles (LA). About 80% of the IM trains on this route are loaded or unloaded only at Chicago and LA, so there is little container shifting occurring enroute *[13, 14, 15]*.

Intermodal loads, i.e. trailers or containers, range in length from 20 to 57 ft. There is considerable variety in the design and capacity of intermodal railcars with different numbers of units and slots, and thus loading capabilities. An intermodal railcar can have one or more units permanently attached to one another (via articulation or drawbar). A unit is a frame supported by at least two trucks, providing support for one or more platforms (a.k.a. slots). For example, Figure 1a shows an articulated 3-unit well car, and Figure 1b is a 5-unit spine car. A platform (or slot) is a specific container/trailer loading location. As a result, each well-car unit has two slots because of their accommodation of two containers, one stacked on the other (a.k.a. "double stack"), and each spine-car unit has one slot (Figure 1).



(a)

(b)

FIGURE1 (a) a 3-unit well car with 6 slots (b) a 5-unit spine car with 5 slots

There are also a number of loading rules developed for safety purposes and various feasible and infeasible combinations of IM load and car configuration. Because intermodal cars in a train are not generally switched in and out at terminals, managers primarily control the assignment of loads but not the configuration of the equipment in a train. Consequently, we treat the train make-up as given in this study.

Aerodynamic drag is a major component of train resistance, particularly at high speeds *[16, 17, 18]*. The Association of American Railroads (AAR) supported research on wind tunnel testing of rail equipment, including large-scale intermodal car models *[19]*. The results were used to develop the Aerodynamic Subroutine of the Train Energy Model (TEM) *[20]*. These experiments showed that gap length between IM loads and position-in-train were the two important factors affecting train aerodynamics *[21]*. Larger gaps result in a higher aerodynamic coefficient and greater resistance. The front of the train experiences the greatest aerodynamic resistance due to headwind impact. Therefore, to incorporate both the gap length and position-in-train effect, the objective function of the model is to minimize the adjusted gap length within the train. Adjusted gap length is equal to the adjustment factor times gap length. The adjustment factor associated with each gap is computed by dividing the drag area of a given unit by the drag area of the 100$^{th}$ unit (Table 1) to account for position-in-train effect.

TABLE 1 Adjustment factor for each gap in the train *[7]*

| k | Drag area (ft$^2$) | Adjustment factor |
|---|---|---|
| 1 (locomotive) | 31.618 | 1.5449 |
| 2 | 28.801 | 1.4073 |
| 3 | 26.700 | 1.3046 |
| 4 | 25.133 | 1.2280 |
| 5 | 23.963 | 1.1709 |
| 6 | 23.091 | 1.1283 |
| 7 | 22.440 | 1.0964 |
| 8 | 21.954 | 1.0727 |
| 9 | 21.591 | 1.0550 |
| 10 | 21.320 | 1.0418 |
| 100 | 20.466 | 1.0000 |

## Static Aerodynamic Efficiency Model

The following notation is used in the algebraic model: $i$ is an index referring to the type and size of the load (namely, 40' container, 48' trailer, 53' trailer, etc.); $C$ is the subset of i for containers. We group loads of the same type together with an index, $j$ ($j$ = 1, 2, 3…$J_i$); $J_i$ is the number of loads of a specific type and size $i$ ($i$ = 40C, 48'T, 53T, etc.), for instance, $J_{48T}$ =10 means that there are ten 48' trailers in the storage area. $t$ is the index for outgoing trains ($t$ = 1, 2…$T$). The symbol $k$ defines the position of each unit in the train ($k$ = 1, 2, 3…$N$), where $k$=1 corresponds to the first intermodal unit of the train. The slot position in each unit is denoted by $p$, where $p$ = 1 represents the upper (top) platform in a well-car unit or the single platform in a spine-car or flat-car unit, and $p$ = 2 represents the lower (bottom) platform in a well-car unit (Figure 2). The following symbols represent the parameters used in the model: $A_k$ is the adjustment factor of the $k^{th}$ gap shown in Table 1, where $A_1 > A_2 > \ldots > A_N$; $U_{tk}$ is the length of $k^{th}$ unit of train t; $\delta_k$ indicates the type of $k^{th}$ unit, where $\delta_{tk} = 1$ when the unit is a well-car unit, and $\delta_{tk} = 0$ otherwise; $L_i$ is the length of the $i^{th}$ type load; $Q_{tkp}$ is the length limit of position $p$ in the $k^{th}$ unit of train $t$; $w_{ij}$ is the weight of the $j^{th}$ load of type $i$; $C_{tk}$ is the weight limit of the $k^{th}$ unit of train t; and $R_{itpk}$ is a four dimensional matrix for loading capabilities of each slot, where $R_{itpk}$ = 1 if the $i^{th}$ type of load can be assigned to position $p$ in unit $k$ of the train $t$, and it is 0 otherwise. Finally, $\Phi$ represents an arbitrarily specified large number introduced for modeling purposes as will be explained in the model description below.

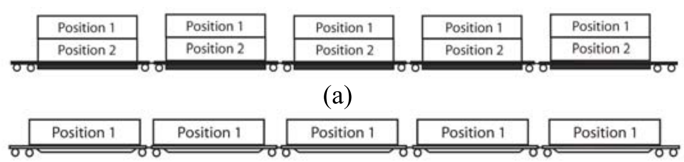



FIGURE 2 The available slots in (a) 5-unit well car (b) 5-unit spine car

Two sets of binary decision variables are included in the IP model. The first variable is denoted by $y_{ijtpk}$ where:

$$y_{ijtpk} = \begin{cases} 1, & \text{if } j^{th} \text{ load of type } i \text{ is assigned to position } p \text{ in } k^{th} \text{ unit of train } t \\ 0, & \text{otherwise} \end{cases}$$

The second binary variable, denoted by $x_{tk}$, determines whether the top slot in a well unit can be used, namely:

$$x_{tk} = \begin{cases} 1, & \text{if the top slot of the } k^{th} \text{ unit in train } t \text{ can be used} \\ 0, & \text{otherwise} \end{cases}$$

According to the loading rules, the top slot can be used when the bottom slot is filled by containers whose total length is at least 40' *[22]*.

The loading problem is formulated as a linear integer program to minimize fuel consumption (i.e., the total adjusted gap length) of all outgoing trains. For train *t*, the objective function is

$$z_t = \frac{A_1}{2}\left(U_{t1} - \sum_i \sum_j y_{ijt11} L_i\right) + \frac{\sum_{k=1}^{N} A_{k+1}}{2}\left[\left(U_{tk} - \sum_i \sum_j y_{ijt1k} L_i\right) + \left(U_{tk+1} - \sum_i \sum_j y_{ijt1k+1} L_i\right)\right] \tag{1}$$

The complete mathematical program for all *T* trains is as follows:

$$\text{Min} \sum_{t=1}^{T} z_t \tag{2}$$

Subject to:

$$\sum_t \sum_p \sum_k y_{ijtpk} R_{itpk} \le 1 \quad \forall i, j \tag{3}$$

$$y_{ijtpk} \le R_{itpk} \quad \forall i, j, t, p, k \tag{4}$$

$$40 - \sum_{i \in C} \sum_j y_{ijt2k} L_i \le \Phi(1 - x_{tk}) \quad \forall t, k \text{ (such that } \delta_{tk} = 1) \tag{5}$$

$$\sum_{i \in C} \sum_j y_{ijt1k} \le x_{tk} \quad \forall t, k \text{ (such that } \delta_{tk} = 1) \tag{6}$$

$$\sum_i \sum_j \sum_p y_{ijtpk} w_{ij} \le C_{tk} \quad \forall t, k \tag{7}$$

$$\sum_i \sum_j y_{ijtpk} L_i \le Q_{tkp} \quad \forall t, k, p \tag{8}$$

$$y_{ijtpk}, x_k = 0, 1 \tag{9}$$

The objective function (total adjusted gap length for train *t*) is comprised of two parts. The first part, representing the gap length between the locomotive and the first load (Figure 3), is the difference between the length of the first unit *($U_{tl}$)* and the length of the load in position 1 of the 1st unit *($\sum\sum y_{ijt11}L_i$)*, which is then divided by 2. Multiplying the gap length by the adjustment factor $A_1$ results in the first adjusted gap length. Each of the subsequent gaps is half of the difference in length between the current unit and the load *($U_{tk}$ - $\sum\sum y_{ijt1k}L_i$)/2* plus half of the length difference between the next unit and the load *($U_{tk+1}$ - $\sum\sum y_{ijt1k+1}L_i$)/2* multiplied by the appropriate adjustment factor, $A_k$. Thus, the second part of the objective function computes the sum of the subsequent adjusted gap lengths. Note that we only take into account the loads in position 1 of all units in the train. This is reasonable since they are the only loads in spine or flat cars; and for well cars, the upper level gaps have a more significant aerodynamic effect than the lower level gaps *(23, 24, 25)*. A schematic representation is given in Figure 3.

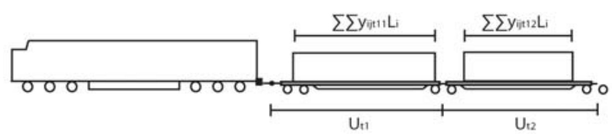


FIGURE 3 Locomotive and first two intermodal units in a train

Minimizing total adjusted gap length creates the most efficient train configuration. However, not all loads can be assigned to all slots. The loading assignment must conform to the loading capability of each unit as well as length and weight constraints. Constraint (3) ensures that each load can be assigned to no more than one slot, and it has to obey the loading assignment rules ($R_{itpk}$); constraint (4) eliminates the possibility of assigning loads ($y_{ijtpk}$ = 1) to unavailable slots ($R_{itpk}$ = 0). Constraints (5) and (6) together state that if the bottom slot (position 2) in a well-car unit ($\delta_{tk}$ = 1) is not filled with containers greater than 40 ft, in which case equation (5) requires that $x_{tk}$ = 0, then no load can be assigned to the top slot (position 1) for the same unit, i.e., $\sum\sum y_{ijt1k} = 0$ and therefore $y_{ijt1k} = 0$ for all *i,j*. Note that constraint (6) allows a bottom load without a top load ($y_{ijt1k}$ = 0). Constraint (7) is the weight limit that is imposed for each car unit in order to reflect its total carrying capacity ($C_{tk}$). And, constraint (8) is the length limit imposed for each slot to guarantee that the total length of loads in a given slot (position) does not exceed the length of that slot ($Q_{tkp}$). Constraint (9) states that both $y_{ijtpk}$ and $x_{tk}$ are binary variables. Note that the trivial solution, namely $y_{ijt1k} = 0$ and $x_{tk} = 0$, satisfies all the constraints of the model. However, this would result in the largest total adjusted gap since all gaps would be at their maximum value. This case is ruled out because of the minimization of the total gap. Thus, the model prefers not to leave a load behind if a suitable slot is available.

Conventional practice is to consider loads only for the current outgoing train. This scenario is a special case of the general model we developed here by setting $t$ equal to 1. The model thus optimizes the aerodynamic efficiency of one outgoing train for a given set of loads. However, if advance information about outgoing trains and loads is available, optimizing more trains and loads together may lead to even more aerodynamically efficient loading patterns.

**Dynamic Aerodynamic Efficiency Model**

Optimization of multiple trains will often lead to more efficient loading overall if complete information on all trains and loads is available at the time of optimization (i.e., static current information). However, the information requirement on future loads and trains often imposes some degree of uncertainty in practice, because information about some loads may not be immediately available (i.e., dynamic future information). Under some circumstances, optimizing the loading pattern of a later train will reduce the efficiency of the immediate outgoing train. For example, the two trains may compete for the same "suitable" load, and the later train may get it. There is some possibility that after the dispatch of the immediate train, another suitable load with the same characteristics becomes available for the later train. In this case, the earlier optimal solution (without future information) turns out to be suboptimal (overall). Therefore, uncertainty about future loads introduces some degree of risk that the overall optimum for multiple trains will not be achieved. In a dynamic setting, there is a trade-off between the benefit of optimizing multiple trains simultaneously versus the risk of making wrong decisions for the uncertain future.

To address this trade-off, we propose a dynamic loading approach with rolling horizons, where loading decisions are updated over time as new information becomes available. Carrying out this approach poses three questions: (1) when to optimize loading patterns for one (or more) trains; (2) how many trains to optimize each time; and (3) how many trains to load after each optimization.

The first and third questions are relatively simple to answer. In principle, it is always better to postpone an optimization to the last moment possible (before loading a departing train), because it maximizes the available information, thereby reducing uncertainty. On the other hand, loading a train early will often lead to a suboptimal loading pattern because less information is available. Therefore, to the extent practicable train loading should be delayed until just before its departure. For the same reason, it is always better to load just the immediately outgoing train based on the optimal loading pattern even though multiple trains may be optimized together. Hence, we should always load the minimum number of trains if the solution speed is efficient enough to frequently update the optimal loading patterns. The only remaining question is how many trains should be optimized each time.

The dynamic train loading problem can be related to dynamic vehicle routing problems, where vehicles deliver loads to dynamically emerging customers with schedules. See Gendreau *[26]*, Ghiani et al. *[27]*, and Larsen *[28]* for reviews. Psaraftis *[29]* proposed a rolling horizon approach where decisions are made at any time $t$ with regard to loads in a future time interval $[t, t+L]$ while only the loads in $[t, t+\alpha L]$, $0<\alpha<1$, are actually assigned to departing vehicles. Later, Mitrovic et al. *[30]* proposed a double horizon heuristic. The objective function is formulated as a linear combination of actual costs in a short horizon and indirect costs (slack time in vehicle schedule) in a long horizon.

The modification of the objective function has been shown to improve the optimal solution by balancing short-term versus long-term costs.

Based on similar ideas, we propose an exponential smoothing approach to the dynamic train loading problem. Before loading the $t^{th}$ train, we optimize the following weighted average of objectives for trains departing in a future horizon:

$$\text{Min } \sum_{s=t}^{t+\tau} \alpha_{t,s} z_s \quad (10)$$
$$\text{s.t. } (3)\text{-}(9)$$

In (10), $\tau$ is the maximum number of future trains that can be filled with currently known loads, and $\alpha_{t,s}$ is a weight assigned to a future train $s \geq t$. The set of weights, $\tilde{\alpha}(t) := (\alpha_{t,t}, \alpha_{t,t+1}, \ldots, \alpha_{t,\tau}, \ldots)$, specifies how future trains are considered in the loading decision. For example, $\tilde{\alpha}(t) = (1, 0, 0, \ldots)$ corresponds to the trivial case where we optimize and load the departing train $t$ only, while $\tilde{\alpha}(t) = (1, 1, 0, \ldots)$ corresponds to optimizing two trains $t$, $t+1$ together and loading $t$ only. Ideally, we want to use $\tilde{\alpha}(t)$ such that the objective in (10) is a weighted average of short-horizon and long-horizon objectives. To achieve this, we propose to use exponentially decreasing weights:

$\tilde{\alpha}(t) = (1, \alpha_t, \alpha_t^2, \ldots)$ for $0 < \alpha_t < 1$,

where $\alpha_t$ is a scalar used to define vector $\tilde{\alpha}(t)$.

Then, (10) becomes

$$\text{Min } \sum_{s=t}^{t+\tau} (\alpha_t)^{s-t} z_s = \quad (11)$$
$$\min\left[(\alpha_t)^{\tau}(z_t + \ldots + z_{t+\tau}) + \sum_{r=0}^{\tau-1}(1-\alpha_t)(\alpha_t)^{r}(z_t + \ldots + z_{t+r})\right]$$

which is a weighted average of $z_t$, $(z_t + z_{t+1})$, …, $\sum_{s=t}^{t+\tau} z_s$. If most load information is already known and there is little uncertainty in the future, we should choose $\alpha_t \approx 1$, such that $\tilde{\alpha}(t) \approx (1, 1, 1, \ldots)$ and $\sum_{s=t}^{t+\tau} \alpha_{t,s} z_s \approx \sum_{s=t}^{t+\tau} z_s$, to exploit the efficiency from optimizing multiple trains together. On the other hand, if future loads are highly uncertain, we should choose $\alpha_t \approx 0$, such that $\tilde{\alpha}(t) \approx (1, 0, 0, \ldots)$ and $\sum_{s=t}^{t+\tau} \alpha_{t,s} z_s \approx z_t$, to avoid penalty due to the uncertain future.

Weight scalar $\alpha_t$ can vary across the train index $t$. Its appropriate value can be estimated over repeated field experiments or simulations for any existing intermodal facility. When empirical data are not available, a reasonable value must be estimated. Note from (11) that $\alpha_t$ reflects the relative importance of the short-term objective (regarding the train departing right away) based on static information, and that of long-term importance (future trains to be loaded) based on dynamic information. This concept is closely related to the "degree of dynamism" (DOD) defined in Lund et al *[31]* and Larsen *[28]* — the proportion of dynamic information at the time of decision. We propose to use 1 – DOD for the scalar weight; i.e.,

$$\alpha_t = \frac{n_{k,t}}{n_{k,t} + n_{u,t}} \quad (12)$$

where $n_{k,t}$ and $n_{u,t}$ are the numbers of unassigned loads known and unknown before loading the $t^{th}$ train, respectively.

## EMPIRICAL CASE STUDY

The cutoff times for loads to make the next departing train are generally 7-8 hours for international and domestic container stack trains, and 3-4 hours for domestic trailer trains *[32]*. The nature of advance information differs for containers versus trailers. It may often be possible to know about containers as much as 12 hours prior to departure, whereas information about trailers is generally not available until about 4 hours before departure. However, loading containers from storage area to assigned slot is more time consuming than loading trailers; so the loading assignments and process must start earlier if containers are involved. Based on these parameters and train departure intervals typical of a busy intermodal line operation with both domestic and international IM loads, we conducted an analysis of 6 trains with 1,380 loads in an 8-hr window.

We first apply the static model to evaluate aerodynamic efficiency obtained from optimal loading at the system level, assuming fully static information of trains and loads. Then, we use the dynamic model to carry out continuous terminal operations when information is dynamic.

**Static Benchmark Case**

In the static case, four scenarios were conducted to show the benefit of optimizing more trains together. They are to optimize one, two, three, or six trains at a time, assuming that all the information regarding these trains and loads is available. Optimizing one train at a time is consistent with current terminal practice. However, the more trains that are optimized at a time, the better the aerodynamic efficiency (Figure 4); although, the marginal benefit declines considerably beyond three trains. Depending on the particular train configuration, the potential fuel savings of the first scenario over a 2,200-mile BNSF Transcon route would be 1,500 gallons per train *[7]*. Since scenario 2, 3 and 4 are more beneficial than scenario 1, the fuel savings is also more significant.

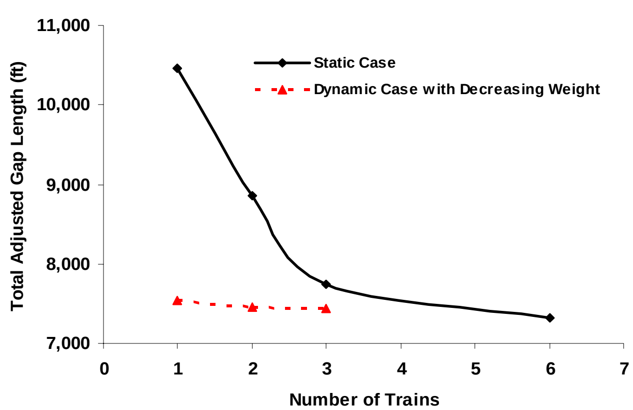



FIGURE 4 Effect of number of trains simultaneously optimized on adjusted gap length

In practice, loads often arrive and depart at terminals quickly with little lead time on their size or configuration. Each train has its own departure scheduled about every 70 ~ 90 minutes. Therefore, it is rarely the case that reliable load information will be available for more than 3 trains. Consequently, a rolling horizon framework is needed for continuous terminal operations such as this.

**Rolling Horizon Operations**

The same dataset in the static case (6 trains with 1,380 loads) is used. At the beginning of the 8-hour time window, 705 loads are available, while the rest becomes known at a constant rate of approximately 120 loads/hour. The degree of dynamism (DOD) of this system is almost equal to 0.5 at any time within the time window.

We start with a set of scenario analyses with six different cases; see Table 2. Scenario 1 is the base case in which only one train is optimized and loaded at a time. In scenario 2 we optimize two trains at a time and load only the first outgoing train. In scenario 3 we optimize two trains at a time and load both trains. The other scenarios follow the same logic.

We clearly see that scenario 3 is worse than scenario 2, and scenarios 5, 6 are worse than scenario 4. This verifies our qualitative analysis that it is not desirable to assign loads to trains other than the immediately outgoing one. We also see in this example that optimizing two trains together is better than optimizing three trains. This confirms our argument that making decision for future trains and loads may incur penalties due to future uncertainty. Overall, the best scenario is scenario 2, yielding a total adjusted gap length of 7,470 ft. It is already very close to the static benchmark from loading six trains with full information (7,317 ft).

Table 2 Scenario analysis of rolling horizon without decreasing weights

| Scenario | Operations | Adjusted Gap Length (ft) |
|---|---|---|
| Scenario 1 | opt 1 train, load 1 train | 7,539 |
| Scenario 2 | opt 2 train, load 1 train | 7,470 |
| Scenario 3 | opt 2 train, load 2 train | 7,535 |
| Scenario 4 | opt 3 train, load 1 train | 7,713 |
| Scenario 5 | opt 3 train, load 2 train | 7,828 |
| Scenario 6 | opt 3 train, load 3 train | 8,851 |

Finally, we test the exponentially smoothing approach with the same data. Since DOD≈0.5 at all time, we use $\alpha_t = 1 - 0.5 = 0.5$, such that $\tilde{\alpha}(t) = (1, 0.5, 0.25, \ldots)$, $\forall t$. Optimizing (10) yields an adjusted gap length of 7,445 ft, which is closer to the static benchmark case with full information (Figure 4).

## DISCUSSION

To compare the performance of the optimal loading patterns and the current operations, we constructed a random loading simulation and applied it to the same example. In this case, the average adjusted gap length of the random assignments was 9,578 (ft). Our optimal result from the realistic rolling horizon case is 22.27% lower than this current case. This shows a substantial benefit from optimizing the aerodynamic efficiency of the loading pattern of intermodal trains.

The objective of our optimization model – "matching intermodal loads with cars" -- has not currently been widely adopted in terminal operations. This is not because it requires significant changes in operations but because there has been little understanding of the benefit prior to the energy efficiency research in Lai & Barkan [18]. The model and solution technique proposed in this paper would not increase the operating costs at the rail terminals because there are no additional operational constraints or requirements.

Rail intermodal business in North American is different from the general freight business. Railroads try hard to avoid intermediate switching and stops because the intermodal business is highly time sensitive. For example, approximately 80% of the intermodal trains on the BNSF Transcon route have no intermediate operations. Most of the other 20% have no more than 2 intermediate stops and these are generally close to the final destination. Therefore, in this work, we focused on the intermodal services between Chicago and LA assuming that no intermediate yard operations were present. If intermediate operations are necessary, the model can be modified to incorporate the time value of adjusted gap lengths in the objective function. It is likely that the resulting problem will become too complex to solve using integer programming. Further research should explore the possibility of using heuristic methods to compute second best (slightly sub-optimum) loading patterns in a practical way.

## CONCLUSION

We extend a previously developed aerodynamic efficiency model [7] to be able to optimize aerodynamic efficiency at the system level, and develop a rolling horizon scheme for continuous terminal operation. For the deterministic case, the system optimum can be reached by optimizing as many trains as possible; however, terminals actually operate in a stochastic environment due to uncertainty regarding incoming loads and trains. Attempting to optimize the loading of too many trains in this environment will reduce the ability to achieve the most efficient loading configuration because of imperfect information. Therefore, a modified model with descending weight assigned to each train is proposed to counterbalance the effect of uncertainty. Appropriate weights are determined by simulations based on real data from an existing intermodal facility. Depending on the particular train configuration, the potential fuel savings of optimizing one train at a time over a 2,200-mile BNSF Transcon route would be 1,500 gallons per train. Since using the dynamic model is even more beneficial, the necessary additional planning or handling may be worthwhile.

## ACKNOWLEDGEMENTS

The authors are grateful to Mark Stehly, and Larry Milhon of the BNSF Railway, for their support and help on this project. The first author was supported by a research grant from the BNSF Railway and a CN Research Fellowship in Railroad Engineering at the University of Illinois.

## REFERENCES

1. Gallamore, R.E. *State of the Art of Intermodal Freight Transport in the United States*. Intermodal Freight Transport in Europe and the United States, Eno Transportation Foundation, Inc., Lansdowne, 1998.

2. AAR. *Yearbook of Railroad Facts*. Association of American Railroads, 2005.

3. Engdahl R., R. L. Gielow, and J. C. Paul. *Train Resistance - Aerodynamics Volume I of II Intermodal Car Application.* Proceedings of Railroad Energy Technology Conference II, Association of American Railroad, 1986, pp. 225-242.

4. Engdahl R., R. L. Gielow, and J. C. Paul. *Train Resistance - Aerodynamics Volume II of II Open top Car Application*. Proceedings of Railroad Energy Technology Conference II, Association of American Railroad, 1986, pp. 243-266.

5. UP. *Intermodal Loading Guide*. Union Pacific Railroad Corporation. http://www.uprr.com/customers/dam-prev/loading/intguide/ Accessed July 31, 2006.

6. Optimization Alternatives Ltd. Inc. *Optimization Alternatives' Strategic Intermodal Scheduler (OASIS).* Optimization Alternatives Ltd. Inc. http://www.oax.com Accessed July 31, 2006.

7. Lai, Y.C., C.P.L. Barkan and H. Onal. *Optimizing the Aerodynamic Efficiency of Intermodal Freight Trains*, Working Paper, 2006.

8. AAR, Analysis of Class I Railroads 2005, Association of American Railroads, 2006.

9. BNSF. *Every Drop Counts in Goal to Improve Fuel Efficiency*. The Burlington Northern Santa Fe Railway Company. www.bnsf.com/media/articles/2004/03/2004-03-05-b.html? index=/media/articles/index.html Accessed July 15, 2004.

10. AAR. *Railroad Energy Technology: The Alternatives*. Proceedings of Railroad Energy Technology Conference I, Association of American Railroads, 1982, pp. 179-204.

11. Smith M. E. *Economics of Reducing Train Resistance*. Proceedings of Railroad Energy Technology Conference II, Association of American Railroad, 1986, pp. 269-305.

12. BNSF. *Intermodal Loading Guide.* The Burlington Northern and Santa Fe Railway Company, 2004.

13. Avriel M., M. Penn, N. Shpirer, and S. Witteboon. Stowage planning for container ships to reduce the number of shifts. *Annals of Operations Research* 76, 1998, pp. 55-71.

14. Giemsch P. and A. Jellinghaus. Optimization Models for the Containership Stowage Problem. *Operations Research Proceedings*, Heidelberg, 2003.

15. Utterback M. Personal Communication. The Burlington Northern Santa Fe Railway Company, 2006.

16. Hay W. W. *Railroad Engineering*, Second Edition ed, 1982.

17. AREMA, *Manual for Railway Engineering. Vol 2, Train Performance*, American Railway

Engineering and Maintenance-of-Way Association, Landover, 2001.

18. Lai Y.C. and C.P.L. Barkan. Options for Improving the Energy Efficiency of Intermodal Freight Trains, *Transportation Research Record 1916*, Transportation Research Board, 2005, pp. 47-55.

19. Gielow M. A., and C. F. Furlong. *Results of Wind Tunnel and Full-Scale Tests Conducted from 1983 to 1987 in Support of The Association of American Railroad's Train Energy Program*. Publication R-685. Association of American Railroad, 1988.

20. Drish W. F. *Train Energy Model Version 2.0 Technical Manual*. Publication SD-040. Association of American Railroads, 1992.

21. Engdahl R. *Full-Scale Rail Car Testing to Determine The Effect of Position-in-Train on Aerodynamic Resistance*. Publication R-705. Association of American Railroad, 1987.

22. AAR, *Loading Capabilities Guide,* Association of American Railroads, Washington, DC, 2004.

23. Furlong C. F. *Aerodynamic Subroutine Users Guide*. Publication R-683. Association of American Railroads, 1988.

24. Storm B. Personal Communication, NASA Ames Research Center, Moffett Field, CA, 2005.

25. Airflow Science, *Aerodynamic Subroutine Version 4.0,* Association of American Railroads, Washington D.C., 1988.

26. Gendreau, M. and J.-Y. Potvin, Dynamic Vehicle Routing and Dispatching, In Fleet Management and Logistics, T.G. Crainic and G. Laporte (Eds.), Kluwer Academic Publishers, Norwell, MA, 1998, pp. 115-126.

27. G. Ghiani, F. Guerriero, G. Laporte, R. Musmanno, Real-time vehicle routing: Solution concepts, algorithms and parallel computing strategies, *European Journal of Operational Research 151(1)*, 2004, pp. 1-11.

28. Larsen A. *The dynamic vehicle routing problem*, Ph.D. Thesis, Institute of Mathematical Modelling, Technical University of Denmark, 2001.

29. Psaraftis, H. Dynamic Vehicle Routing Problems, *Vehicle Routing: Methods and Studies*, Golden, BL and AA Assad (eds.). North-Holland, Amsterdam, 1988.

30. Mitrovic-Minic S., R. Krishnamurti, and G. Laporte, Double-horizon based heuristics for the dynamic pickup and delivery problem with time windows, *Transportation Research Part B* 38, 2004, pp. 669-685.

31. K. Lund, O. B. G. Madsen, and J. M. Rygaard, *Vehicle routing problems with varying degrees of dynamism*, Technical report, Institute of Mathematical Modelling, Technical University of Denmark, 1998.

32. Anderson F. Personal Communication. The Burlington Northern Santa Fe Railway Company, 2006.