INCREASING RAILWAY EFFICIENCY AND CAPACITY THROUGH IMPROVED
OPERATIONS, CONTROL AND PLANNING

BY

YUNG-CHENG LAI

B.S., National Taiwan University, 2002
M.S., University of Illinois at Urbana-Champaign, 2004

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2008

Urbana, Illinois

Doctoral Committee:

Professor Christopher P.L. Barkan, Chair
Professor Hayri Önal
Assistant Professor Yanfeng Ouyang
Professor Ravindra K. Ahuja, University of Florida

ABSTRACT

INCREASING RAILWAY EFFICIENCY AND CAPACITY THROUGH IMPROVED OPERATIONS, CONTROL AND PLANNING

Yung-Cheng Lai, Ph.D.
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
Christopher P. L.Barkan, Ph.D., Advisor

The focus of my dissertation is to increase railway efficiency and capacity through improved operations, control and planning. Various analytical approaches were developed and carried out using operations research techniques, and capacity analysis methodologies for optimization of the aerodynamic efficiency of intermodal freight trains and railway capacity planning.

Intermodal freight recently surpassed coal to become the largest source of revenue for US freight railroads. However, intermodal trains are among the least fuel efficient types of freight trains operated. This is due to a combination of the physical constraints imposed by the loads and railcar designs, which tend to have poor aerodynamics, and the high operating speeds required for this traffic. The importance of intermodal freight and these fuel efficiency characteristics suggest that investigation of options to improve aerodynamic efficiency has important economic and environmental implications for this segment of rail freight traffic.

In chapter 2, I consider basic options for improving the energy efficiency of intermodal trains. It is concluded that slot utilization, slot efficiency, and filling empty slots all have beneficial effects. A train can be more efficiently operated if loads are assigned not only based on slot utilization but also better matching of intermodal loads with cars, which we term "slot efficiency". I then develop the Aerodynamic Loading Assignment Model (ALAM), an integer programming (IP) model that incorporates aerodynamic characteristics of loads and railcar combinations to enable optimization of loading patterns to maximize fuel efficiency (chapter 3). Several policy recommendations regarding railway intermodal operations are also developed in chapter 3 based on a series of scenario analyses. The potential annual savings in fuel consumption through use of ALAM by one large railroad on one of its major intermodal routes is estimated to be approximately 15 million gallons with a corresponding value in 2007 of 29 million dollars.

ALAM was further developed to optimize multiple trains simultaneously if advance information about outgoing trains and loads is available. In chapter 4, I present static and dynamic aerodynamic efficiency models for the loading of multiple intermodal trains with a rolling horizon scheme for continuous train terminal operations. Numerical results show that the rolling horizon scheme significantly reduces the adjusted gap length compared to current practice, thereby leading to further improvement in the aerodynamic efficiency of intermodal trains. Correspondingly greater savings in fuel, emissions and expense are possible if this methodology is applied to all North American intermodal trains.

Railways around the world are facing capacity constraints. In North America, railway freight traffic has increased nearly 30% over the past 10 years, and this demand is projected to increase another 88% by 2035. It is clear that network capacity must be increased using various engineering options to upgrade the infrastructure, but the question is how railroads can allocate these investments in the best possible way to maximize network capacity.

In chapter 5, I develop a new decision support framework to help capacity planners determine how to optimize the allocation of capital investment for capacity expansion projects. This framework has three stand-alone tools: (1) an "Alternatives Generator" that enumerates possible expansion options along with their cost and capacity effects; (2) an "Investment Selection Model" that determines which portions of the network (at the subdivision level) need to be upgraded with what kind of capacity improvement options; and (3) an "Impact Analysis Module" that evaluates the tradeoff between capital investment and delay cost. Based on network characteristics, estimated future demand, and available budget, the proposed decision support framework can successfully determine the optimal solution regarding which subdivisions need to be upgraded and what kind of engineering options should be conducted. This will help railroads maximize their return from capacity expansion projects and thus be better able to provide reliable service to their customers, and return on shareholder investment. Such a decision support framework can be used to optimize the efficiency and effectiveness of railroad capacity expansion programs.

To Father, Mother, Brother, and My Fiancée

ACKNOWLEDGEMENTS

This dissertation is dedicated to my parents, Hong-Wei Lai and Huai-Ling Li. As two great educators, they always made sure that I had the best learning environment and constantly broadened my horizons by providing me with up-to-date articles and books on a variety of subjects.   One of the most important recent influences they have had on me is the Buddhist Philosophy.   I fully apply the wisdom in the Heart Sutra to my daily life, and find myself living a stable and peaceful life – the key to my ability to conduct good research.

This dissertation is also dedicated to my brother, Chien-Ming Lai, who is also a civil engineer and has always been my role model.   Due to the five-year age difference, he was always ahead of me in his education and career.   He has tirelessly provided me with help and critical suggestions in my life.   I could not have accomplished what I have done without him – the most competent role model I could ever ask for.

This dissertation is also dedicated to my fiancée, Hsiu-Wei (Jasmine) Chang.   I found you on my first day in high school 16 years ago, and have never let you go. Thank you very much for providing me the greatest trust and support throughout the 16 years.   Without you, I could never have the strength to face the difficult challenges on the road ahead.   You are no doubt the most important treasure in my life.

I would like to express my most profound and honest gratitude to my advisor, Dr. Christopher P.L. Barkan, for critical encouragement and significant support; and for his

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Railways uniquely combine energy efficiency with high speed overland transport of heavy freight (Hay, 1977). The combination of a low friction guided pathway provided by railroad tracks allows low rolling resistance and the formulation of trains which reduce rolling resistance and aerodynamic resistance respectively. Furthermore, the strong roadbed allows heavy loads enabling high capacity vehicles and therefore substantial economies of scale (Hay, 1982). Nevertheless, competitive and environmental pressures mean that further improvement is needed (UIC, 2003).

Researching energy efficiency is particularly timely in light of recent increases in fuel prices and their impact on industry operating costs, and the need to conserve energy. Class I railroads spent more than $8 billion on fuel in 2006 making it their largest operating expense and fuel cost continues to increase (AAR, 2007); from 2002 to 2005, North American railroad fuel cost doubled, and since 1999 it is up by nearly a factor of three. This trend is impacting railroads all over the world, consequently fuel efficiency is more important than ever (Stodoloski, 2002; UIC, 2004; BNSF, 2004). The sharp increase in energy costs, combined with railways' growing interest in improving their role as an environmentally sustainable transport mode has stimulated renewed interest in research on all aspects of energy efficiency (Smith, 2003; Wierderkehr, 2004). This

includes investigation of technologies to improve the efficiency of motive power, recover kinetic energy of moving trains, energy efficient design of railway vehicles, more efficient operations, prevention of fuel spillage and various approaches to reducing train resistance (Stodoloski, 2002; UIC, 2004; Barkan, 2004).

The North American railroad industry is increasingly experiencing capacity problems, and these directly impact transportation efficiency. Many railroad mainlines are currently at or near capacity, and the future demand is projected to increase by 88% by 2035 (AASHTO, 2007). This would not be as important if alternative modes were able to handle the traffic but highway construction is not keeping up with the growth in demand, either. Furthermore, much of rail traffic is not economically transported by truck even if the capacity was available. Rail is also safer, more land-use and energy efficient, and has less environmental impact than highway transport. Therefore, public officials increasingly see rail as an alternative transport mode needed to handle the increasing freight traffic that will accompany sustained economic growth (TRB, 2006; ASSHTO, 2007). This raises the question, how can railroads handle additional traffic on a network that is already experiencing constrained capacity in many locations.

The focus of my dissertation research is investigation of how to increase railway efficiency and capacity through improved operations, control and planning. Analyses using operations research (OR) techniques, and capacity analysis methodologies were conducted for optimizing the aerodynamic efficiency of intermodal (IM) freight trains (Chapter 2, 3, & 4), and railway capacity planning (Chapter 5).

OR techniques have been applied to capacity planning research in various fields (Ordonez, 2004) including electric utilities (Murphy and Weiss, 1990), telecommunication (Balakrishnan et al., 1995; Riis and Andersen, 2004), manufacturing (Eppen et al., 1989), inventory management (Hsu, 2002), and transportation (Magnanti and Wong, 1984; Minoux, 1989).   Problems in different domains can generally be formulated as network flow models with additional decision variables to determine whether or where to increase arc capacities.   Although the basic structures are similar, the real elements within the models often differ.

Among the key elements that make railroads unique are that they are a single degree of freedom system (i.e. trains can move only forward or backward).   Traffic is generally handled over discrete, fixed elements in the network called blocks and trains can pass only at specific locations.   The benefits of a one degree of freedom system are its high capacity and safety, but the drawback is its inflexibility (Department for Transport, 2001).   As a result, the interactions among trains, infrastructure, and operational strategies must be highly structured and make railway problems different and more complicated compared to many other network systems.   Consequently, it is essential to understand these similarities and differences across different domains.

1.1 Motivation

IM freight recently surpassed coal to become the largest source of revenue of US railroad freight transportation (Gallamore, 1998; AAR, 2005).   However, IM trains are

the least fuel efficient trains due to the physical constraints imposed by the combination of loads and railcar design (Engdahl et al., 1987). Ironically, IM trains are typically the fastest freight trains operated. This calls for investigation of the aerodynamic effects and options to improve aerodynamic efficiency, which have important economic and environmental implications for rail freight transportation (AAR, 1981; Smith, 1987).

Railroads are approaching the limits of practical capacity, and estimated future demand is substantial. Railway line capacity can generally be improved through operations and/or engineering options. Operations options should be considered first because they are generally less expensive and more quickly implemented than building new infrastructure. However, the considerable increase in demand is unlikely to be satisfied by changing operating strategies alone. Hence, railway capacity must be increased by upgrading the infrastructure through multiyear capacity expansion projects.

The North American railroad industry generally relies on experienced personnel and simulation software to identify bottlenecks and propose alternatives to reduce congestion (HDR, 2003; CN, 2005; Vantuono, 2005). Experienced railroaders often identify good solutions, but this does not guarantee that all good alternatives have been evaluated or that the best one has been found. Furthermore, the aging demographics of the railroad industry means that many experienced capacity analysts will soon retire. Simulation can model a section of the network in great detail but it is not suitable for network capacity planning. Instead of solving the real problem, solutions based on corridor-based simulation analyses may move bottlenecks to other places in the network. The emphasis

of my work is to look at the network as a whole. I account for each link as well as the whole network, and account for different capacity upgrade options, their relative cost effectiveness, and possible budget constraints to determine a set of optimal solutions.

1.2 Objectives of Study

I intend to achieve the following objectives in my study.

1. Optimizing the aerodynamic efficiency of intermodal (IM) freight trains:

   a) Investigate options for improving the energy efficiency of IM trains

   b) Establish a scoring system to quantify the loading efficiency of IM trains

   c) Develop a loading assignment model to minimize the aerodynamic resistance of IM trains and hence their fuel consumption

   d) Develop a rolling horizon scheme to optimize multiple trains simultaneously for continuous terminal operations

2. Optimizing railway capacity planning:

   a) Review of capacity analysis methodologies

   b) Develop a framework for evaluating railway capacity expansion projects

      i. Establish the decision support framework of railway capacity planning

      ii. Develop an, Alternatives Generator (AG) tool, to automatically generate possible capacity expansion options for each link and estimate the corresponding cost and capacity improvement

iii. Develop an, Investment Selection Model (ISM), that uses optimality techniques to determine the best set of investment options at the network level

iv. Develop a sensitivity analysis procedure, the Impact Analysis Module (IAM), to evaluate the benefit of upgrading infrastructure

1.3 Contribution Summary

I summarize the potential contributions of my dissertation work below:

1. Contributions to Civil (Railroad) Engineering

    a) A loading assignment model that helps terminal managers make up more fuel efficient intermodal trains,

    b) A scoring system to monitor and determine the efficiency of the loading patterns of intermodal trains,

    c) An enhanced capacity evaluation tool that can be used for strategic planning,

    d) A new decision support framework to determine the best strategy for network capacity expansion.

2. Contributions to Capacity Research

    a) A review of railway capacity analysis methodologies to determine the appropriate model for the specific task,

    b) A complete parametric model that is able to generate possible capacity expansion options for each subdivision,

c) A sensitivity analysis procedure to evaluate the tradeoff between capital investment and delay cost.

3. Contributions to Operations Research

   a) First application of optimization techniques to improve the energy efficiency of intermodal trains,

   b) A new rolling horizon method to simultaneously optimize the energy efficiency of multiple IM trains for continuous terminal operations,

   c) A new investment selection model to determine the optimal set of investment options for railway capacity expansion projects.

1.4 Dissertation Organization

This dissertation is designed as a series of chapters, four of which are intended as individually publishable papers.   As a result, each chapter may cover some material discussed in previous chapters.   Chapters 2–5 form the core of the dissertation.   Chapter 6 presents a summary of findings, and identifies additional research needs.

Chapter 1: Introduction

This chapter presents the motivation, objectives, and summary of this dissertation.

Chapter 2: Options for Improving Energy Efficiency of IM Trains

This chapter investigates options to improve the energy efficiency of intermodal trains. Three options for improving intermodal train energy efficiency were evaluated; slot utilization, slot efficiency, and filling empty slots.   All of them have a beneficial effect.

Slot efficiency, in which intermodal loads are matched with cars of an appropriate length reduces the gap length between loads, thereby improving airflow. Compared to slot utilization, maximizing slot efficiency offers additional potential to reduce fuel consumption and intermodal train operating costs. Filling empty slots with empty containers or trailers also reduces aerodynamic resistance thereby improving energy efficiency. Despite the additional weight penalty and consequent increase in bearing and rolling resistance that would accrue, the reduction in aerodynamic resistance more than offsets at the speeds intermodal trains typically operate. Depending on the particular train configuration, I found that train resistance can be reduced by as much as 27% and fuel savings of approximately 1 gal per mile per train are possible.

This chapter was published as: Lai, Y.C., and Barkan, C.P.L. (2005) Options for Improving the Energy Efficiency of Intermodal Freight Trains, *Transportation Research Record*, Vol 1916, 47–55.

Chapter 3: Optimizing the Aerodynamic Efficiency of IM Freight Trains

This chapters presents an aerodynamic loading assignment model (ALAM) for intermodal freight trains based on an integer-programming framework to help terminal managers make up more fuel-efficient trains. This is the first use of optimization modeling to address the aerodynamics and energy efficiency of railroad intermodal trains. The model developed in this research can be adapted to a variety of other intermodal train loading assignment problems through modification of the objective function. This is a novel contribution to the literature and enhances its generality because the formulation can

be solved efficiently and thus serve as a basis for other intermodal load assignment problems.

There are substantial potential fuel and cost savings benefits that railroads can achieve thorough implementation of ALAM at intermodal terminals. These benefits can be further enhanced through several additional steps including: (a) better matching of railcars and loads for international intermodal trains; (b) simultaneous optimization of multiple trains to take greater advantage of the potential to improve the energy efficiency of intermodal trains through use of more aerodynamic loading patterns; and (c) uncoupling empty railcars from the end of loaded intermodal trains when practical. The potential annual savings in fuel consumption through use of ALAM by one large railroad on one of its major intermodal routes is estimated to be approximately 15 million gallons with a corresponding value in 2007 of 29 million dollars. Correspondingly larger savings in fuel, emissions and expense are possible if the methodology described in this chapter were applied to all North American intermodal trains.

This chapter was published as: Lai Y.C., Barkan CPL, and Önal H. (2007) Optimizing the Aerodynamic Efficiency of Intermodal Freight Trains, *Transportation Research Part E*, doi:10.1016/j.tre.2007.05.011.

__Chapter 4: Optimizing the Aerodynamic Efficiency of IM Freight Trains with__

__Rolling Horizon Operations__

In this chapter, ALAM was further developed to allow simultaneous optimization of multiple trains when advance information about outgoing trains and loads is available.   In this research, I first developed 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 in a stochastic environment in which complete information on future trains and incoming loads may not be available.   This study 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.   A rolling horizon scheme was developed 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 can be obtained using these optimization models.

This chapter was published as: Lai, Y.C., Ouyang, Y., and Barkan, C.P.L. (2008)   A Rolling Horizon Model to Optimize Aerodynamic Efficiency of Intermodal Freight Trains with Uncertainty, *Transportation Science*, Accepted.

__Chapter 5: Optimizing Railway Capacity Planning__

This chapter demonstrates a new decision support framework to help capacity planners determine how to optimize the allocation of capital investment for capacity

expansion projects.   This framework has three stand-alone tools: (1) an "Alternatives Generator" that enumerates possible expansion options along with their cost and capacity effects; (2) an "Investment Selection Model" that determines which portions of the network (at the subdivision level) need to be upgraded with what kind of capacity improvement options; and (3) an "Impact Analysis Module" that evaluates the tradeoff between capital investment and delay cost.   Based on network characteristics, estimated future demand, and available budget, the proposed decision support framework can successfully determine the optimal solution regarding which subdivisions need to be upgraded and what kind of engineering options should be conducted.   The decision support framework can help railroads maximize their return from capacity expansion projects and thus be better able to provide reliable service to their customers, and return on shareholder investment.

Chapter 6: Conclusions and Future Research

This chapter provides a summary of chapters 2–5, and proposed future research topics following by my research work in this dissertation.

CHAPTER 2

OPTIONS FOR IMPROVING THE ENERGY EFFICIENCY
OF INTERMODAL TRAINS

Intermodal (IM) freight is among the largest segments of the US railroad freight transportation, and is definitely the fastest growing portion (nearly 80% in last 15 years) (Gallamore, 1998; AAR, 2005). However, intermodal trains are the least fuel efficient trains. This inefficiency is due to the physical constraints imposed by the combination of loads and the railcar design (Engdahl et al., 1987). The pair of photographs (Figure 2.1) contrasts the close spacing of the hopper cars with the much larger gaps due to the empty slots in the intermodal train. The aerodynamic drag of the IM train is thus higher than that of the fully loaded coal train by 25%, and this difference increases exponentially with speed. It is particularly ironic given that intermodal trains are typically the fastest freight trains operated.

(a)

(b)

Figure 2.1:   Pictures of (a) general coal cars (b) trailers on spine cars

At intermodal terminals, containers or trailers are assigned to available well, spine or

flat cars (UP, 2004). Although computer software (Optimization Alternatives Ltd Inc,

2006) is often used by terminal managers to assist in this task, it is still a largely manual

process. The principal metric used to measure the efficiency of loading is "slot

utilization" (Buriss, 2003). Although the details vary depending upon the particular

combination of intermodal load and car being considered, slot utilization is basically a

metric used to measure the percentage of the spaces (a.k.a. slots) on intermodal cars that

are used for loads. For example, Figure 2.2a is an example of a 5-unit well car with

100% slot utilization because all 10 slots are filled with containers whereas the railcar in

Figure 2.2b has 80% slot utilization.

(a)

(b)
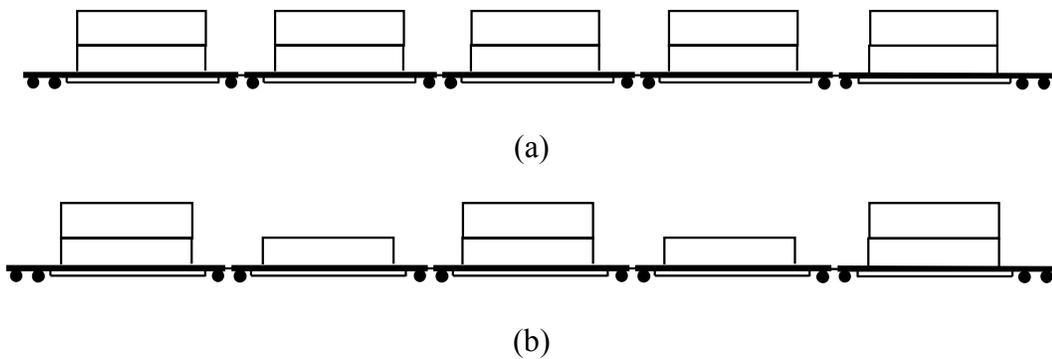
Figure 2.2:   (a) 100% slot utilization (b) 80% slot utilization

Slot utilization does not take into account the size of the space compared to the size

of the load. Although perfect slot utilization indicates maximal use of spaces available,

it is not intended to, nor does it ensure, that intermodal cars are loaded to maximize the

energy efficient operation of intermodal trains. Two trains may have identical slot

utilization, but different loading patterns and consequent aerodynamic resistances. For instance, both railcars in Figure 2.3 have 100% slot utilization but the railcar in Figure 2.3a is more aerodynamically efficient than the railcar in Figure 2.3b due to the shorter gap lengths.

(a)

(b)
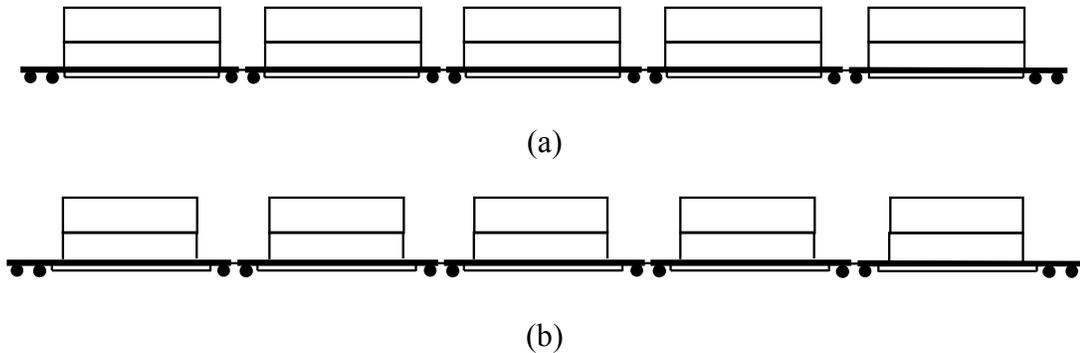
Figure 2.3:   (a) Loads match slots' capability (b) loads do not match slots' capability

During the 1980's, a number of studies focused on technologies to reduce train resistance and therefore reduce fuel costs (AAR, 1981; Smith, 1987). Aerodynamic drag was known to be a major component of the total tractive resistance particularly at higher speeds, so the Association of American Railroads (AAR) supported research on wind tunnel testing of rail equipment, including large-scale intermodal car models (Gielow and Furlong, 1988; Engdahl, 1987). The results were used to develop the Aerodynamic Subroutine of the AAR's Train Energy Model (TEM) (Drish, 1992).

From these wind tunnel tests, it was found that the lead locomotive experienced the highest drag and this decreased until about the 10th unit or car in the train, after which, drag remained roughly constant per unit for the remainder of the train. They also found

that closely-spaced containers or trailers behave as one long load. Conversely, as shown in Figure 2.4, loads spaced equal to or greater than 12' behave as distinct objects on whose surfaces boundary layers are reinitialized (Engdahl et al., 1987). Consequently, Engdahl et al (1987) suggested that filling empty slots with empty containers might have potential advantages; however, this suggestion did not consider the effect of the increased weight of the additional loads.
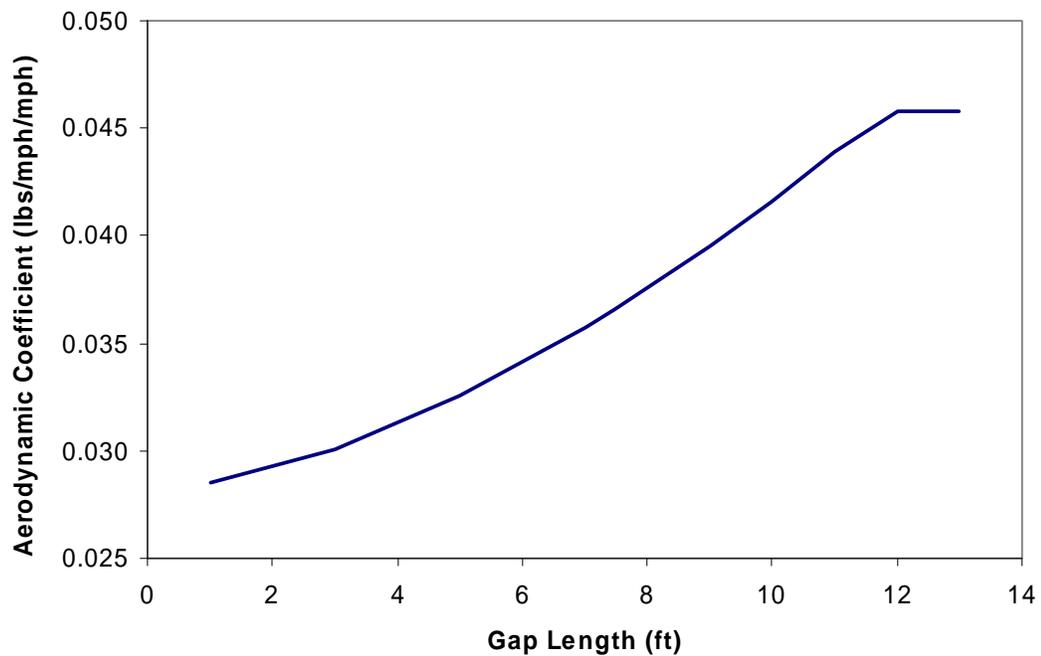
Figure 2.4: Critical gap length of well cars

Improving the loading patterns of intermodal trains has the potential to improve railroad fuel efficiency and reduce emissions. Maximizing slot utilization will enhance energy efficiency, but matching intermodal loads with appropriate length intermodal car slots can further reduce gap length between loads, and thus improve airflow. Filling

empty slots with empty containers or trailers also reduces aerodynamic resistance; however, the additional weight penalty generates more bearing and rolling resistance. In this chapter, we conducted a series of analyses to compare both the relative and absolute effects of different loading patterns and operating practices on train make-up and energy efficiency.

2.1 Train Resistance Calculation

Several approaches are considered to maximize intermodal train energy efficiency: slot utilization, improved equipment matching, and use of empty intermodal loads to improve train aerodynamics. Train resistance and the aerodynamic coefficient are computed for a series of different train scenarios using TEM (Drish, 1992) and the Aerodynamic Subroutine (Furlong, 1988). Train resistance is the sum of the forces opposing the movement of a train (Hay, 1982). The greater the resistance, the more energy is required to move the train. Therefore, it is a major factor affecting fuel economy.

The general expression for calculating train resistance is (Hay, 1982):

$$R = A + BV + CV^2 \qquad\qquad (2.1)$$

Where:

R = Train resistance (lbs)

V = Train speed (mph)

A = Bearing and rolling resistance independent of train speed (lbs)

B　= Coefficient used to define train resistance dependent on train speed

(lbs/mph)

C　= Aerodynamic coefficient (lbs/mph^2)

The B term is generally small and sometimes ignored (Gielow and Furlong, 1988; AREMA, 2001).　The C term can be computed from the Aerodynamic Subroutine by specifying a train consist.　For bearing and rolling resistance, the equations in TEM were used.　TEM requires input regarding bearing type and condition, and truck design and condition.　Based on information from railroads and intermodal equipment engineering personnel (TTX, 1999), we made the following assumptions in our analyses:

- 50% of the bearings are manufactured by Timken and the other 50% by Brenco

- 50% of the bearings are worn and the other 50% new

- 50% of the trucks are worn three-piece and the other 50% are new three-piece

- Ambient temperature is 60 ^{o}F

- No side wind effect (yaw angle = 0°)

According to these assumptions, the bearing resistance is calculated as follows (Drish, 1992):

$$R_{Bk} = n_k C_{Bk} \tag{2.2}$$

$$C_{Bk} = 6.2334 \times W_k^{0.20194} \tag{2.3}$$

Where:

k = Ordinal number of vehicle in consist

R_{Bk} = Bearing resistance acting on vehicle k (lbs)

n_k = Number of axles on vehicle k

C_{Bk} = Bearing resistance coefficient for vehicle k (lbs/axle)

W_k = Total weight for vehicle k (tons)

And the rolling resistance is computed as (Drish, 1992):

$$R_{Rk} = 0.0005 w_k C_{Rk}$$ (2.4)

Where:

R_{Rk} = Rolling resistance acting on vehicle k (lbs)

C_{Rk} = Rolling resistance coefficient for vehicle k (lbs/ton)

w_k = Total weight for vehicle k (lbs)

If $w_k < \gamma_k$

$$C_{RK} = 2.25 - [2.25 - \lambda] \left[\frac{w_k - \tau_k}{\gamma_k - \tau_k} \right]$$ (2.5)

Otherwise ($w_k \geq \gamma_k$)

$$C_{Rk} = \lambda$$ (2.6)

Where:

γ_k = Gross rail load for vehicle k (lbs)

τ_k = Vehicle k tare weight (lbs)

λ = Loaded rolling resistance coefficient for vehicle k (lbs/ton) (λ is 2.13 lbs/ton for three-piece worn truck and 1.57 lbs/ton for three-piece new truck.)

As a result, the resistance equation in this study can be represented as:

$$R = R_{Bk} + R_{Rk} + CV^2 \qquad (2.7)$$

2.2 Matching Intermodal Loads with Cars

A typical intermodal train has 3 locomotives and 80 to 120 units.   Therefore, a train of 3 locomotives and 100 units (20 5-unit cars) was chosen as suitably representative for our analyses.   The capacity of well and spine cars is usually constrained by the length of the loads.   For example, a 5-unit articulated double stack well car with a 40-foot well cannot handle containers greater than 40 feet long in the bottom position, whereas a 5-unit car with a 48-foot well can handle containers up to 48 feet in length.   Similarly, a 5-unit articulated spine car with 48-foot slots cannot handle containers or trailers greater than 48 feet, while a 5-unit car with 53-foot slots is able to handle trailers of any length up to 53 feet (UP, 2004; BNSF, 2004; TTX, 1999; Armstrong, 1998).   Consequently, cars with longer slots are more flexible; however, if loaded with trailers or containers less than the maximum they allow, then the gaps between loads are correspondingly larger,

and less aerodynamically efficient.   We conducted the following analyses to illustrate the potential differences in resistance for different train loading configurations.

In the case of the well car, a 40-foot container can be assigned to a car with either 40-foot or 48-foot wells; however, only use of a car with 40-foot wells would result in the shortest gap and the best aerodynamics.   In this example, the gap between two double stack 40-foot containers would increase by 8 ft if 48-foot-well cars were used (Figures 2.5a, 2.5b).

(a)

(b)

(c)

(d)
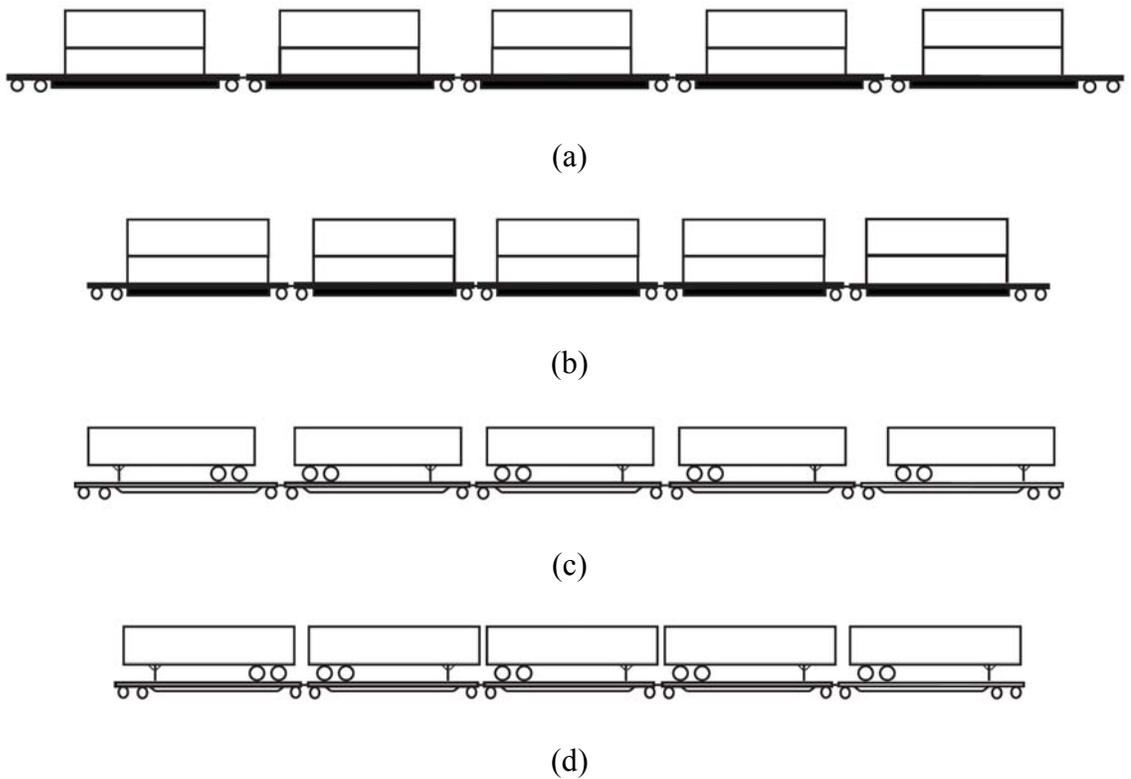
Figure 2.5:   (a) Double stack 40' containers in a 48-foot-well car; (b) double stack 40' containers in a 40-foot-well car; (c) 48' trailers in a 53-foot-slot spine car; (d) 48' trailers in a 48-foot-slot spine car

For a train of 20 cars with 40-foot double stack containers, the aerodynamic coefficient increases from 4.82 to 5.05 lbs/mph^2 when 48-foot-well cars are used instead of 40-foot.   We calculated the resistance for these two train configurations for speeds up to 70 mph.   As expected the train with 40-foot-well cars had lower resistance at all speeds (Figure 2.6a).   The difference in resistance for the 40-foot-well car with 40-foot containers, compared to the 48-foot-well car with 40-foot containers (Figure 2.6b), increases from 1.03% to 2.96% as speed increases to 70 mph.

Similarly, a 48-foot trailer can be placed on a spine car with either 48-foot or 53-foot slots.   The gaps between trailers are shortest when using cars with 48-foot-slots (Figures 2.5c, 2.5d).   For a train made up of 20 spine cars with 48-foot trailers, the aerodynamic coefficient increases from 5.90 to 9.12 lbs/mph^2 when 53-foot-slot cars are used instead of 48-foot-slot cars.   The difference in resistance ranges from 0.07% to 26.72% depending on speed (Figures 2.7a, 2.7b).

In the analyses above, each datum represents the effect on train resistance at a specific speed; however, a train's speed will actually vary as it traverses a route.   In addition to resistance, the power to ton ratio, route characteristics, and train schedule will all affect fuel consumption.   Therefore, the distribution of speed profiles and throttle setting is needed to more accurately estimate fuel saving.   TEM was used to compute and compare the fuel consumption for each case using a representative rail line.   A typical intermodal route in the Midwest was chosen for this analysis.   It is 103-miles in

length with gently rolling topography, grades generally under 0.6% and curves less than 3 degrees.
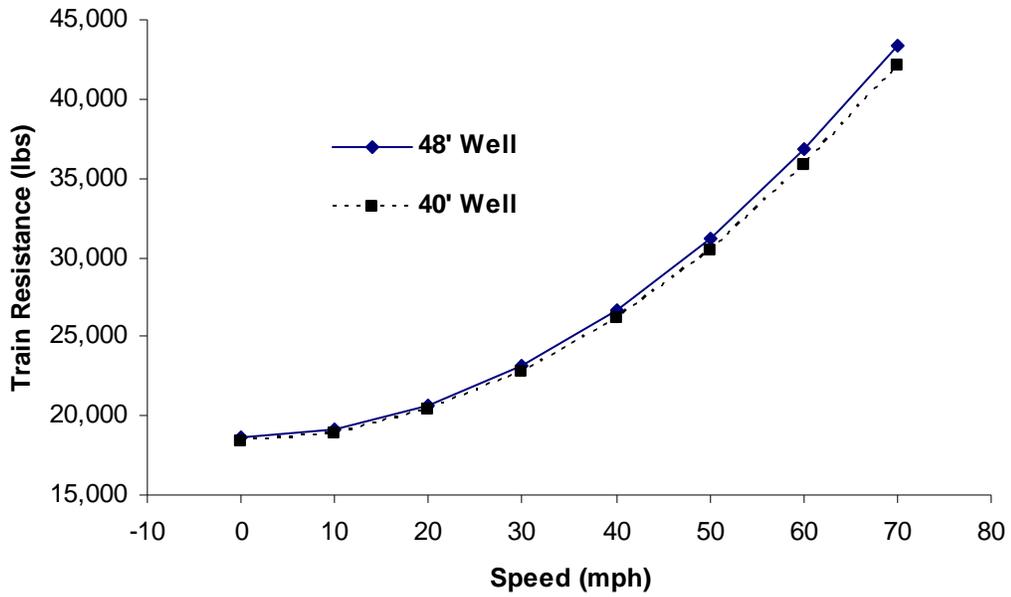
(a)
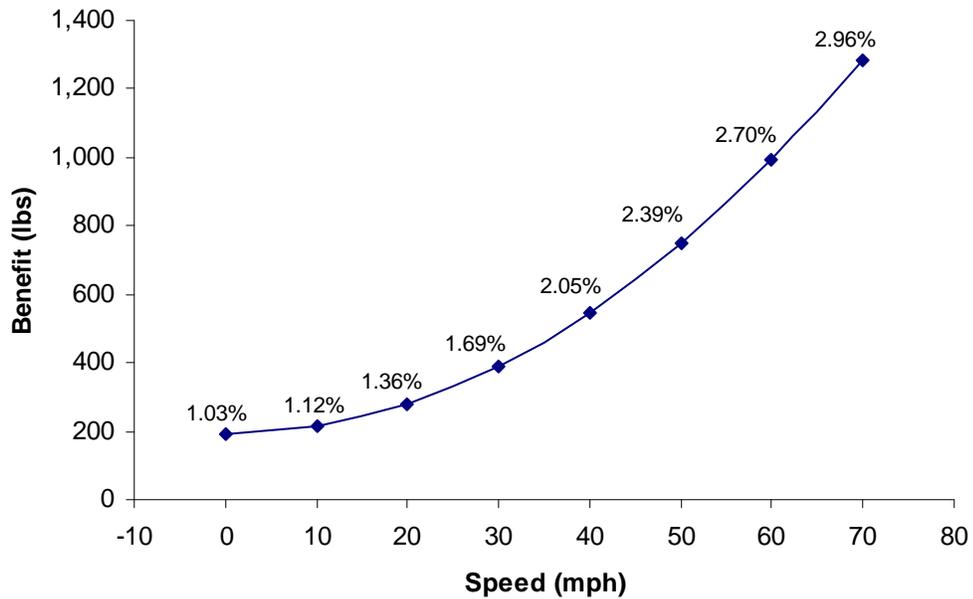
(b)

Figure 2.6:   (a) Resistance of 40' containers on 48-foot-well cars or 40-foot-well cars; (b) the benefit of using 40-foot-well cars
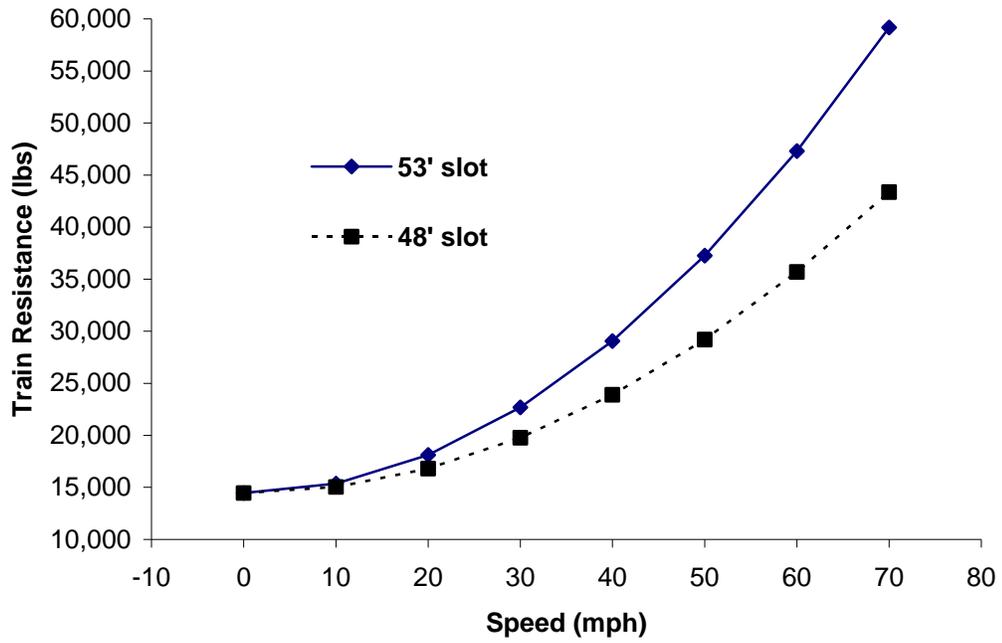
(a)
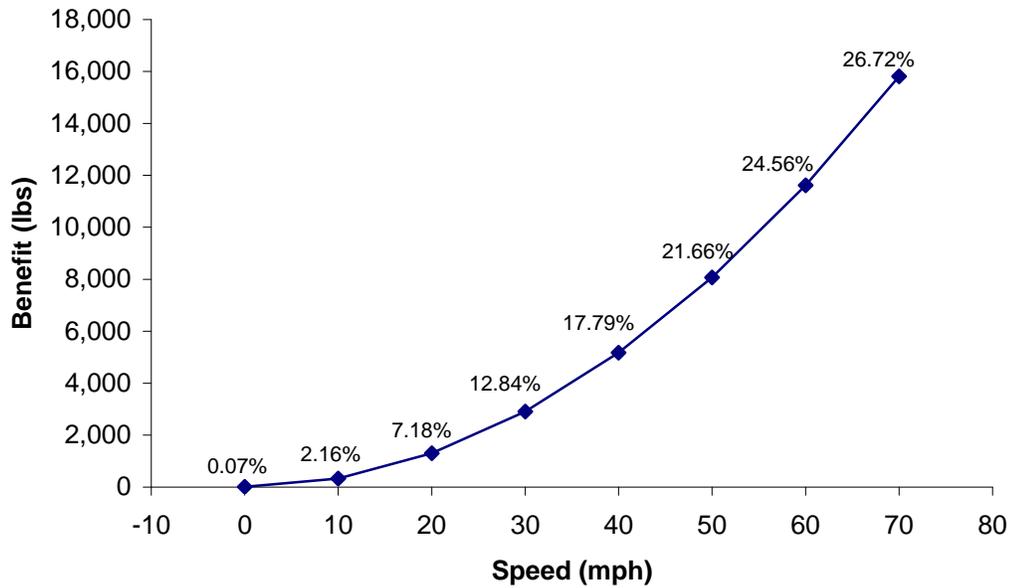
(b)

Figure 2.7:   (a) Resistance of 48' trailers on 53-foot-slot spine cars or 48-foot-slot spine

cars; (b) the benefit of using 48-foot-slot spine cars

In the first case, placing 40-foot double stack containers on cars with 40-foot wells would save 13 gallons of fuel per train on this route compared to cars with 48-foot wells. In the second case, placing 48-foot trailers in spine cars with 48-foot slots would save over 100 gallons of fuel per train.   The resultant fuel savings in the first case is 0.13 gal/mile, and in the second is over 1 gal/mile.   The reason for the difference is because in the first case the gaps are reduced in length, whereas in the latter case the gaps are almost completely eliminated.

2.3   Slot Utilization vs. Equipment Matching

Maximizing slot utilization has a positive effect on train energy efficiency because it eliminates empty slots and the consequent large gaps that would otherwise occur. However, as should be evident from the prior example in which all the trains considered had 100% slot utilization there is still the potential for substantial improvement in efficiency depending on the specific load-and-car combinations that are used.   Simply maximizing slot utilization does not ensure that the lowest aerodynamic resistance is achieved, whereas proper matching of intermodal loads with cars can.   Consequently, matching is a better metric for energy efficiency than slot utilization.

For example, the aerodynamic coefficient for a train of 20 48-foot-well cars loaded with 40-foot containers will be reduced by 23% if slot utilization is improved from 90% to 100% (Figure 2.8).   However, if the 48-foot-well cars are replaced with 40-foot-well

cars, the aerodynamic coefficient would be reduced by another 5%.   Note that in both

cases, slot utilization is 100%.

Similarly, the aerodynamic coefficient decreases by 3% if slot utilization is increased

from 90% to 100% for a train of 20 53-foot-slot spine cars with 48-foot trailers (Figure

2.8).   Replacing 53-foot-slot spine cars with 48-foot-slot spine cars reduces the

aerodynamic coefficient by another 36%.
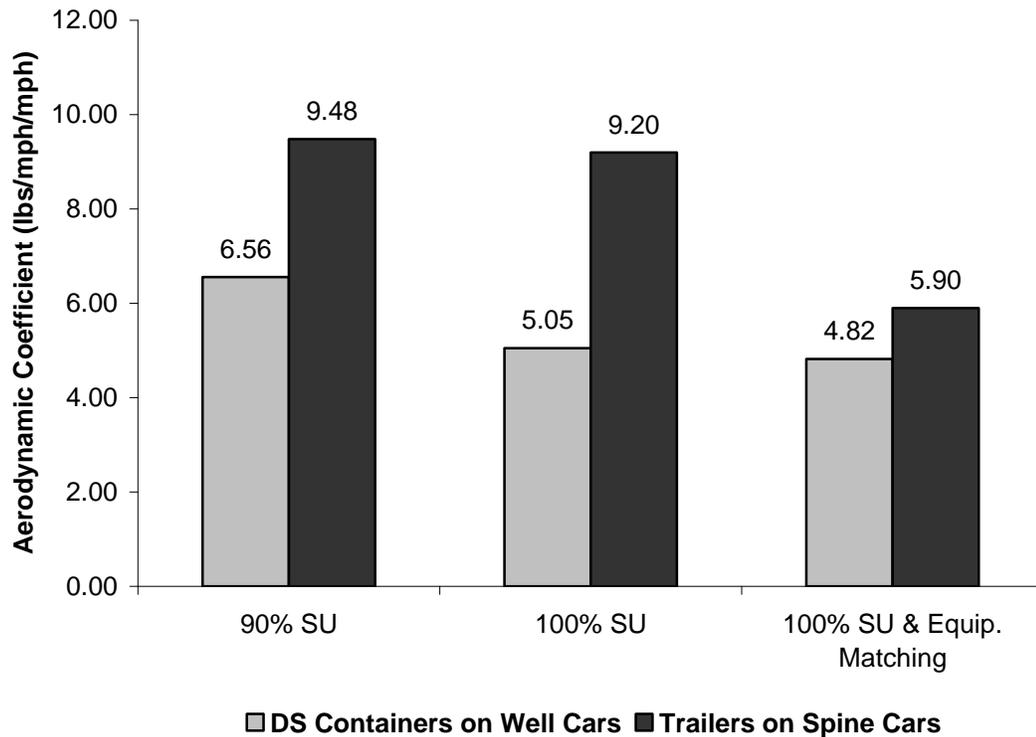
Figure 2.8:   The aerodynamic coefficient of 90% slot utilization (SU), 100% slot

utilization or equipment matching for double stack containers on well cars and trailers on

spine cars

Accordingly, a train can be more efficiently operated if loads are assigned not only based on slot utilization but also better matching of intermodal loads with cars, which we term "slot efficiency" (Milhon, 2004).   This effect will be especially pronounced for the units in the front of the train where the aerodynamic effect is greater.   Therefore, I proposed to use "slot efficiency" to evaluate the loading efficiency of IM train; it represents the difference between the actual and ideal loading configuration given the particular set of railcars in the train and the loads available.   The slot efficiency of each slot is calculated as follows:

$$Slot\ Efficiency\ =\ \frac{Length\ of\ Actual\ Load}{Length\ of\ Ideal\ Load} \times 100\% \qquad (2.7)$$

For example, the slot efficiency of a 45-ft trailer on a 53-ft-slot spine car unit is 85%, whereas placement of a 53-ft trailer on a 53-ft-slot spine car unit generates the lowest aerodynamic resistance and thus the highest score (100%) for this size slot.   Slot efficiency is similar to slot utilization except that it also factors in the energy efficiency of the load/slot combination.

2.4 Filling Empty Slots with Empty Loads

We recorded the load configuration of over 30 intermodal trains on a high density intermodal route of a Class 1 railroad and observed that empty slots usually occur as a single container on a well car, or an empty slot on a spine car.   Therefore, three different loading combinations were analyzed in three scenarios to evaluate the effect of placing empty loads in empty slots (Figure 2.9).   They are double stack containers on well cars,

trailers on spine cars and containers on spine cars.   For each scenario, the "baseline"

case represents empty slots, and the "alternative" case represents filling empty slots with

empty intermodal loads.

Both the number of empty slots and train speed affect the train resistance

computation.   In the following analysis of three scenarios train speed was held constant

at 50 mph.   Resistance values were computed for each case by changing the number of

discrete empty slots.   We restricted these changes to the last 90 units of the train to avoid

the complicating effects of factoring in the different aerodynamic effects characteristic of

the front of the train (Gielow and Furlong, 1988).   In this respect, our results understate

the potential benefits to a small extent because the aerodynamic benefit of improvements

in the front of the train is slightly higher.

(a)

(b)

(c)
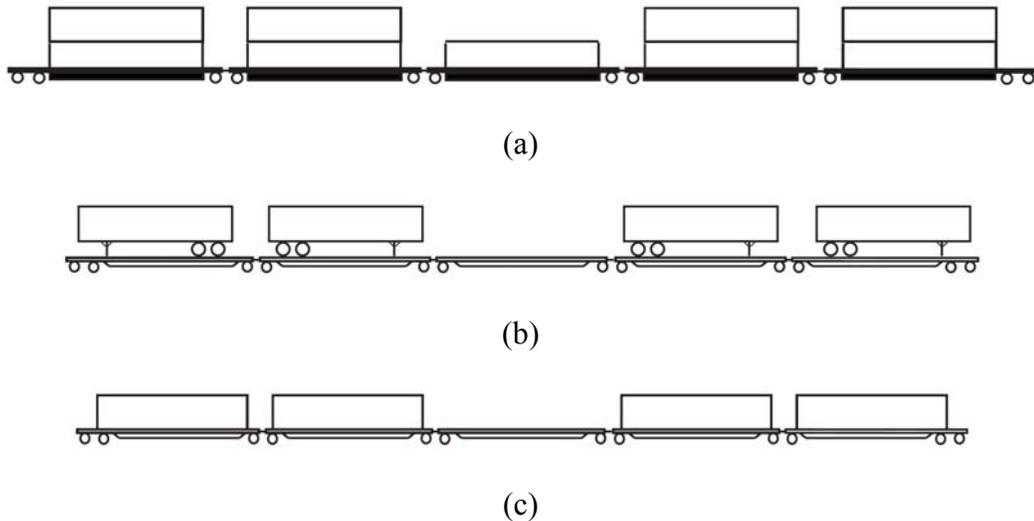
Figure 2.9:   (a) Containers on 5-unit articulated well car (scenario 1); (b) trailers on

5-unit articulated spine car (scenario 2); (c) containers on 5-unit articulated spine car

(scenario 3)

2.4.1 Scenario 1: Double-Stack Containers on Well Cars

In scenario 1, the train consists of 20 5-unit articulated 48-foot-well cars with 48-foot double stack containers and from one to ten empty slots.   A single empty slot in a 5-unit well car is shown in Figure 2.9a.   We compared the baseline condition with empty slots to the alternative condition, in which empty containers are placed in the previously empty slots (Figure 2.10).
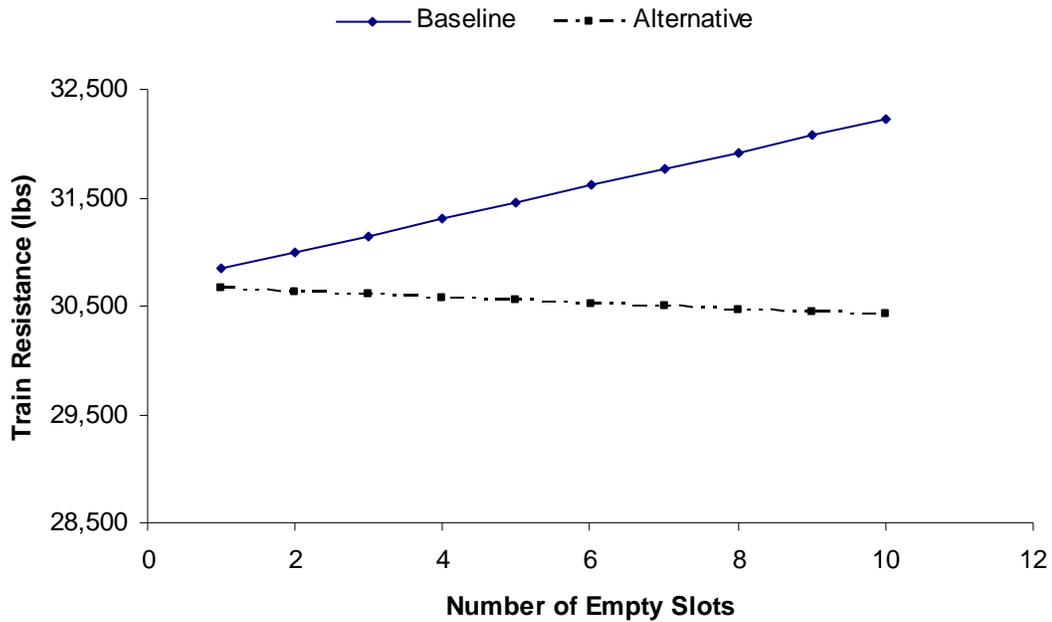
Figure 2.10:   Resistance of placing or not placing empty containers on empty slots in well cars

In the baseline case, the greater the number of empty slots, the higher the train resistance despite the reduction in train weight.   This is due to the increased number of large gaps and the consequent greater turbulence.   In the alternative case, empty

containers are placed in the open slots.   The results in a small increase in bearing and

rolling resistance due to the extra weight, but this resistance is more than offset by the

reduction in aerodynamic resistance.   This result is an inverse relationship between

resistance and the number of empty slots filled with empty containers because of both the

lighter train and improved aerodynamics.   The benefit is the difference between the

baseline and alternative cases, and it increases with the number of empty slots filled with

empty containers.   For all the conditions, the alternative method results in a reduction in

resistance (Figure 2.11).
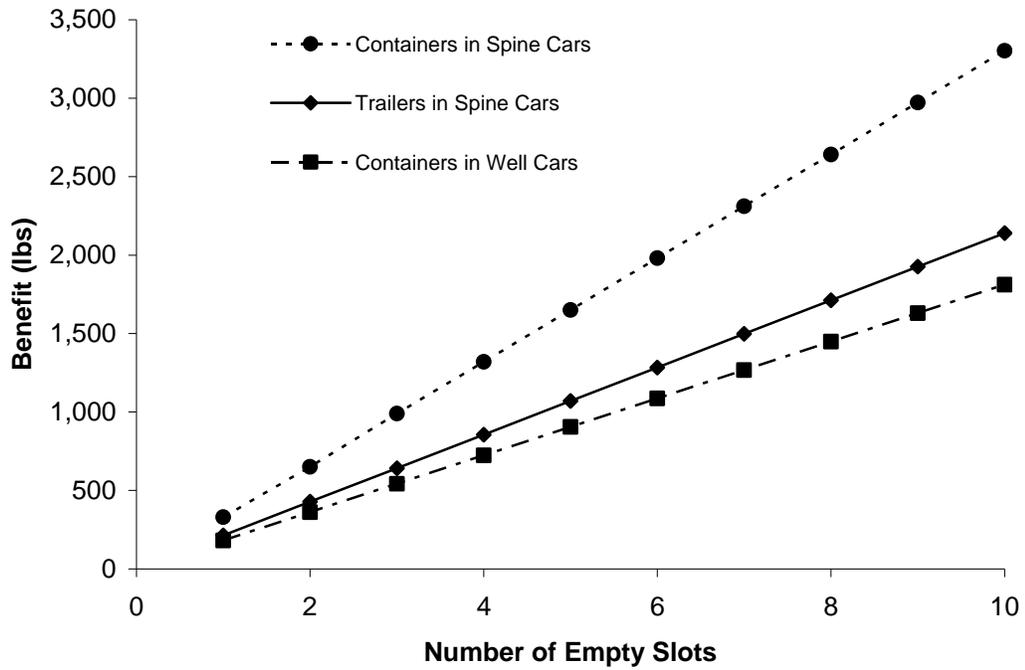
Figure 2.11:   The benefit of placing empty containers on empty slots in well cars,

placing empty trailers on empty slots in spine cars, and placing empty containers on

empty slots in spine cars

2.4.2 Scenario 2: Trailers on Spine Cars

In scenario 2, the train consists of 20 5-unit articulated 48-foot-slot spine cars with 48-foot trailers. As in scenario 1, the number of cars with empty slots was varied from one to ten (an example of a car with a single empty slot is shown in Figure 2.9b). The resistance of the baseline condition is compared to the alternative in which empty trailers are placed on empty slots on spine cars (Figure 2.12).
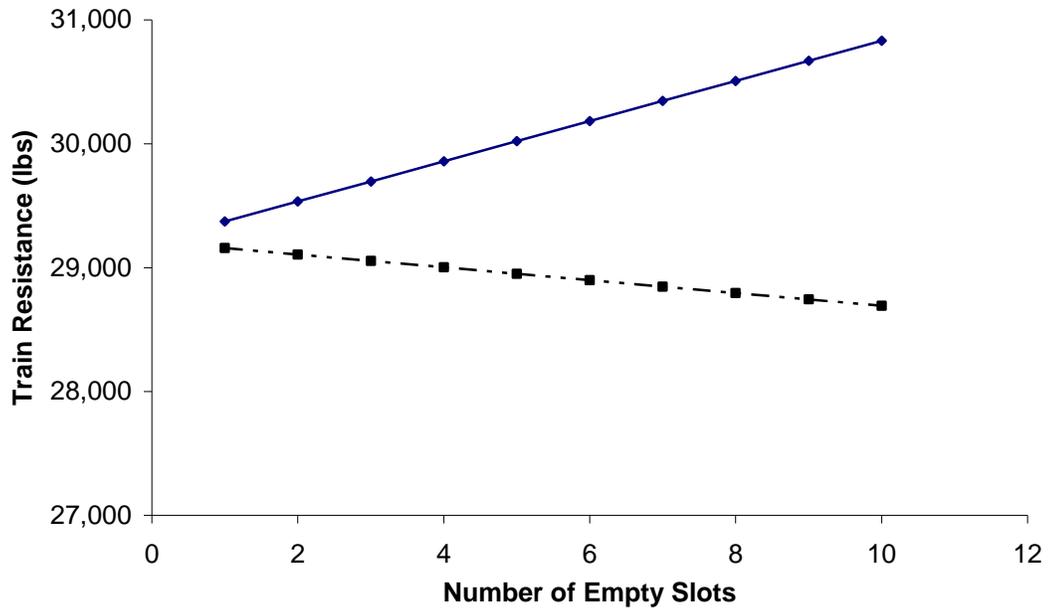
Figure 2.12: Resistance of placing or not placing empty trailers on empty slots in spine cars

The resistance of the baseline case increases and that of alternative declines with the greater number of empty slots, and the overall benefit increases with number of empty slots filled with empty trailers. As in scenario 1, the alternative method reduced resistance for all the conditions (Figure 2.11). The values are consistently higher than

for comparable numbers of empty slots in scenario 1. As discussed above, this is because the spacing between trailers on spine cars is closer than the spacing between containers in well cars, consequently the difference in the aerodynamics is greater between the baseline and alternative conditions. In fact, it is possible to space trailers so closely as to appear as one continuous body.

2.4.3 Scenario 3: Containers on Spine Cars

In scenario 3, the train consists of 20 5-unit articulated 48-foot-slot spine cars with 48-foot containers. The number of empty slots is varied as in the previous scenarios (Figure 2.9c). Again, the resistance of the baseline case increases and that of alternative case decreases as the number of empty slots goes up (Figure 2.13).
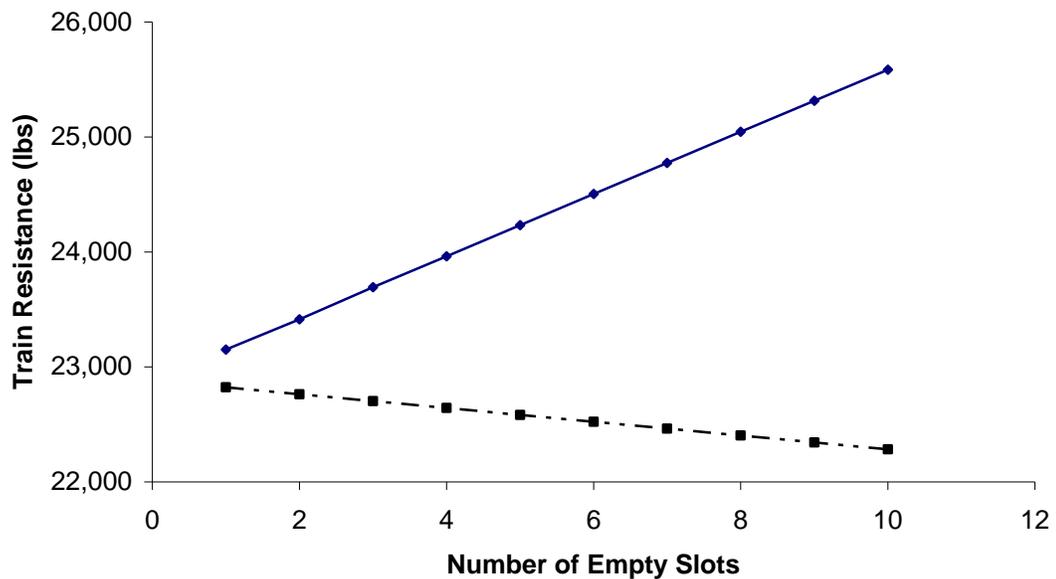
Figure 2.13: Resistance of placing or not placing empty containers on empty slots in spine cars

In this scenario, the corresponding benefits are even higher than in scenarios 1 & 2 (Figure 2.11).   There is more benefit from filling empty slots with empty containers on spine cars compared to the other two train configurations.   This is not only because the spacing between containers on spine cars is closer than on well cars; but also because closely-spaced containers can be regarded as a single long box. By contrast even closely spaced trailers create drag due to the presence of the hitch, trailer landing gear, and wheels below the floor of the trailer.

2.4.4   Effect of Speed

The scenario analyses demonstrated the effect of the number of empty slots at a single speed (50 mph).   The aerodynamic term in the train resistance model is a squared function of speed.   Consequently, we expect a greater aerodynamic benefit at higher speeds.   Conversely at lower speeds the relative benefit is expected to be smaller.   We conducted sensitivity analyses on the effect of train speed on resistance while holding the number of empty slots constant (Figures 2.14 & 2.15).

We first consider a train configured as in scenario 1 with 20 double stack well cars and five empty slots in the train.   Figure 2.14 compares the resistance of the train in the baseline and alternative conditions as a function of speed.   The resistance in both increases exponentially with speed and the difference between them also increases.   If the number of empty slots in the train is increased to ten, the resistance and corresponding benefit is also greater (Figures 2.15, 2.16a, 2.16d).   There is no benefit

when the speed is less than 10 mph because the reduction in aerodynamic resistance is
not enough to offset the increase in bearing and rolling resistance due to the extra weight
of the empty intermodal loads.   However, above this speed, there is a net benefit that
increases with speed so that at 70 mph, filling five empty slots with containers results in a
4% reduction in train resistance, and filling ten empty slots reduces train resistance by
8%.
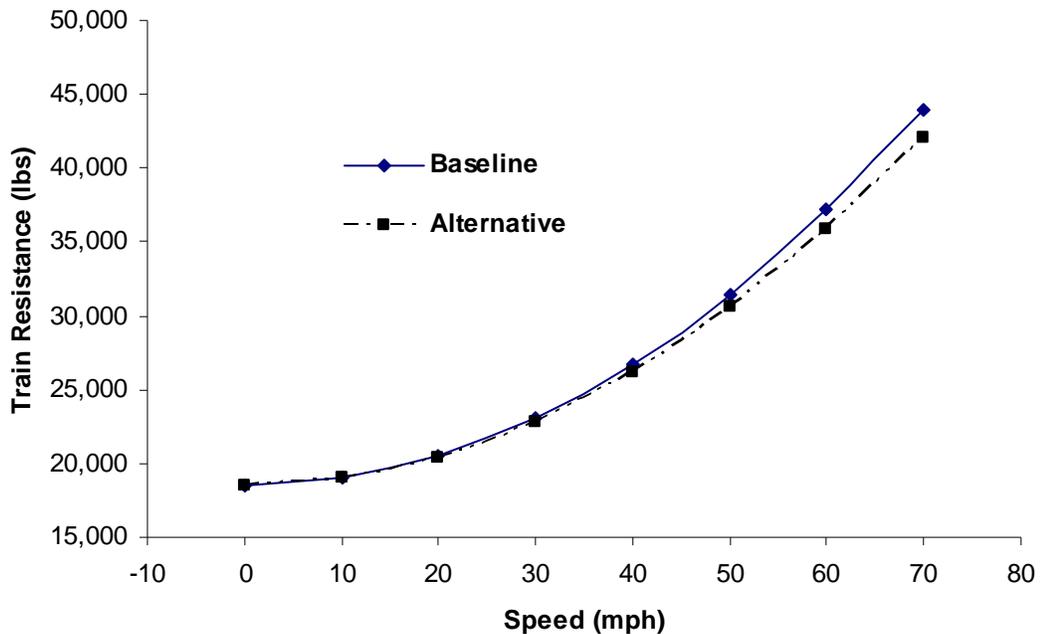
Figure 2.14:   Sensitivity analysis of speed in resistance of placing or not placing empty
containers on five empty slots in well cars

Similar analyses were conducted on trains configured as in scenarios 2 and 3, with
similar results.   Figures 2.16b and 2.16e show the effect of placing empty trailers on five
empty slots and ten empty slots respectively.   For the spine cars with trailers, the trend is
the same as for containers on the well cars but the benefit is greater.   At 70 mph, filling

five empty slots with trailers, reduced train resistance by 5%, and filling ten empty slots

reduced train resistance by 9%.   The greatest benefit comes from placing empty

containers on empty slots in spine cars, with a benefit for filling five empty slots at 70

mph of 10%, and 18% for filling ten empty slots (Figures 2.16c, 2.16f).
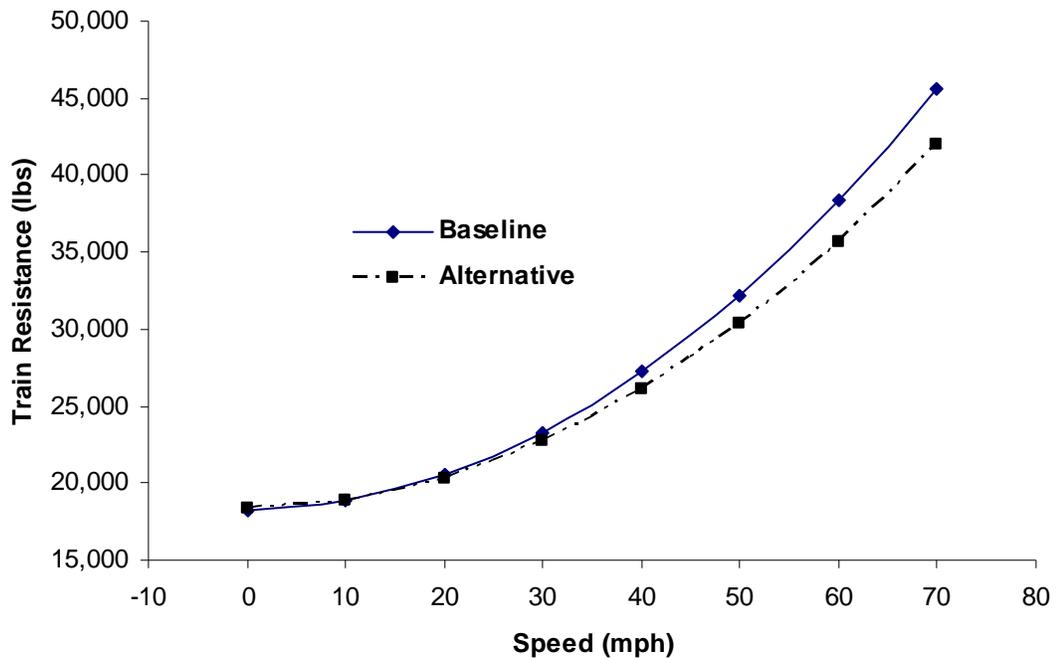
Figure 2.15:   Sensitivity analysis of speed in resistance of placing or not placing empty

containers on ten empty slots in well cars

In conclusion, the practice of loading empty intermodal equipment in empty slots

will generally have a beneficial effect on train resistance.   Use of this practice for

containers on spine cars offers the greatest benefit, followed by trailers on spine cars and

then containers on well cars.
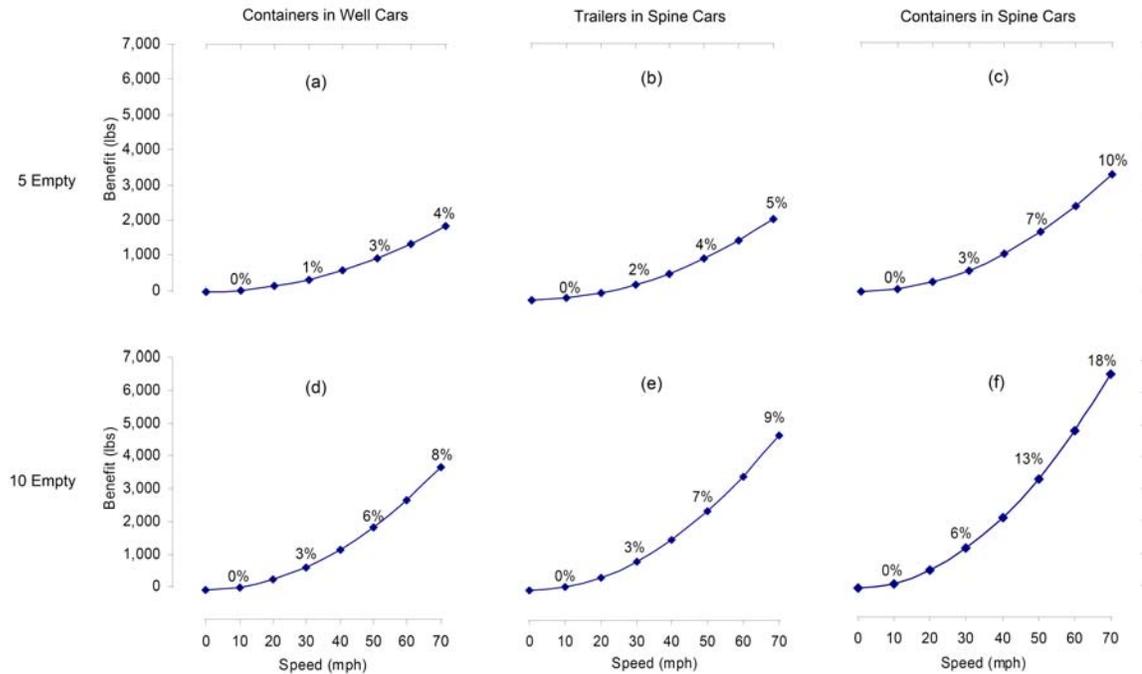
Figure 2.16:   The benefit of placing (a) empty containers on five empty slots in well cars;

(b) empty trailers on five empty slots in spine cars; (c) empty containers on five empty

slots in spine cars; (d) empty containers on ten empty slots in well cars; (e) empty trailers

on ten empty slots in spine cars; (f) empty containers on ten empty slots in spine cars

2.4.5   Fuel Consumption Computation

Four trains were analyzed for each of the three scenarios representing the baseline and alternative cases with five or ten empty slots each (Table 2.1).   Simulations using TEM were conducted for each train configuration over the same 103-mile route described above.   Filling five empty slots with loads resulted in a savings of about 22 gallons of fuel per train in scenario 1 (double stack containers on well cars), 24 gallons of fuel in scenario 2 (trailers on spine cars) and 68 gallons of fuel in scenario 3 (containers on spine cars) (Table 2.1).   Filling ten empty slots with loads would save 47 gallons of fuel in

scenario 1, 53 gallons of fuel in scenario 2 and 104 gallons of fuel in scenario 3. The fuel savings ranged from 0.21 gal/mile to 1.01 gal/mile.

Table 2.1:   Fuel consumption of baseline and alternative cases for double stack
containers on well cars (scenario 1), trailers on spine cars (scenario 2), and
containers on spine cars (scenario 3)

	5 Empty Slots				10 Empty Slots			
	Baseline	Alternative	Fuel Savings		Baseline	Alternative	Fuel Savings	
	(gallons)	(gallons)	(gallons)	(gal/mile)	(gallons)	(gallons)	(gallons)	(gal/mile)
Scenario 1	765	743	22	0.21	787	740	47	0.46
Scenario 2	786	762	24	0.23	811	758	53	0.51
Scenario 3	635	567	68	0.66	669	565	104	1.01

2.4.6   Extra Costs of Filling Empty Slots

Maximizing slot efficiency involves better matching of equipment and loads, but does not require transportation of extra equipment; however filling empty slots does. Consequently, the extra costs associated with this should be accounted for. These include the extra grade resistance and the opportunity cost of the equipment.

Over the 103-mile route analyzed, grade resistance was not a significant factor. However, there are substantial elevation changes over many intermodal routes. Empty 40' containers weigh approximately 8,500 lbs and the added grade resistance should be accounted for in calculating the savings due to the improved aerodynamics. We used TEM and Poole's fuel consumption formula (AREMA, 2001; Poole, 1962) to evaluate the importance of this effect over a typical western transcontinental route with

approximately 15,000' of total elevation rise.   We found that lifting the weight of 10

empty containers would require approximately 38.5 additional gallons of fuel.   Using

the figures in Table 2.1, the estimated fuel savings due to the improved aerodynamics

over a 2,000-mile route range from about 400 to 2,000 gallons per train, which is

considerably more than the fuel penalty due to the extra weight.

Regarding the opportunity cost of empty intermodal loads we assumed that an empty

container is worth about $2,000 each and that the time value of 10 empty containers is

about $5.50 per day.   Thus, for a three-day trip, the total would be about $16.50 which is

less than 1% of the value of the fuel that would be saved.

2.5 Discussion

The current practice of measuring intermodal loading efficiency uses the metric slot

utilization.   For example, improving slot utilization from 90% to 100% on some typical

intermodal trains reduced the aerodynamic coefficient by 3% to 23% depending on train

type.

Beyond this, matching intermodal loads with cars of an appropriate length to

maximize slot efficiency results in further improvement in train aerodynamics and can

provide greater energy efficiency than slot utilization alone.   If the loads and cars are

matched, the additional aerodynamic benefit ranged from 5% to 36%.   Over the

103-mile long route considered, this benefit was estimated to reduce fuel consumption by

0.13 to 1.0 gal/mile, depending on the load-and-car combinations analyzed.   When these

amounts are extrapolated to the 800 to 2,000 mile distances typical of many intermodal routes, the potential for fuel savings are substantial.   Intermodal trains can be more efficiently operated if loads are assigned not only based on slot utilization, but also better matching of intermodal loads with cars.   Although not considered in this chapter, the effect will be even greater for the units in the front of the train where the aerodynamic effect is greater.

Filling empty slots with empty loads also reduces aerodynamic resistance and improves energy efficiency, despite the additional weight penalty and consequent increase in bearing, rolling and grade resistance.   A series of analyses for double stack containers on well cars (scenario 1), trailers on spine cars (scenario 2) and containers on spine cars (scenario 3) were conducted.   These scenario analyses show that filling empty slots with empty loads is beneficial, and the magnitude of this benefit increases with the number of empty slots to be filled.

Based on sensitivity analyses of speed, filling empty slots with empty intermodal loads will generally reduce train resistance at the speeds typical of intermodal trains. The container-on-spine car scenario is the most beneficial followed by trailer on spine car and then container on well car.   The fuel savings generally ranged from 0.21 to 1.01 gal/mile over the route considered.

Although these options appear to offer potential benefit in terms of energy efficiency, they also introduce logistical challenges regarding rail car use, positioning and

availability, terminal operations and design, and placement of empty containers or trailers. The cost-effectiveness of implementing new practices based on the results presented here would have to consider all of these factors.

2.6 Conclusion

Three approaches for improving intermodal train energy efficiency were evaluated; slot utilization, slot efficiency, and filling empty slots.   All have a beneficial effect. Slot efficiency in which intermodal loads are matched with cars of an appropriate length reduces the gap length between loads, thereby improving airflow.   Compared to slot utilization, maximizing slot efficiency offers additional potential to reduce fuel consumption and intermodal train operating costs.   Filling empty slots with empty containers or trailers also reduces aerodynamic resistance thereby improving energy efficiency.   Despite the additional weight penalty and consequent increase in bearing and rolling resistance that would accrue, the reduction in aerodynamic resistance more than offsets this at speeds typical of intermodal trains.

CHAPTER 3

OPTIMIZING THE AERODYNAMIC EFFICIENCY OF INTERMODAL FREIGHT TRAINS

In chapter 2, I conducted a series of analyses to compare both the relative and absolute effects of different loading patterns and operating practices on train make-up and energy efficiency. It was found that aerodynamic characteristics significantly affect IM train fuel efficiency; and, a train can be more efficiently operated if loads are assigned not only based on slot utilization but also by properly assigning loads to cars, which we referred to as "slot efficiency". Depending on the particular train configuration, train resistance for a fully loaded train can be reduced by as much as 27%, and fuel savings by 1 gallon per mile per train, simply by better matching loads and railcars.

Our previous work (Lai and Barkan, 2005) quantified the aerodynamic and energy penalties of specific load and car combinations under idealized conditions. We did not consider the actual make-up of train consists or the wide variety in available loads and car types that a terminal manager must contend with in trying to implement more energy efficient loading practices. In Lai et al. (2007), we describe a wayside machine vision system that automatically monitors the gap lengths between IM loads on passing trains so the railroad can evaluate how aerodynamically that trains were loaded. However, no previous work has addressed the question of how to select among the wide variety of

loads and railcars actually available to load aerodynamically efficient trains. This is an essential element of achieving the potential fuel and costs savings. In this chapter, I develop an aerodynamic loading assignment model (ALAM) using an integer programming (IP) framework to optimize aerodynamic efficiency under various constraints regarding loading assignments. The model can help terminal managers load trains more efficiently and can be incorporated into software used to automate or expedite the loading process inside IM terminals.

Previous researchers have considered various other aspects regarding optimization of the loading process and equipment utilization. Feo and Gonzales-Velrade (1995) proposed an integer-linear programming model to maximize the utilization of trailers to railcar hitches. Powell and Carvalho (1998) developed a dynamic model to optimize the flow of flat cars over a network. Corry and Kozan (2006) presented an assignment model to dynamically assign containers to IM trains so as to minimize excess handling time and optimize the weight distribution of the train. Each of the above studies focused on certain types of IM loads or railcars. However, none of them considered the energy efficiency of IM train loading. In this chapter, we present the first application of optimization techniques to improve the energy efficiency of IM trains. The proposed model can deal with all types of IM loads (11 different types of trailers and containers), and railcars (hundreds of different types of well, spine, flat cars) operated in North America (TTX, 1999).

This study is particularly timely in light of increasing fuel prices and their impact on industry operating costs, as well as the need to conserve energy and reduce greenhouse gas emissions.   In 2006, the major North American railroads spent over $8 billion on fuel in the United States making it their largest operating expense.   From 2002 to 2005, North American railroad fuel cost doubled, and since 1999 it is up by nearly a factor of three (AAR, 2006).   This trend is impacting railroads all over the world, consequently fuel efficiency is more important than ever (Stodolski, 2002; BNSF, 2004; UIC, 2004). Investigation of options to improve aerodynamic efficiency is a promising avenue with widespread potential economic and environmental benefits for rail freight transportation efficiency (AAR, 1981 and 1987; Smith, 1987).

3.1   Intermodal Rail Terminal Operations

At IM terminals, managers assign containers and trailers of a variety of lengths to available well, spine or flat cars (BNSF, 2004; UP, 2004).   Railroad IM business in North America is different from the general carload freight business; IM business often competes directly with trucks and as a consequence is very time sensitive.   Because of this, railroads try to avoid intermediate switches and stops on most IM trains.   In this study, we used the IM operation of the BNSF Railway's route between Chicago and Los Angeles (LA) (aka "the Transcon") as the basis for our analyses.   This is one of the busiest IM corridors in North America and approximately 80% of the IM trains on this route have no intermediate operations.   Of the remaining 20%, most have no more than two intermediate stops and these are generally close to the final destination (Utterback, 2006).   Given that for most trains there is little or no container shifting occurring en

route, the initial loading pattern will be the principal factor affecting their aerodynamic performance.

IM loads, i.e. trailers or containers, range in length from 20 to 57 ft (Muller, 1999; TTX, 1999).   There is considerable variety in the design and capacity of IM railcars with different numbers of units and slots, and thus loading capabilities.   An IM railcar may 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 3.1a shows an articulated 3-unit well car, and Figure 3.2b 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 ability to accommodate two containers, one stacked on the other (a.k.a. "double stack"), and each spine-car unit has one slot (Figure 3.1).

(a)

(b)
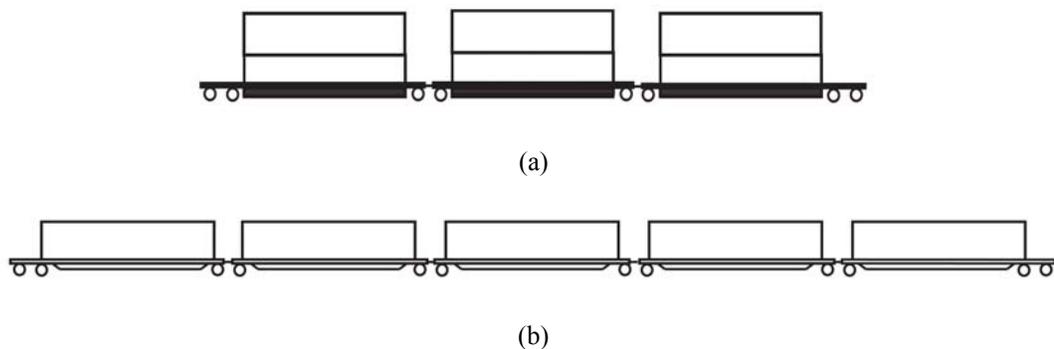
Figure 3.1:   (a) A 3-Unit Well Car with 6 Slots (b) a 5-Unit Spine Car With 5 Slots

There are also a number of safety-related loading rules and various feasible and infeasible combinations of IM load and car configurations.   Because IM 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, it is reasonable to treat train make-up as given.   Terminal managers often use computer software (Optimization Alternatives Ltd Inc, 2006) as a decision-making tool to assist them in complying with loading rules; nevertheless load assignment is still a largely manual process in which aerodynamic efficiency is not considered.

3.2   Methodology

To develop ALAM we need to quantify the IM train aerodynamic characteristics in order to incorporate them into the integer programming model for optimal loading.

3.2.1   Evaluation of Intermodal Aerodynamic Efficiency

The principal metric currently used to measure the efficiency of IM train loading is "slot utilization" (Burriss, 2003).   It measures the percentage of available slots on IM cars that are used for loads.   Slot utilization does not take into account the size of the space compared to the size of the load.   Although perfect slot utilization indicates maximal use of available spaces, it is not intended to, nor does it ensure, that IM cars are loaded to optimize their aerodynamic characteristics and hence maximize energy efficiency.   Two trains may have identical slot utilization, but significantly different energy efficiency due to different loading patterns and consequent aerodynamic resistance (Lai and Barkan, 2005).

Aerodynamic drag is known to be a major component of train resistance, particularly at high speeds (Hay, 1982; AREMA, 2001).   The Association of American Railroads

(AAR) supported research on wind tunnel testing of rail equipment, including large-scale IM car models (Gielow and Furlong, 1988). The test results were used to develop the Aerodynamic Subroutine of the AAR's Train Energy Model (TEM) (Drish, 1992). These experiments showed that the gap length between IM loads, position-in-train, and yaw angle of wind are three important factors affecting train aerodynamics (Engdahl, 1987). Although yaw angle of wind is important at specific locations, over a long route this effect will tend to be canceled out by winds in all directions. Consequently, this study focuses on the first two factors, namely gap length and position-in-train effects.

The greater the gaps between loads, the larger the aerodynamic penalty because closely-spaced containers or trailers behave as one long load. In contrast, loads spaced 72″ or more apart, behave as distinct objects as boundary layers on their surfaces are reinitialized (Engdahl, 1987). The wind tunnel tests showed that the lead locomotive experiences the highest drag due to headwind impact. After the head end, resistance declines until about the 10th unit in the train, after which drag remains nearly constant for the remaining units. Therefore, minimizing the total gap length and placing loads with shorter gaps near the front of the train will result in lower aerodynamic resistance. Consequently, the objective of optimal loading can be stated as minimization of the "total adjusted gap length (TAGL)" within the train. The adjusted gap length is defined as the gap length weighted by the position-in-train effect, where the weight associated with each unit gets smaller as the unit gets farther from the head end. The relationship between aerodynamic resistance and position-in-train effect is represented in the following equation derived from wind tunnel testing (Engdahl, 1987):

$$C_D A = 14.85824e^{-0.29308k} + 9.86549e^{-0.00007k} + 10.66914 \qquad (3.1)$$

where k is the unit position in the train and $C_D A$ is the drag area which represents the

aerodynamic resistance in ft^2.   The adjusted factor associated with each gap can be

computed by dividing the drag area of a given unit by the drag area of the 100th unit

(Table 3.1).

Table 3.1:   Adjusted factor of each gap in the train

k	Drag area (ft^2)	Adjusted 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

3.2.2   Aerodynamic Loading Assignment Model (ALAM)

The following notation is used in the algebraic model: $I=\{i\}$ is an index referring to

the type or size of the load (for instance, i=C40 represents a 40' container, i=T48

represents a 48' trailer, etc.). C_L and T_L are a subset of I representing containers and

trailers, respectively.   We group loads of the same type together and denote each load of

type i with j = 1, 2, 3…J_i, where J_i is the total number of loads of type i (for instance, J_{T48}

=10 means that there are ten 48'-trailers in the storage area).   The symbol k (k = 1, 2,

3…N) defines the position of each unit in the train.   For instance, k = 1 corresponds to

the first IM unit of the train, k = 2 corresponds to the second unit, etc.   The slot position

in every unit is denoted by p, where $p = 1$ represents the upper (top) slot (platform) in a well-car unit or the (only) slot in a spine-car or flat-car unit, and $p = 2$ represents the lower (bottom) slot in a well-car unit (see Figure 3.2).

The following symbols represent the parameters used in the model: A_k is the adjusted factor of the kth gap (see Table 3.1), where $A_1 > A_2 > ...>A_N$; U_k is the length of the kth unit; δ_k indicates the type of the kth unit, where $\delta_k = 1$ when the unit is a well-car unit, and $\delta_k = 0$ otherwise; L_i is the length of an ith type load; Q_{kp} is the length limit of position p in the kth unit; w_{ij} is the weight of the jth load of type i; W_k is the weight limit for the kth unit; and R_{ipk} is a three dimensional matrix for loading capabilities of each slot, where $R_{ipk} = 1$ if the ith type of load can be assigned to position p in unit k, otherwise it is 0.   Finally, Φ represents an arbitrarily specified large number introduced for modeling purposes as will be explained in the model description below.

(a)

(b)
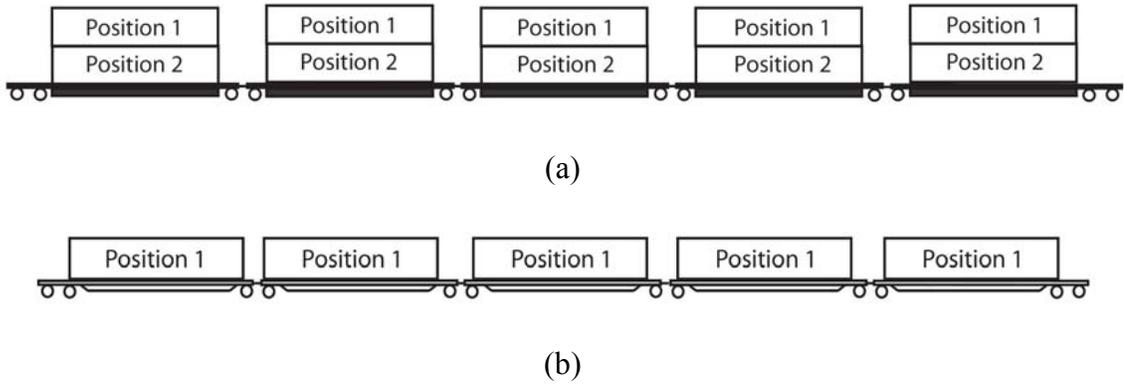
Figure 3.2:   The available slots in (a) 5-unit well car (b) 5-unit spine car

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

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

The second binary variable, denoted by x_k, determines whether the position 1 (top slot) in a well unit can be used, namely:

$$x_k = \begin{cases} 1, & \text{if top platform of the } k^{th} \text{ unit can be used} \\ 0, & \text{otherwise} \end{cases}.$$

According to the loading rules, position 1 of a well-car unit (top slot) can be used when the bottom slot is filled by containers whose total length is at least 40' (AAR, 2004).

The loading problem is formulated here as a linear integer programming model. This model minimizes aerodynamic resistance of an IM train (thus the fuel consumption), which is assumed to be a function of the train's total adjusted gap length, subject to the train characteristics and loading possibilities for a given set of loads. The algebraic expression is given below:

Min

$$0.5 \times \left\{ A_1 \left(U_1 - \sum_i \sum_j y_{ij11} L_i \right) + \sum_{k=1}^N A_{k+1} \left[\left(U_k - \sum_i \sum_j y_{ij1k} L_i \right) + \left(U_{k+1} - \sum_i \sum_j y_{ij1k+1} L_i \right) \right] \right\} \quad (3.2)$$

The constraints of the model are as follows:

$$\sum_p \sum_k y_{ijpk} R_{ipk} \leq 1 \quad \forall i, j \tag{3.3}$$

$$y_{ijpk} \leq R_{ipk} \quad \forall i, j, p, k \tag{3.4}$$

$$40 - \sum_{i \in C_L} \sum_j y_{ij2k} L_i \leq \Phi(1 - x_k) \quad \forall k \text{ (such that } \delta_k = 1) \tag{3.5}$$

$$\sum_{i \in C_L} \sum_j y_{ij1k} \leq x_k \quad \forall k \text{ (such that } \delta_k = 1) \tag{3.6}$$

$$\sum_{i \in T_L} \sum_j y_{ij2k} \leq 2 \times (1 - \sum_{i \in C_L} \sum_j y_{ij1k}) \quad \forall k \text{ (such that } \delta_k = 1) \tag{3.7}$$

$$\sum_i \sum_j \sum_p y_{ijpk} w_{ij} \leq W_k \quad \forall k \tag{3.8}$$

$$\sum_i \sum_j y_{ijpk} L_i \leq Q_{kp} \quad \forall k, p \tag{3.9}$$

$$y_{ijpk}, x_k = 0, 1 \tag{3.10}$$

The objective function (TAGL) represents the total adjusted gap length, which is comprised of two parts.   The first part involves the gap length between the locomotive and the first load (Figure 3.3), which is given by the difference between the length of the first unit (U_1) and the length of the load in position 1 of the 1st unit $(\sum\sum y_{ij11} L_i)$ divided by 2.   Multiplying the gap length by the adjusted factor A_1 results in the first adjusted gap length.   The second part of the objective function computes the sum of the subsequent adjusted gap lengths.   Each of the subsequent gaps is half of the difference in length between the current unit and the load $(U_k - \sum\sum y_{ij1k} L_i)/2$ plus half of the length difference between the next unit and the load $(U_{k+1} - \sum\sum y_{ij1k+1} L_i)/2$ multiplied by the appropriate adjusted factor, A_k.   Note that we only take into account the loads in position 1 of all units in the train.   This is reasonable since these are the only loads on spine or flat cars; and for well cars, the upper level gaps have a more significant aerodynamic effect than the lower level gaps (Furlong, 1988; Storms, 2005; Airflow Science Corporation, 2006).
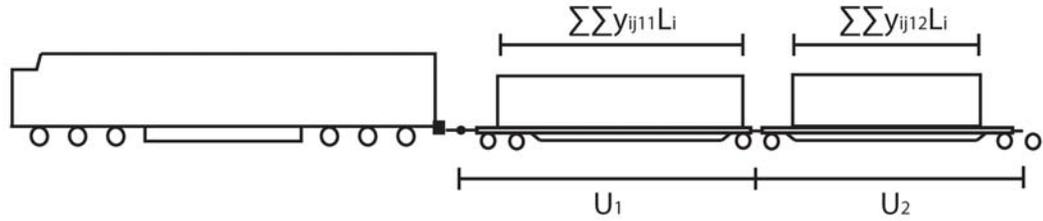
Figure 3.3:   Locomotive and first two intermodal units in a train

Minimizing total adjusted gap length creates the most efficient train configuration. However, not all loads can be assigned to all slots.   Loading assignment must conform to the loading capability of each unit as well as length and weight constraints.   These are expressed in the model constraints above.   The first two constraints, (3.3) and (3.4), ensure that each load is assigned to no more than one slot following the loading rules. Constraints (3.5) and (3.6) together state that if the bottom slot (position 2) in a well-car unit ($\delta_k = 1$) is not filled with loads greater than 40 ft, in which case equation (3.5) requires that $x_k = 0$, then no load can be assigned to the top slot (position 1) for the same unit, i.e., $\sum\sum y_{ij1k} = 0$ and therefore $y_{ij1k} = 0$ for all i, j.   Constraint (3.7) ensures that containers cannot stack on top of trailers in well-car units.   Constraints (3.8) and (3.9) are for weight and length limits, respectively.   The weight constraint (3.8) is imposed for each car unit in order to reflect the total carrying capacity of that unit.   The length constraint (3.9) is 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.

3.2.3   Solution Algorithm

When assigning loads to IM trains, there are 3 possible scenarios that terminal managers may encounter: (1) number of loads = number of slots, (2) number of loads >

number of slots, and (3) number of loads < number of slots.   Scenario 1 and 2 are more common than scenario 3, and they can be solved directly by using the IP model to select the best loads for the available slots.   When there are fewer loads than available slots (scenario 3), in some cases ALAM assigned two loads to two spine-car units (one load per unit) instead of double stacking them in a frontal well-car unit although a well-car unit was available.   This is because the model only accounts for loads in position 1 (top) of all units for determining the adjusted gap length calculation.   Clearly, such a loading pattern is not the most efficient alternative.   To solve with this we developed an algorithm that can deal with all possible loading scenarios and successfully implement ALAM to determine the most energy efficient loading pattern.

A stepwise indirect approach is presented below based on the idea that placing loads towards the front of the train and leaving the rear cars empty is generally more aerodynamic and thus preferable.   This approach ensures that only the front part of the train is first made available for loading; thereby leaving the rear of the train empty.   The algorithmic details are provided below:

Step 1: $k = 1$, *total number of loads* $= N_L$, *total number of units in the train* $= N_U$

Step 2:   *Count number of slots from 1st unit to k^{th} unit* $\rightarrow$ *if this value is less than N_L and $k < N_U$ then go to Step 3; otherwise, go to Step 4*

Step 3: $k = k + 1$ *and go to Step 2*

Step 4:   Solve by ALAM → if $k = N_U$ or there is no unassigned loads then go to

Step 5; otherwise, go to Step 3

Step 5:   Stop and output the optimal loading pattern

The algorithm starts with identifying the "loaded section" by increasing the number

of available slots (in "loaded section") until it is equal to the number of total loads.

With the "loaded section" determined, we can then use ALAM to solve the IP problem.

After implementing ALAM, if there are no unassigned loads or if k is equal to the total

number of units in the train, we can output the optimal loading pattern; otherwise, we

have to increase the number of available slots to accommodate the unassigned loads.

Although the number of slots is equal to the number of loads, some loads might still be

unassigned because of possible weight or length restrictions.   If there are any unassigned

loads and k is not the last unit in the train, the $(k+1)^{\text{st}}$ unit is made available.

With this solution procedure, ALAM is able to deal with all kinds of scenarios of IM

loading which means the loading assignment model is now complete.   This stepwise

indirect algorithm may require several iterations before reaching the optimum; however,

our computational experience shows that the solution time of ALAM is efficient enough,

and therefore is a practical decision tool for real-time terminal operations.

3.3   Empirical Application

Most IM trains operating in North America can be categorized into four general

types: (I) International Stack Trains; (II) Domestic Stack Trains; (III)

Trailer-on-Flat-Car (TOFC)/Container-on-Flat-Car (COFC) Trains; and, (IV) Mixed IM

Trains consisting of both TOFC/COFC and Double Stack equipment (Armstrong, 1998). International and Domestic Stack Trains (Type I & II) have only well cars, TOFC/COFC Trains (Type III) have spine and flat cars, and mixed IM Trains (Type IV) are comprised of all types of IM railcars (well, spine, and flat cars).

We have developed a database of approximately 250 IM trains and loads that operated on the BNSF Chicago – Los Angeles route (BNSF, 2005). Based on our assessment of their typical makeup, four representative trains were selected for detailed comparison of current terminal operations and the optimal loading pattern, given the loads and cars available (Table 3.2). Train 1 represents a double stack train transporting mostly international containers. Train 2 is also a double stack train but it is used for domestic containers. Train 3 is a TOFC/COFC train with a variety of trailers, and train 4 was comprised of well, spine and flat cars, and it also has a variety of differently-sized containers and trailers (Table 3.3).

As a result of the different characteristics of equipment and loads of these types of trains, there are varying degrees of flexibility in loading options. For example, train 1 has primarily 20' and 40' containers with little flexibility in alternative loading assignments compared to train 2, which had a much greater diversity of load types due to the variation in domestic container size (Table 3.3). Consequently, the potential improvement in the aerodynamics of train 2 is greater than for train 1.

Table 3.2:   Four representative IM trains

Train	Type	Number of Cars	Number of Units	Total Loads
1	International Stack Train	30	104	224
2	Domestic Stack Train	37	115	244
3	TOFC/COFC Train	37	110	131
4	Mixed Train	32	104	173

Table 3.3:   Number of loads in the example trains

Train	C20	C40	C45	C48	C53	T20	T28	T40	T45	T48	T53	Total Loads
1	32	184	8	0	0	0	0	0	0	0	0	224
2	28	88	9	17	102	0	0	0	0	0	0	244
3	0	0	0	0	0	11	31	0	30	35	24	131
4	32	22	0	6	59	0	2	2	2	15	33	173

In the following sections, we present a series of scenario analyses for the four types of IM trains described above.   In section 3.3.1, we consider the scenario in which the number of loads equals the number of available slots in the train, and in sections 3.3.2 and 3.3.3, we consider scenarios in which the number of loads is more, or less, than the number of slots, respectively.   In section 3.3.4, we compute fuel consumption for each of the scenario analyses, and the results and policy suggestions based on our analyses are presented in 3.3.5.

3.3.1   Scenario 1: Number of Loads Equals Number of Slots

ALAM was used to analyze the optimal loading pattern for each of the four representative types of IM train and loads shown in Tables 3.2 and 3.3.   Certain restrictions were applied to the loading pattern when assigning loads to slots (AAR, 2004; Armstrong, 1998; TTX, 1999).   For example, a 48'-well car cannot handle containers or trailers greater than 48' in the bottom position (position 2), although it can accommodate

containers up to 53' in the top position (position 1).   53'-slot spine cars have only one platform (position 1) per unit and can handle either containers or trailers up to 53' long. Besides length and weight constraints, some of the IM units can accommodate only containers, some can accommodate only trailers, and the others can handle both.   To ensure that load assignment follows the loading rules, possible loading combinations are specified for each IM unit.   In order to clearly illustrate the aerodynamic effects, we assumed that none of the units were constrained by weight limits and the optimization process was based solely on minimization of the total adjusted gap length (TGAL).

For any given train and pool of loads, there is at least one loading pattern in which TAGL is minimized (the most aerodynamic pattern), and conversely, another loading pattern in which it is maximized (worst case).   At present, terminal managers' goal in load assignment is to maximize slot utilization; therefore, they are largely indifferent to alternative loading patterns as long as they achieve 100% slot utilization (and comply with applicable loading rules).   Consequently, with 100% slot utilization, average terminal operating practice will be some intermediate value between these two extremes. Therefore, I assumed that the mean of the TAGL for the minimum and maximum cases represents average terminal performance in the scenarios with 100% slot utilization (scenarios 1 and 2).

CPLEX 9.0 incorporated with GAMS (Brooke et al., 1998) was used to solve the model.   Our computational experience showed that the optimal solution for each case could be obtained in less than five seconds.   Thus, computational complexity is not a

problem for real-time terminal operations. We then compared the results of ALAM with our estimate of average terminal loading (Figure 3.4). In this scenario, the optimal results were generally close to the mean due to the inflexibility in possible loading combinations. This was especially evident for train 1 and train 3. Since train 1 has mostly international IM loads (i.e. 20' and 40' containers), the difference between the most and the least aerodynamic loading patterns is small. For train 3, the small range in TAGL is due to the characteristics of trailers and equipment in TOFC/COFC (Type III) trains. Train 4 is really a combination of trains with characteristics like trains 1 and 3 so it is not surprising that it also demonstrates a relatively small possible range of TAGL values for this scenario. By contrast, train 2 has a considerably greater TAGL range because of the wider variation in loads that is typical of this type of train.
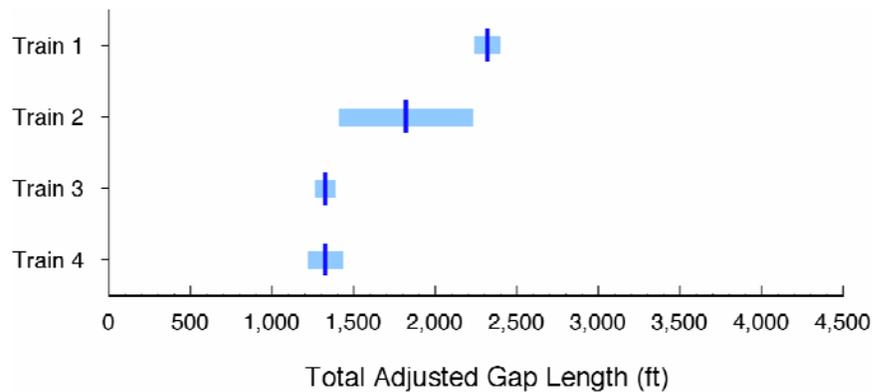
Figure 3.4: Minimum, maximum and estimated average TAGL for the four example trains when loads = slots (scenario 1)

3.3.2 Scenario 2: Number of Loads Exceeds Number of Slots

At IM terminals, managers often have more loads than available slots in the next outgoing train (scenario 2) and this offers greater flexibility in load assignment. Scenario 2 is more common at IM terminals than scenarios 1 and 3 so these results are particularly important in assessing the potential benefits of using ALAM. The task in scenario 2 is to select the best set of loads from the load-pool to match the current outgoing train. Unfortunately, specific data on the composition of the pool of loads are not available because IM load assignments are only recorded when an outgoing train is loaded and ready to depart. However, it is reasonable to assume that the load pool is proportional to the distribution actually loaded onto a train. For the purpose of comparison among the four train types in this scenario analysis, I assumed that the number of each type of load was increased by approximately 50% (Table 3.4). Train consists and configurations were assumed to be unchanged.

Table 3.4: Number of available loads for example trains in scenario 2

Train	C20	C40	C45	C48	C53	T20	T28	T40	T45	T48	T53	Total Loads
1	48	276	12	0	0	0	0	0	0	0	0	336
2	42	132	14	26	153	0	0	0	0	0	0	367
3	0	0	0	0	0	17	47	0	45	53	36	198
4	48	33	0	9	89	0	3	3	3	23	50	261

Due to the greater number of potential loads to choose from compared to scenario 1, there is some reduction in TGAL for all four train types; however, the magnitude of potential improvement varies widely (Figure 3.5). Only train 1 (international double-stack container trains) experiences little potential improvement, again due to the inflexibility in types of loads. By contrast, all three of the other train types show

57

substantial potential to reduce TAGL, with train 2 showing the greatest range.   The wide

range in possible TAGL means that terminal managers have a higher chance of loading

aerodynamically inefficient trains and that use of ALAM has more potential to improve

the energy efficiency of these types of trains.
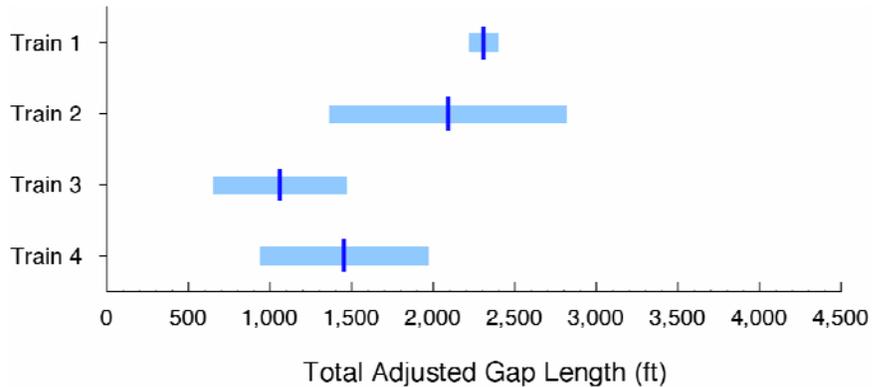
Figure 3.5:   Minimum, maximum and estimated average TAGL for the four example

trains when loads > slots (scenario 2)

3.3.3   Scenario 3: Number of Loads Less Than Number of Slots

In scenario 3, I analyzed the same four trains but the number of loads available was

reduced by approximately 50% (Table 3.5).   Since slot utilization is not 100% in this

scenario, a larger adjusted gap length does not necessarily represent the poorest

aerodynamics.   According to Engdahl et al. (1987), the worst aerodynamic pattern is to

have empty units uniformly distributed throughout the train.   If terminal managers are

loading without regard to aerodynamics, we expect loading patterns to be somewhere

between the best and the worst.   Therefore, I estimated their performance by assuming

they placed half of the empty units at the end of the train, which would not affect

aerodynamics, and the other half uniformly distributed in the "loaded section" (from the 1st unit to the last loaded unit).

Table 3.5:   Number of available loads for example trains in scenario 3

Train	C20	C40	C45	C48	C53	T20	T28	T40	T45	T48	T53	Total Loads
1	16	92	4	0	0	0	0	0	0	0	0	112
2	14	44	5	9	51	0	0	0	0	0	0	123
3	0	0	0	0	0	6	16	0	15	18	12	67
4	16	11	0	3	30	0	1	1	1	8	17	88

The number of empty units for each of the four trains can be determined based on the optimal patterns obtained from ALAM (50, 57, 52, 38 units for trains 1, 2, 3, 4, respectively).   For example, the best aerodynamic loading pattern for train 1 is to assign all available loads to the first 54 units (loaded section), and leave the last 50 units empty. The worst case for train 1 is to distribute the 50 empty units uniformly throughout the entire 104-unit train which results in 4,446-ft adjusted gap length.   Therefore, the average terminal performance was estimated by setting the last 25 units empty and distributing the other 25 empty units in the available loaded section.   The differences in TAGL among optimal loading, worst loading and estimated average terminal performance are substantially greater than those in scenarios 1 and 2 because the calculation of adjusted gap lengths ignores units not in the loaded section (Figure 3.6).
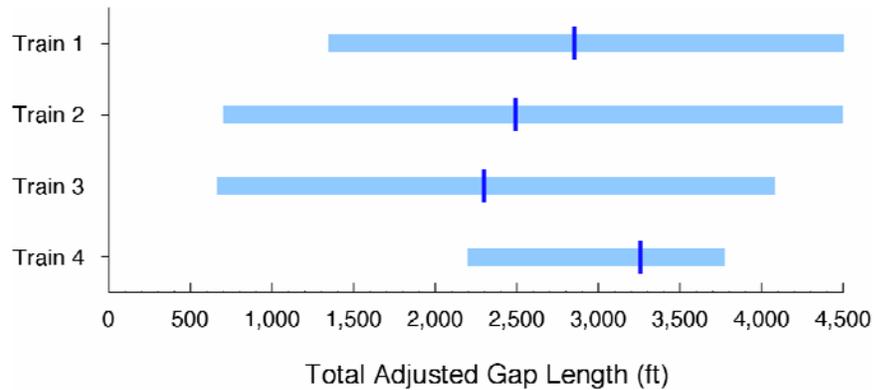
Figure 3.6: Minimum, maximum and estimated average TAGL for the four example trains when loads < slots (scenario 3)

3.3.4 Fuel Consumption Computation

To quantify the potential fuel savings resulting from the optimal loadings obtained from ALAM, I computed the aerodynamic coefficients and the corresponding fuel consumption using the Aerodynamic Subroutine and the AAR Train Energy Model (TEM) (Furlong, 1988; Drish, 1992). The BNSF Transcon is a high speed freight rail route primarily with gentle grades, curves and rolling topography, so we selected a typical 100-mile segment to estimate fuel consumption and extrapolated this to develop estimates for the entire route.

Table 3.6 summarizes the computed fuel consumption values and associated savings if trains are loaded optimally using ALAM compared to estimated-average-loaded trains. The aerodynamic benefits of trains in scenario 2 are generally higher than scenario 1 due to the increased flexibility in loading patterns. However, there is almost no added benefit from optimizing loading train 1 compared to normal terminal practices because of the inflexibility in loading patterns for this type of train. Even though the flexibility of

60

train 1 in scenario 2 is relatively higher than scenario 1, it still makes almost no difference in fuel consumption.   I will further discuss the implication of this result in section 3.3.5 below.

The aerodynamic benefit for train 3 (TOFC/COFC train) is generally higher than the others even though the range in TAGL in scenario 1 is not as large as the others (Figure 3.4).   This is because loads in spine or flat cars can be placed relatively closer to each other compared to the other types of trains comprised of well cars (Lai and Barkan, 2005), resulting in a greater difference in the aerodynamic efficiency of loading patterns.

Scenario 3 is a relatively unique case (slot utilization << 100%) compared to scenarios 1 and 2 (slot utilization = 100%).   The differences between optimal patterns and estimated average terminal practices in scenario 3 are substantial because there are quite a few long gaps in the terminal case caused by empty slots in the train (Table 3.6c).   The fuel savings for trains 1 and 2 are higher here since the aerodynamic drag caused by empty slots between two double-stack units is greater than empty slots between single-level units in trains 3 and 4.

Table 3.6: Fuel consumption and estimated savings for optimally loaded trains compared to average loaded trains (estimated terminal performance) in (a) scenario 1 (b) scenario 2 (c) scenario 3

(a)

| Train | Fuel Consumption (gallons) | | Fuel Savings | |
	Optimal	Terminal	(gallons/mile)	%
1	961	961	0.00	0.0%
2	964	973	0.09	0.9%
3	937	996	0.59	5.9%
4	944	986	0.42	4.3%

(b)

| Train | Fuel Consumption (gallons) | | Fuel Savings | |
	Optimal	Terminal	(gallons/mile)	%
1	961	961	0.00	0.0%
2	955	988	0.33	3.3%
3	937	1,033	0.96	9.3%
4	945	967	0.22	2.3%

(c)

| Train | Fuel Consumption (gallons) | | Fuel Savings | |
	Optimal	Terminal	(gallons/mile)	%
1	827	992	1.65	16.6%
2	825	1,036	2.11	20.4%
3	823	926	1.03	11.1%
4	838	928	0.90	9.7%

3.3.5 Policy Recommendations for Railway Intermodal Operations

The scenario analyses presented above demonstrate the potential benefit of implementing ALAM at IM terminals. In this section, we discuss the implications of these results and some options to improve the energy efficiency of IM train operations.

**Use of ALAM for intermodal train loading**

The most obvious recommendation stemming from the work described here is that ALAM be integrated into railroads' IM terminal loading operations. Use of ALAM to match loads to slots would have reduced fuel consumption by as much as 20%. The exact amount will vary for individual train consists and loads available, but these figures provide insight into the magnitude of potential savings available. The percentage of each of the four types of trains on the BNSF Transcon are approximately equal with a total of about 50 per day. Accounting for the different potential savings for these train types if they were loaded according to ALAM guidelines, compared to the average loading configuration, translates to a total annual savings of about 29 million dollars based on the most common scenario (scenario 2).

Major railroad IM terminals already use software to assist and expedite the loading process, so integration of ALAM objectives and methodology into this software would not require substantial institutional or process change. It also should have little if any impact on operating cost because it will not generally require more work, but rather, inform the loading process with a new quantitative parameter, minimization of TAGL. Implementation of ALAM would automate terminal managers' consideration of the large variety of loads and railcar types available thereby enabling them to load trains in a more aerodynamically efficient configuration. These benefits can be further enhanced by several additional factors as described in the following three subsections.

Better Matching of Loads with Railcars

The aerodynamic benefit of optimizing the loading pattern of train 1 in scenarios 1 and 2 is small. This is due to the inflexibility in loading patterns of train 1, and also the characteristics of the "long" well cars. Train 1 is primarily transporting 20' & 40' international containers with just a few 45' containers. Nevertheless, approximately 70% of its railcars were designed for 53' containers. Placing 45' containers in the frontal positions in the train would generally be better than 40' containers; however, for this kind of long well-car unit, it makes little aerodynamic difference. This is because using 53' cars to transport either 40' or 45' loads always results in gaps greater than the critical gap length (12 ft), necessary to gain aerodynamic benefit (Engdahl et al., 1987) so reconfiguring the shorter loads on the longer cars has little effect (Lai and Barkan, 2005).

Using 53'-well cars is convenient for managers because of these cars' flexibility; however, placing shorter international containers in 53'-well cars causes greater aerodynamic resistance than if 40'-unit well cars are used. For example, without changing the placement of loads, if train 1 used well cars designed for 40' containers (45' containers can still be placed in at the top), the aerodynamic coefficient reduces from 7.04 to 5.50 lbs/mph^2, and the weight of the train would be reduced by 18% as well. The corresponding fuel savings would be 0.88 gallons per mile, an approximately 9% reduction. Consequently, we suggest better matching IM loads with the railcars used to transport them, specifically acquisition and use of well cars with 40' slots that are designed for international loads.

Optimize Loading for More Than a Single Train Simultaneously

The analysis thus far has focused on optimizing the aerodynamic efficiency of a single outgoing train for a given set of loads. We formulated the problem this way because most loads arrive shortly before loading begins. However, if advance knowledge (either empirical or probabilistic) on the composition of outgoing trains and the load pool is available, this information can be used to optimize the loading of multiple trains simultaneously and increase the benefit of applying the model developed in this research.

ALAM can be extended to address the multiple train loading problem by modifying it to consider units of all the loads and units in each of the trains available. This is accomplished by introducing a new index, t, that refers to the set of outgoing trains. The objective function of the extended model is minimization of the total adjusted gap length of all the outgoing trains in a given time horizon:

$$\text{Min} \ \sum_{t=1}^{T} z_t \tag{3.11}$$

Where:

$$z_t = 0.5 \times \left\{ A_1 \left(U_{t1} - \sum_i \sum_j y_{ijt11} L_i \right) + \sum_{k=1}^{N} A_{k+1} \left[\left(U_{tk} - \sum_i \sum_j y_{ijt1k} L_i \right) + \left(U_{tk+1} - \sum_i \sum_j y_{ijt1k+1} L_i \right) \right] \right\}$$

The constraints are the same as ALAM except the decision variables y_{ijpk} and x_k are replaced by y_{ijtpk} and x_{tk}. Compared to the basic ALAM described earlier, this modified

version provides a global optimum solution as opposed to the local optimum based on analysis of a single train.

Uncouple Empty Railcars at the End of IM Trains

The optimal loading pattern aims to achieve the lowest aerodynamic resistance and thereby maximize fuel economy; therefore, when the number of loads is less than the number of slots (scenario 3), the loads are placed in cars at the front of the train, and empty slots in the rear.   According to IM loading rules, the speed of a train with empty cars is restricted to 55 mph due to concerns about dynamic instability of these cars at higher speeds (BNSF, 2004).   Hence, there is a tradeoff between aerodynamic loading pattern and train operating speed.   Although ALAM is intended to maximize aerodynamic efficiency, it can be modified to suit terminal operators' or dispatchers' preferences (higher train speed vs. better fuel efficiency) by adding additional constraints or pre-processes (such as each car must have at least one load so that the speed restrictions can be avoided).

We can also approach this problem by removing the empty cars from the end of the train.   Although this might incur some additional operating costs this should be compared to the energy efficiency benefits.   Uncoupling empty cars would reduce the weight of the train and also eliminate the speed restriction.   Table 3.7 shows the computed fuel consumption and respective savings for the worst case, average terminal practices, and optimal loading pattern with or without an uncoupling policy.   Fuel savings are computed by comparing terminal practices to the optimal results if the

66

uncoupling strategy is used.   The savings with implementation of the practice of uncoupling are greater than without uncoupling.   Comparison of Table 3.6c and Table 3.7 shows that the uncoupling strategy can increase fuel savings by 14 ~ 25%. Therefore, if repositioning empty IM equipment can be efficiently accomplished without transporting it in IM trains with loads, uncoupling railcars from the end of trains should be considered.

Table 3.7:   Fuel consumption and estimated savings for the worst case, average loaded trains (estimated terminal performance), and optimal results without or with uncoupling policies in scenario 3

Train	Fuel Consumption (gallons)				Fuel Savings	
	Worst Case	Terminal	Optimal without Uncoupling	Optimal with Uncoupling	(gallons/mile)	%
1	1,090	992	827	598	3.94	39.7%
2	1,135	1,036	825	562	4.74	45.8%
3	976	926	823	578	3.48	37.6%
4	976	928	838	706	2.22	23.9%

3.4 Discussion

ALAM can be extended to optimize the aerodynamic efficiency at the system level instead of optimizing the loading of a single train.   The time horizon can be extended to several days (or more if desired) depending on the availability of advance train-and-load information and how difficult it is to solve the problem using IP.   An interesting question we intend to address in the future is to consider the time horizon required to approach the global optimum, and how much improvement is possible as that horizon is lengthened.

This study focused on IM services between Chicago and LA in which no intermediate train reconfiguration occurs en route. This is the case for approximately 80% of the trains operating on this route. In the majority of cases, the initial loading pattern could be determined to optimize the train's aerodynamic efficiency without considering unloading sequences. Of the remaining 20%, half of them have no more than two additional terminals and these are generally near the final destination so the same aerodynamic benefits will apply over most of the route. If intermediate operations do occur, then the aerodynamic benefits will apply for the portion of the route prior to the terminal where the train is reconfigured. Beyond that trains may operate at lower efficiency. It would also be possible to modify the model to optimize aerodynamic efficiency for the entire route including intermediate destinations (where the train is reconfigured) in the analysis. However, there is a trade off between the costs and benefits of doing this. The increased modeling complexity would demand more computation time and the resulting problem might be too complex to solve using IP. Further research should be conducted to develop specially designed solution algorithms that can generate very good (if not optimal) loading patterns for trains whose composition changes en route.

3.5 Conclusions

We develop a mathematical programming model (ALAM) by incorporating the aerodynamic characteristics of intermodal trains to optimize load-to-unit assignments. The model can be integrated into current terminal operation software as an additional tool to help terminal managers make better loading decisions. The contributions of this work

to the literature are: (1) It is the first use of optimization modeling with the objective of improving the aerodynamics and consequent energy efficiency of intermodal trains and reveals significant possible savings are possible. (2) The model developed in this chapter can be adapted to a variety of other intermodal train loading assignment problems through modification of the objective function. This is a novel contribution to the literature and enhances its generality because the formulation can be solved efficiently and thus serve as a basis for other intermodal load assignment problems. (3) Several policy recommendations regarding railway intermodal operations are developed based on a series of scenario analyses.

There are substantial potential fuel and cost savings benefits that railroads can achieve thorough implementation of ALAM at intermodal terminals. These benefits can be further enhanced through several additional steps including: (a) better matching of railcars and loads for international intermodal trains (b) simultaneous optimization of multiple trains to take greater advantage of the potential to improve energy efficiency of intermodal trains through use of more aerodynamic loading patterns, and (c) uncoupling empty railcars from the end of loaded intermodal trains when practical. The potential annual savings in fuel consumption through use of ALAM by one large railroad on one of its major intermodal routes is estimated to be approximately 15 million gallons with a corresponding value of 29 million dollars. Correspondingly larger savings in fuel, emissions and expense are possible if the methodology described in this chapter were applied to all North American intermodal trains.

CHAPTER 4

OPTIMIZING THE AERODYNAMIC EFFICIENCY OF INTERMODAL FREIGHT TRAINS WITH ROLLING HORIZON OPERATIONS

In chapter 2, aerodynamic characteristics have been prove to have a significant impact on IM train fuel efficiency; therefore, a train can be more efficiently operated if loads and slots are carefully matched during the process of load assignments (Lai and Barkan, 2005). In order to help terminal managers assemble more fuel efficient trains, in chapter 3, I present an aerodynamic loading assignment model (ALAM) in which the objective was to maximize aerodynamic efficiency (by minimizing the adjusted gap length) of the outgoing IM freight train given any particular static combination of loads and railcar types (Lai et al., 2007). The analysis of one major railroad IM route revealed the potential to reduce fuel consumption by 15 million gallons per year with a corresponding cost saving opportunity of $29 million.

ALAM was developed based on current terminal practices and considers optimization of the loading pattern of a single train at a time. However, if advance information on outgoing trains and loads is available, a better loading plan will often be possible by simultaneously considering more than one train. The larger pool of loads and railcars will

enable better matching, but may also introduce greater uncertainty about the composition of the future load pool.

In this chapter, I extend ALAM to optimize multiple trains simultaneously. This work starts by evaluating the benefit from optimizing multiple trains and loads assuming full static information on trains and loads. I then consider the more realistic case with incomplete future information. A dynamic load assignment model with a rolling horizon scheme is developed for continuous terminal operations, which balances the advantage from optimizing multiple trains together against the risk of making suboptimal decisions due to incomplete future information. This study addresses an important economic and environmental topic in rail transport, and also makes a methodological contribution by introducing rolling horizon operations for IM loading efficiency to the literature.

4.1 Methodology

4.1.1 Loading Assignment at IM Terminals

The rail IM business in North American is quite different from the general freight business and intermediate stops are no longer the norm for most IM trains, the subject of our research. Railroads try to avoid intermediate switches and stops because the IM business is highly time sensitive. For example, approximately 80% of the IM trains on the BNSF Transcon route (Chicago – LA) 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, so there is little container shifting occurring enroute (Avriel et al., 1998;

Giemsch and Jellinghaus, 2004; Utterback, 2006).   Therefore, the initial loading pattern for most trains will be the principal factor affecting their aerodynamic performance for all or most of the trip.

At IM terminals, containers and trailers of a variety of lengths are assigned to available well, spine or flat cars by terminal managers (BNSF, 2004; UP, 2004).   IM loads, i.e. trailers or containers, vary in length from 20 to 57 ft.   There is considerable variety in the design and capacity of IM railcars with different numbers of units and slots, and thus loading capabilities.   An IM 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 4.1a shows an articulated 3-unit well car, and Figure 4.1b 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 stacking"), and each spine-car unit has one slot (Figure 4.1).

(a)

(b)
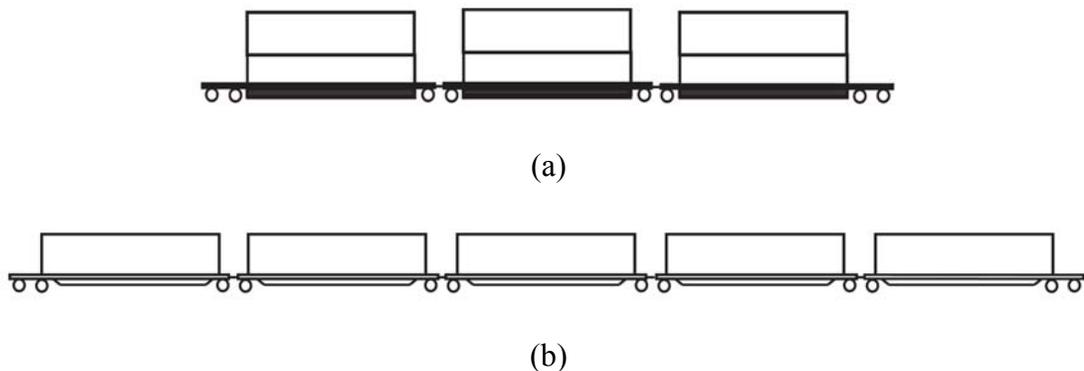
Figure 4.1:   (a) a 3-unit well car with 6 slots (b) a 5-unit spine car with 5 slots

There are also a number of loading rules developed for safety purposes and various feasible and infeasible combinations of IM load and car configurations. Because IM cars in a train are not generally switched in and out at terminals (i.e., cars will not be uncoupled from one train and coupled to another), managers primarily control the assignment of loads but not the configuration of the equipment (i.e., railcars) in a train. Consequently, we treat the train configuration as given.

Aerodynamic drag is a major component of train resistance, particularly at high speeds (Hay, 1982; AREMA, 2001; Lai and Barkan, 2005). In the 1980s, the Association of American Railroads (AAR) sponsored research on wind tunnel testing of rail equipment, including large-scale IM car models (Gielow and Furlong, 1988). The results were used to develop the Aerodynamic Subroutine of the Train Energy Model (TEM) (Drish, 1992). These experiments showed that gap length between IM loads and position-in-train were the two important factors affecting train aerodynamics (Engdahl, 1987). Larger gaps result in a higher aerodynamic coefficient and greater resistance; and, the front of the train experiences the greatest aerodynamic resistance due to headwind impact. Therefore, to incorporate both important aerodynamic factors, the model chooses to minimize the summation of total adjusted gap lengths (i.e., gap lengths multiplied by adjusted factors). The adjusted factors (accounting for the position-in-train effect) are computed by dividing the drag area (representing the aerodynamic resistance in ft^2) of a given unit by the drag area of the 100th unit; the result is listed in Table 4.1.

Table 4.1:   Adjusted factor for each gap in the train (Lai et al., 2007).

k	Drag area (ft^2)	Adjusted 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

4.1.2   Static Aerodynamic Efficiency Model

Placing loads with shorter gaps in the frontal position generates less aerodynamic resistance; therefore, the objective function of the aerodynamic model is to minimize the total adjusted gap length of all trains considered in the decision horizon.   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_L is the subset of i for container loads, and T_L is the subset of i for trailer loads.   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 = C40, T48, T53, etc.), for instance, J_{T48} = 10 means that there are ten 48' trailers in the storage area.   Let t denote 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 IM 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 4.2).   The following symbols represent the parameters used in the

model: A_k is the adjusted factor of the kth gap shown in Table 4.1, where $A_1 > A_2 > ...>A_N$;

U_{tk} is the length of the kth unit of train t; δ_{tk} indicates the type of the kth unit in train t,

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 kth unit of train t; w_{ij} is the weight

of the jth load of type i; W_{tk} is the weight limit of the kth unit of train t; and R_{itpk} is a four

dimensional matrix for loading capabilities of each slot, where $R_{itpk} = 1$ if the ith type of

load can be assigned to position p in unit k of train t, or it equals 0 otherwise.   Finally, Φ

represents an arbitrarily large number introduced for modeling purposes as explained in

the model description below.

(a)

(b)
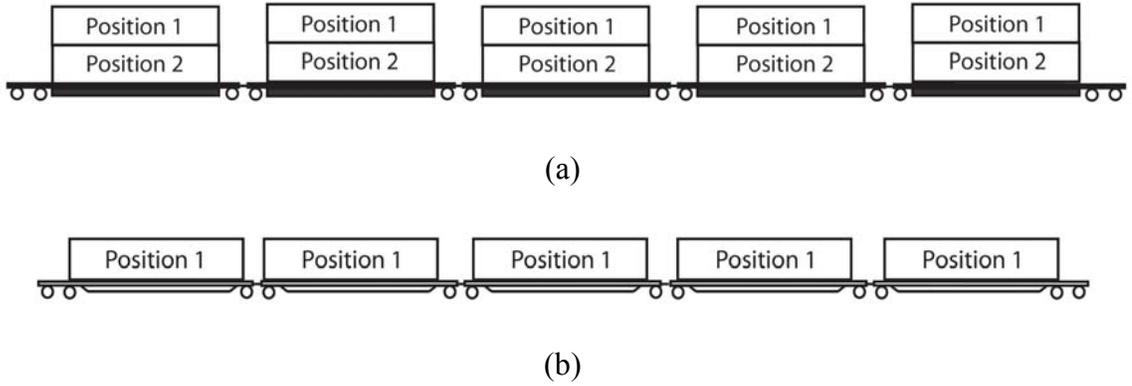
Figure 4.2:   The available slots in (a) 5-unit well car (b) 5-unit spine car

Two sets of binary decision variables are included in the model.   The first variable

is denoted by y_{ijtpk} such that:

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

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' (AAR, 2004).

The loading problem is formulated as a mixed integer program (MIP) that minimizes fuel consumption (i.e., the total adjusted gap length) of T outgoing trains. For train t, the objective function (total adjusted gap length) is

$$z_t = 0.5 \times \left\{ A_1 \left(U_{t1} - \sum_i \sum_j y_{ijt11} L_i \right) + \sum_{k=1}^{N-1} A_{k+1} \left[\left(U_{tk} - \sum_i \sum_j y_{ijt1k} L_i \right) + \left(U_{tk+1} - \sum_i \sum_j y_{ijt1k+1} L_i \right) \right] \right\} \quad (4.1)$$

This objective function is comprised of two parts. The first part, representing the gap length between the locomotive and the first load (Figure 4.3), is the difference between the length of the first unit (U_{t1}) 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 adjusted 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 adjusted factor, A_k. Thus, the second part of the objective function computes the sum of the subsequent adjusted gap lengths. Note that we only

76

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; for well cars, the upper level gaps have a more significant aerodynamic effect than the lower level gaps (Furlong, 1988; Storm, 2005; Airflow Science, 2006). A schematic representation is given in Figure 4.3.
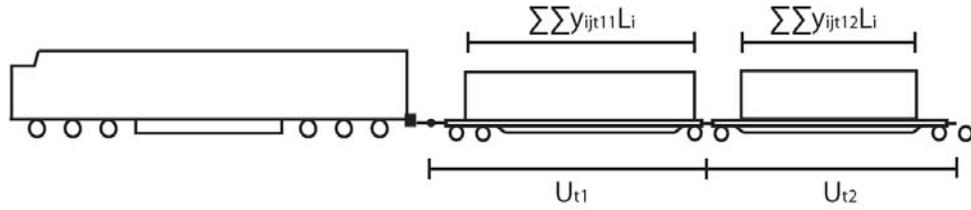
Figure 4.3:   Locomotive and first two IM units in a train.

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

$$\text{Min} \ \sum_{t=1}^{T} z_t \tag{4.2}$$

Subject to:

$$\sum_{t}\sum_{p}\sum_{k} y_{ijtpk} R_{itpk} \leq 1 \quad \forall i,j \tag{4.3}$$

$$y_{ijtpk} \leq R_{itpk} \quad \forall i,j,t,p,k \tag{4.4}$$

$$40 - \sum_{i \in C_L}\sum_{j} y_{ijt2k} L_i \leq \Phi(1-x_{tk}) \quad \forall t,k \text{ (such that } \delta_{tk}=1) \tag{4.5}$$

$$\sum_{i \in C_L}\sum_{j} y_{ijt1k} \leq x_{tk} \quad \forall t,k \text{ (such that } \delta_{tk}=1) \tag{4.6}$$

$$\sum_{i \in T_L}\sum_{j} y_{ijt2k} \leq 2 \times (1 - \sum_{i \in C_L}\sum_{j} y_{ijt1k}) \quad \forall t,k \text{ (such that } \delta_{tk}=1) \tag{4.7}$$

$$\sum_i \sum_j \sum_p y_{ijtpk} w_{ij} \leq W_{tk} \qquad \forall t, k \tag{4.8}$$

$$\sum_i \sum_j y_{ijtpk} L_i \leq Q_{tkp} \qquad \forall t, k, p \tag{4.9}$$

$$y_{ijtpk}, \; x_k \;\; = \;\; 0, 1 \tag{4.10}$$

Minimizing total adjusted gap length creates the most efficient train configuration, but the loading assignment must conform to the loading capability of each unit as well as length and weight constraints. Constraints (4.3) and (4.4) ensure that each load can be assigned to no more than one slot, and must obey the loading assignment rules (R_{itpk}). Constraints (4.5) and (4.6) 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 (4.6) allows a bottom load without a top load ($y_{ijt1k} = 0$). Constraint (4.7) ensures that containers cannot stack on top of trailers in the well car units; the parameter 2 is used for the possible scenario of two trailers in one well-car unit. Constraint (4.8) is the weight limit that is imposed for each car unit in order to reflect its total carrying capacity (W_{tk}). Constraint (4.9) is the length limit imposed for each slot to guarantee that the total length of loads in a given slot does not exceed the length of that slot (Q_{tkp}). 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 length. Thus, the model prefers not to leave a load behind if a suitable slot is available.

The above optimization formulation (4.2)–(4.10) reminds us of certain network flow problems (e.g., assignment problem) that can be solved efficiently. However, the existence of certain constraints (e.g., the weight constraints) makes the problem NP-hard. Optionally we can develop relaxation or decomposition based heuristics for our model, but earlier research (Lai et al., 2007) has shown that existing commercial MIP solver (i.e., CPLEX) can solve this problem within reasonable time (this is also found to be true in our numerical experiments). Therefore, in this study we choose to use CPLEX to solve the problem instances.

The current terminal operational practice and a previous paper by Lai et al. (2007) consider loading plans for the current outgoing train only. This scenario is a special case of the general model developed here in which $T = 1$. The model thus optimizes the aerodynamic efficiency of one outgoing train for a given set of loads. However, some degree of advance information about outgoing trains and loads is often available (Anderson, 2006). This provides an opportunity to achieve even more aerodynamically efficient loading patterns by optimizing more trains and loads together.

4.2 Dynamic Aerodynamic Efficiency Model

Obviously, optimizing multiple trains simultaneously will lead to more efficient loading plans if complete information on all trains and loads is available at the time of optimization (i.e., static current information). In practice, however, information about some loads may not be immediately available (i.e., future information). Under some

circumstances, including the loading pattern of a later train in the optimization will reduce the efficiency of the immediate outgoing train. For example, suppose the two trains compete for the same "suitable" load, and the later train gets this load in the optimization (with the objective of minimizing the total adjusted gaps in both trains). It is possible, however, that after the dispatch of the immediate train, another suitable load with the same characteristics becomes available. As a result, the earlier optimal solution (before knowing the future load information) turns out to be suboptimal (overall). Therefore, uncertainty about future loads introduces some degree of risk on optimizing multiple trains; i.e., 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 and 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 with 'smoothed' objectives 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) train; (2) how many trains to optimize each time and how to optimize them; 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 decision to the last moment possible (before loading a departing train), because it maximizes the available information, thereby reducing uncertainty. 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

only the next 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, assuming that each optimization process can be conducted efficiently to update

the optimal loading patterns in time.   The only remaining question is how many trains

should be optimized each time and how to optimize them.

We further propose an exponential smoothing approach under the rolling horizon

framework, where future trains are considered simultaneously with the current train.

Before loading the tth train, suppose we have known information on $n_k(t)$ unassigned

loads, and these loads can fill a maximum number of $T(t)+1$ sequential trains (i.e., trains t,

$t+1$, ..., $t+T(t)$), where $T(t)+1$ is the number of future trains considered in an assignment.

The time horizon is defined to be from the departure time of train t to that of train $t+T(t)$.

The loading decision for train t will be directly relevant to the trains departing in this

horizon.   Meanwhile, these trains are also directly influenced by the future loads

incoming within this horizon; assume there are $n_u(t)$ such future loads.   We optimize the

following weighted average of objective functions:

$$\text{Min} \quad \sum_{s=t}^{t+T(t)} \alpha_{t,s} z_s \tag{4.11}$$

s.t.   (4.3) – (4.10)

In (4.11), parameter $\alpha_{t,s}$ is a nonnegative weight assigned to a future train s, for $t \leq s$

$\leq t+T(t)$.   The vector of weights, $\widetilde{\alpha}(t) := (\alpha_{t,t}, \alpha_{t,t+1}, \ldots, \alpha_{t, t+T(t)})$, specifies how future

trains are included in the loading decision. For example, $\tilde{\alpha}(t) = (1, 0, 0, \ldots, 0)$

corresponds to the trivial case where we optimize and load the departing train t only,

while $\tilde{\alpha}(t) = (1, 1, 0, \ldots, 0)$ corresponds to optimizing two trains t, $t+1$ together and

loading train t only. Ideally, we want to define $\tilde{\alpha}(t)$ in a way such that the objective in

(4.11) is a weighted average of short-term (currently departing train) and long-term

(future trains) objectives. To achieve this, we propose to use exponentially decreasing

weights:

$$\tilde{\alpha}(t) = (1, \alpha_t, \alpha_t^2, \ldots, \alpha_t^{T(t)}), \tag{4.12}$$

where α_t is a scalar such that $0 \leq \alpha_t \leq 1$.

Then, (11) becomes

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

(4.13)

which is a weighted average of z_t, $(z_t + z_{t+1})$, $\ldots$, and $\sum_{s=t}^{t+T(t)} z_s$. If most load information

is already known and there are few unknown loads on the horizon (i.e., $n_u(t) << n_k(t)$), we

should choose $\alpha_t \approx 1$, such that $\tilde{\alpha}(t) \approx (1, 1, \ldots, 1)$ and $\sum_{s=t}^{t+T(t)} \alpha_{t,s} z_s \approx \sum_{s=t}^{t+T(t)} z_s$, to exploit

the efficiency from optimizing multiple trains together. In the limit, this scenario

converges to the static optimization case where full future information is available. On the other hand, if we expect a large number of unknown loads on the horizon (i.e., $n_u(t) >> n_k(t)$), we should choose $\alpha_t \approx 0$, such that $\tilde{\alpha}(t) \approx (1, 0, 0, \ldots, 0)$ and $\sum_{s=t}^{t+\tau} \alpha_{t,s} z_s \approx z_t$, to avoid the penalty due to future uncertainty.

The weight scalar α_t can vary over time and across the train index t. Its appropriate value can be calibrated from historical data over repeated experiments or simulations for any existing IM facility. When empirical data are not available, a reasonable value need be estimated. Note from (4.11) and (4.13) that the value of α_t controls the balancing between the short-term objective (regarding immediate train departure) and long-term importance (future trains to be loaded). It reflects the relative significance of static information versus dynamic information, which is closely related to the concept of "degree of dynamism" (DOD) introduced in Lund et al. (1998) and Larsen (2001) — the proportion of dynamic information at the time of decision. We propose an "adjusted DOD," defined as follows:

$$DOD = {n_u(t)} \Big/ {(n_k(t) + n_u(t))} \qquad \forall t. \tag{4.14}$$

And, we propose using α_t as the following:

$$\alpha_t = 1 - DOD = {n_k(t)} \Big/ {(n_k(t) + n_u(t))} \qquad \forall t. \tag{4.15}$$

For every optimization, the value of DOD is determined based on not only the numbers of known loads, n_k, but also the estimated number of unknown loads, n_u. For example, if n_k is large enough for three consecutive outgoing trains, then n_u will be the number of estimated future incoming loads from now until the decision time for the third outgoing train. Therefore, a uniform arrival of loads and trains' departure time would result in a constant DOD whereas a non-uniform arrival and departure time would lead to adaptive DODs.

4.3 Model Extensions and Operating Rules

4.3.1 Level of Service

The model thus far treats all loads as equally important by assuming each load can be placed on any of the trains. This assumption is reasonable given the frequent service on many IM routes; however, sometimes there may be certain loads with a higher priority than others. Or, railroads may promise their customers that loads making the cutoff time will be loaded onto one of the next several trains. These specific operational practices can be accommodated by adding level of service (LOS) constraints during data preprocessing. For example, if we know there are 30 UPS trailers that must make it onto the first outgoing train, we add constraint $\sum_{i \in UPS} \sum_{j} \sum_{p} \sum_{k} y_{ij1pk} \geq 30$ to the original model.

More generally, the LOS constraints can be enforced as follows:

$$\sum_t \sum_p \sum_k d_t y_{ijtpk} \leq S \quad \forall i, j \tag{4.16}$$

Where d_t is the departure time of train t, and S is the service level in hours. This constraint ensures that every load will be assigned within the service level. The constraints can also be modified to impose this rule on specific loads by changing i to "$i \in$ *specific loads* ". Similarly, if S is defined in terms of the number of train departures (i.e., the load can be delayed by at most S train departures) and load ij makes to the cutoff time of train which is denoted by G_{ij}, the following constraint ensures load ij being assigned in the next S trains.

$$\sum_t \sum_p \sum_k y_{ijtpk} t \leq G_{ij} + S \quad \forall i, j \tag{4.17}$$

4.3.2   Blocking and Loading before Cutoff Time

As mentioned in section 2.1, the rail IM business in North American is different from the general freight business in the sense that Railroads try to avoid intermediate switches and stops because this business is highly time sensitive. Therefore, the initial loading pattern for most trains will be the principal factor affecting their aerodynamic performance for all or most of the trip.

Blocking is usually defined on a car basis prior to loading assignment. Therefore, for the small number of cases in which the IM train does make intermediate stops, we can set y_{ijtpk} to zero if load ij is not supposed to be on the kth unit of train t priori to the optimization. This would possibly expedite the solution process by reducing the size of the decision space.

Occasionally, the actual time required to load the whole train is longer than the time period from cutoff to departure; therefore, terminal managers have to start loading trains before the cutoff time. Some site-specific loading rules can be developed to ensure aerodynamic efficiency under this circumstance. For yards handling only containers or mixed IM loads, managers should try to load the shorter containers (≥ 40') in the lower positions of well cars. This is because the aerodynamics of longer container atop shorter containers are better than the opposite (shorter atop longer). Thus, holding longer containers for upper level slots will generally be a good option. For yards handling only trailers, managers should assign loads that best match the available slots beginning at the front of the train to guarantee an aerodynamic loading pattern. With above rules, managers can still apply the proposed load-assignment models with available loads and slots at cutoff time, and approach to the system optimum.

4.4 Model Implementation and Case Study

In the following sections, we first apply the static model to evaluate aerodynamic efficiency from optimal loading at the system level, assuming static information of trains and loads. Then, we use the dynamic model to analyze continuous terminal operations

when information is dynamic.   The dynamic model is implemented for two different cases: (1) a terminal with uniform arrival rate of incoming loads; and (2) a terminal with non-uniform arrival rate of incoming loads.

4.4.1   Static Case with Perfect Information

In the static case, we conducted an analysis of 16 trains with 90-minute departure intervals ranging from 84 to 122 units (mean = 104), and 4,224 loads (both domestic and international IM loads) in a 24-hr window.   Trains originated from a major IM terminal in one 24-hour period.   The numbers and types of available loads for each train were obtained from data provided by the railroad.   Five possible scenarios were conducted to evaluate the benefit of optimizing more trains together.   They are to optimize one, two, four, eight and sixteen trains at a time, assuming perfect information for all trains and loads.   CPLEX 10.0 incorporated with GAMS (Brooke et al., 1998) was used to solve the model in reasonable time.   For example, the 16-train scenario (with 45,477 variables and 8,605 effective constraints after data preprocessing) was solved to optimality within 0.922 seconds by a 2.26GHz CPU and 1.5 GB of RAM.

Considering only one train at a time is consistent with current terminal practice. However, with perfect information, the more trains that are optimized at a time, the better the aerodynamic efficiency (Figure 4.4); although, the marginal benefit declines considerably beyond four trains.   Terminal managers' goal in load assignment is to maximize slot utilization; therefore, they are largely indifferent to alternative loading patterns as long as they achieve 100% slot utilization and comply with applicable loading rules.   Consequently, with 100% slot utilization, average terminal operating practice will

be some intermediate value between the minimal and maximal total adjusted gap length. We assumed that the mean of the total adjusted gap length for the minimum and maximum cases represents average terminal loading performance in scenarios with 100% slot utilization (Lai et al., 2007), and is therefore equal to 31,822' (Figure 4.4). Compared with this baseline performance in which load assignment is based on maximizing slot utilization only, the benefit of optimizing aerodynamic efficiency of IM trains ranges from 14% to 46%.   Since scenarios 2 to 5 (optimizing multiple trains) are more beneficial than scenario 1, the fuel savings are also more significant.
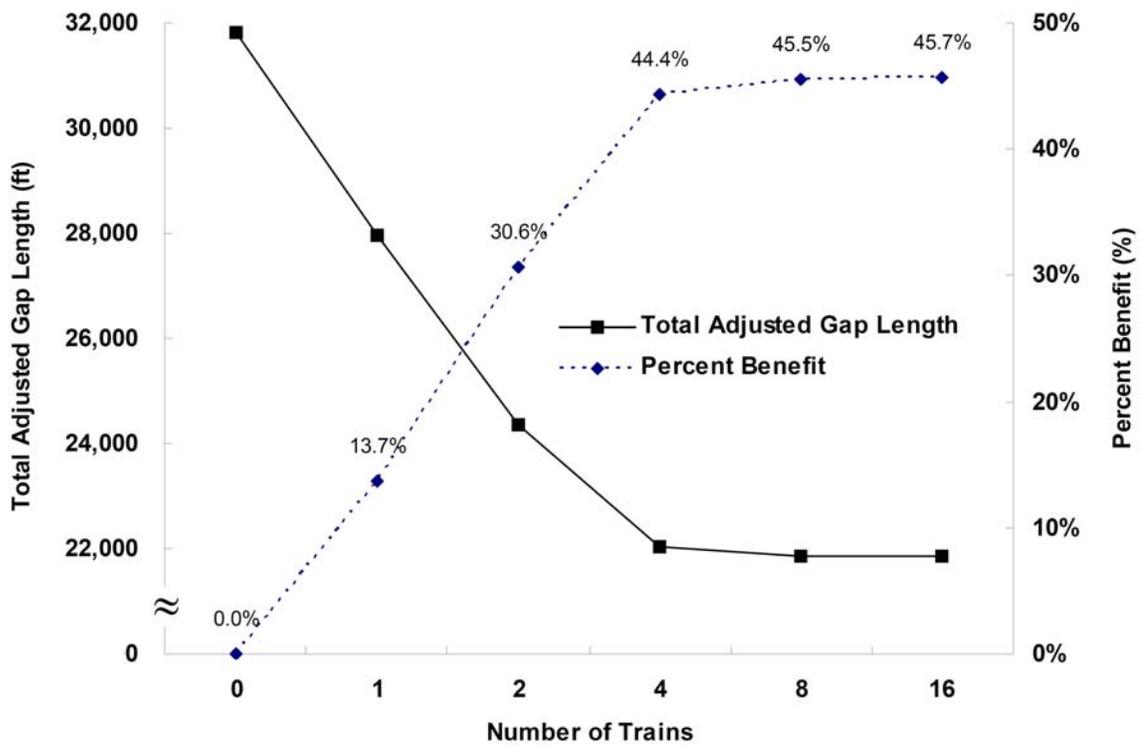
Figure 4.4:   Effect of the number of trains optimized simultaneously on the adjusted gap length.

Placing loads on different trains is generally feasible due to the frequent service of IM operations.   Lai et al. (2007) estimated that optimizing one train at a time can save 15 million gallons of fuel per year compared to current operations; and, optimizing multiple trains together shows the potential to further improve savings by an additional 30% (Figure 4.4).   In other words, if there is no flexibility in placing loads on different trains, the potential 30% savings is lost.

In practice, loads often arrive at and trains depart from terminals quickly, with little lead time. Therefore, it is rarely the case that reliable load information will be available for more than three trains at any time.   Consequently, we implement the rolling horizon framework developed in this study for continuous terminal operations in sections 4.4.2 & 4.4.3.

4.4.2   Rolling Horizon Operations with Uniform Arrival Rate

To implement the rolling horizon scheme, we need to know the number of loads initially available and the arrival pattern of additional loads between consecutive train departures.   According to the cutoff time and load information, we assumed that 690 loads (for three trains) are known at the beginning of the 24-hour period with a constant rate of incoming loads.   The constant rate is computed as the sum of all the loads for the 16 trains divided by 16.   The resultant rate is approximately 230 loads per 90 minutes. In other words, there will be 230 new loads incoming before the next optimization and assignment since the train departure interval is also 90 minutes.   The detailed distribution of load numbers (by type) is shown in Table 4.2.

Table 4.2:   The distribution of number of loads for all types of IM loads.

Load Type	Containers					Trailers						Total
	C20	C40	C45	C48	C53	T20	T28	T40	T45	T48	T53	
Initial number of loads	81	225	18	60	255	6	9	6	6	9	15	690
Incremental loads / 90 min	27	75	6	20	85	2	3	2	2	3	5	230

In this case, $T(t) = 2$ is true for all trains except the final two.   The adjusted DOD within the time window of this system is approximately $230 \times 2/(690+230 \times 2) = 0.4$ most of the time, and we simply recommend using $\alpha_t = 1 - 0.4 = 0.6$, for all t.   As proposed, we consider up to three trains in each optimization, and load only the current outgoing train.   The unassigned loads and future trains will be incorporated into the next optimization, along with any new loads.   Figure 4.5 shows the result of implementing the rolling horizon scheme for this empirical example.   The dotted line is a benchmark representing the objective value for a static assignment case where one train is optimized at a time with 230 available loads (as described in Lai et al., 2007).   The dashed line shows the idealized scenario with full information, where 16 trains are optimized together. This objective value is the best performance possible and serves as a reference point to evaluate the performance of the proposed rolling horizon framework.   The solid line in Figure 4.5 shows the experiments with the rolling horizon scheme where we vary α_t over 0 to 1.

These numerical results verify our qualitative arguments.   When $\alpha_t = 1$, all trains are treated equally in each optimization, and the final objective value is actually about 3% higher than the best possible.   This shows that the optimality of the current outgoing

train is unnecessarily over-compromised by putting too much emphasis on future trains. On the other hand, when $\alpha_t = 0$, the objective value is also about 2% higher than the best possible confirming our argument that there are benefits from considering multiple trains together. The exponential smoothing scheme, however, successfully reduces the objective value to within 0.2% of the best possible for any α_t between 0.1 and 0.6. These values are about 7.5% smaller than the benchmark value for the one-train-a-time strategy, thus demonstrating that the proposed rolling horizon scheme with exponentially decreasing weights is beneficial.
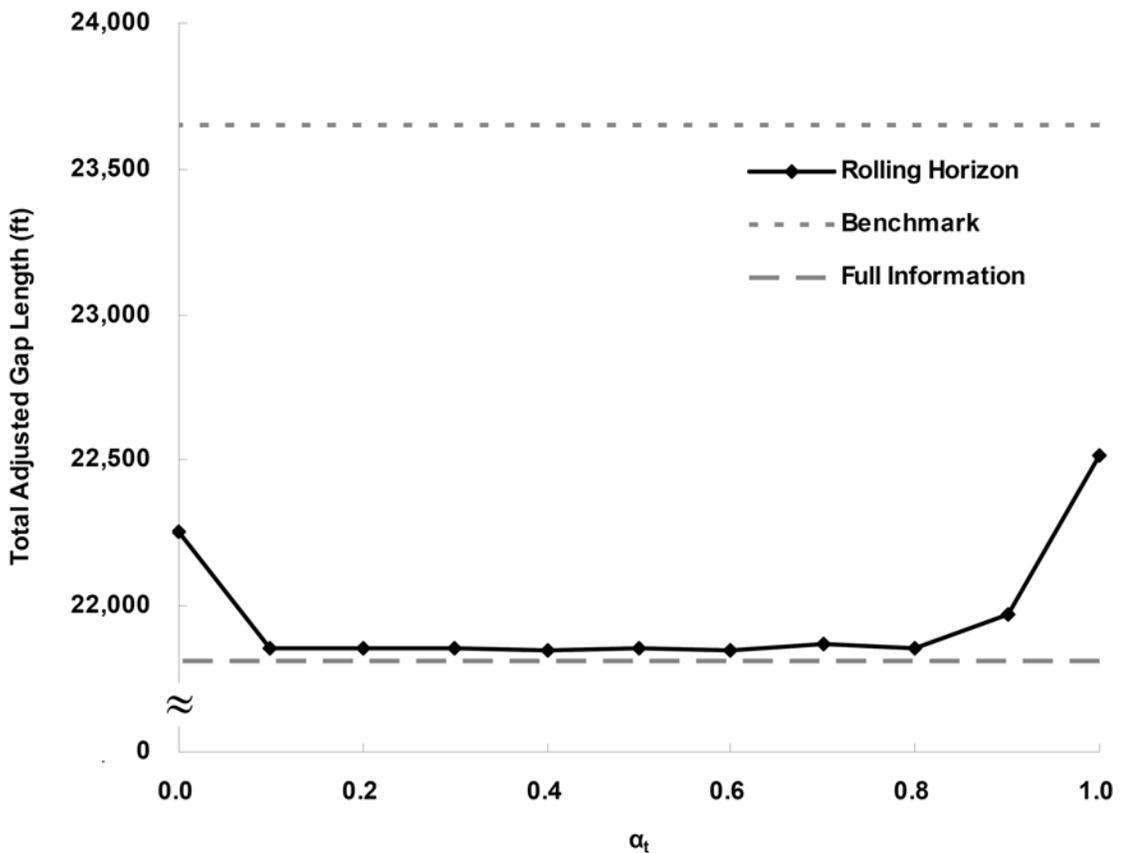
Figure 4.5: Rolling horizon scheme on uniform operation with different α_t values.

The numerical experiments also reveal interesting insights into the choice of the weight parameter. Since DOD ≈ 0.4 at all times, we propose that a possible α_t value be 1–DOD = 0.6, and that $\widetilde{\alpha}(t) = (1, 0.6, 0.36, \ldots), \forall t$. Figure 4.5 shows that the optimal objective value is actually insensitive to α_t for a wide range, $0.1 \leq \alpha_t \leq 0.8$. This finding indicates that an appropriate value of α_t can be calibrated from historical data; and if historical data are not available, $1 - DOD$ would be a good choice. Compared to optimizing one train at a time ($\alpha_t = 0$), rolling horizon operations yield a 2.2% fuel savings in this example, or approximately 160,000 gallons of fuel per year for trains on the single route considered in this analysis.

The computation time and optimality gap by CPLEX 10.0 does vary across optimization instances. For example, when $\alpha_t = 0.6$, 12 out of the 16 instances are solved to within 0.1% relative optimality gap in less than 1 CPU second. Note that only part of the solution for each instance (regarding the current outgoing train) is finally implemented into the overall solution throughout the horizon. Although the other four instances have relative optimality gap between 1% - 3% after 600 CPU seconds (pre-determined upper limit), the overall quality of the rolling horizon solution is close to the known optimum (with full information and zero optimality gap) (Figure 4.5). This computational performance is also found to be true for the example in the next section.

4.4.3 Rolling Horizon Operations with Non-Uniform Arrival Rate

In section 3.2, we implemented the rolling horizon scheme for a terminal with a hypothetical uniform load arrival and train schedule. In this section we consider a

terminal with non-uniform conditions.   IM loads typically do not arrive at terminals

uniformly throughout the 24-hour day cycle (Figure 4.6).   Instead, more loads arrive by

day then at night, with the peak at 12 noon at the terminal we studied.   A detailed

breakdown of the incoming loads by type and intended departure time is presented in is

presented in Tables 4.3 & 4.4.   The number of units in the ten trains range from 93 to

122 units (mean = 110).   The cutoff time is assumed to be 2 hours before departure, so

the 11 a.m. train can draw from the pool of loads not assigned after the 5 a.m. train, plus

those newly arrived from 3 a.m. to 9 a.m.
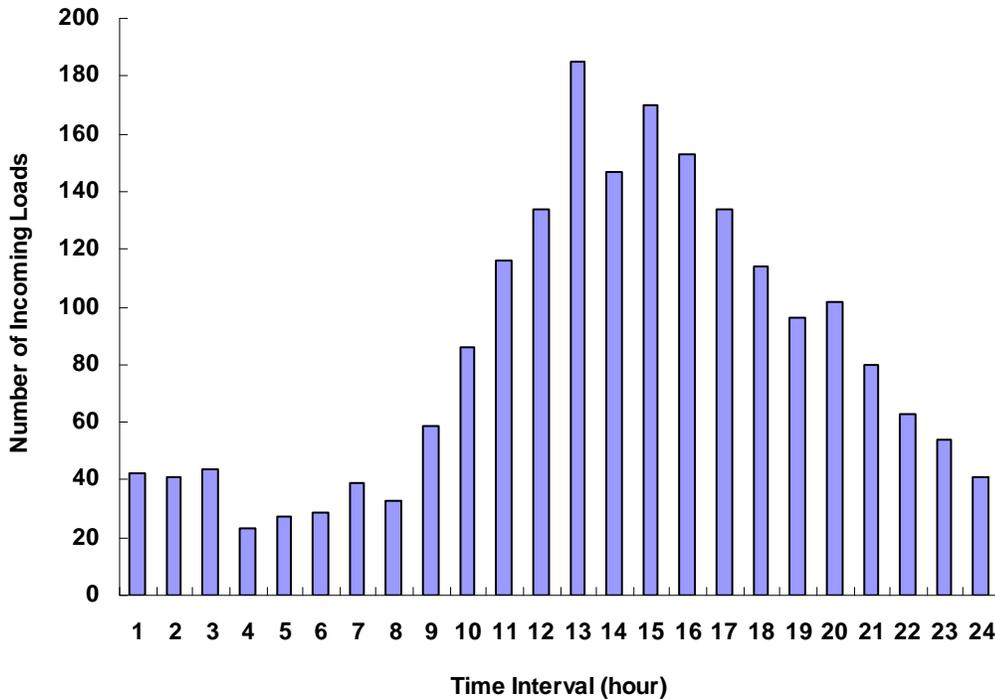
Figure 4.6:   Distribution of incoming loads by time of day.

Table 4.3:   Number of incoming loads by time and type

Time	Load Type					Total
	C40	C45	C48	C53	T40	
Initial	144	4	12	168	72	400
1:00	15	0	1	18	8	42
2:00	15	0	1	17	7	40
3:00	16	0	2	18	8	44
4:00	8	0	1	10	4	23
5:00	10	0	1	11	5	27
6:00	10	0	1	12	5	28
7:00	14	0	1	16	7	38
8:00	12	0	1	14	6	33
9:00	21	1	2	17	11	52
10:00	31	1	3	36	15	86
11:00	42	1	3	49	21	116
12:00	48	1	4	56	24	133
13:00	67	2	6	78	33	186
14:00	53	1	4	62	26	146
15:00	61	2	5	71	31	170
16:00	55	2	5	64	28	154
17:00	48	1	4	56	24	133
18:00	41	1	3	48	21	114
19:00	35	1	3	40	17	96
20:00	37	1	3	43	18	102
21:00	29	1	2	34	14	80
22:00	23	1	2	26	11	63
23:00	19	1	2	23	10	55
0:00	15	0	1	17	7	40

Table 4.4:   Departure time of outgoing trains

	Depature Time
Train 1	1:00
Train 2	5:00
Train 3	11:00
Train 4	13:00
Train 5	15:00
Train 6	16:00
Train 7	18:00
Train 8	20:00
Train 9	21:00
Train 10	0:00

We examine the effect of (1) varying the constant factor ($0 \le \alpha_t \le 1$), and (2) adaptive α_t in this experiment.   The inventory level is always enough for two trains, i.e., $T(t) = 1$, throughout the decision horizon; hence, we consider two future trains in each optimization, and load only the current outgoing train.   The adjusted DOD for each of the 10 assignments depends on the current inventory level and the next outgoing train's departure time.   For instance, the DOD of the 5 a.m. train is $201/(201+324) = 0.617$ because it has 324 loads available in the pool, and there will be 201 new loads incoming to the terminal before the cutoff time of the next outgoing train (11 a.m. train).

Figure 4.7 shows the result of implementing the rolling horizon scheme to this empirical example, for a range of static α_t.   The results of applying the rolling horizon scheme to non-uniform operations are largely similar to uniform operations.   Either placing too much emphasis on future trains (with $\alpha_t = 1$) or myopically ignoring future trains (with $\alpha_t = 0$) compromises the optimality of the solution.   Again, the optimal objective value is relatively insensitive to the value of α_t for a wide range, $0.1 \le \alpha_t \le 0.9$. The static case with applying adaptive DOD (changing α_t over time) yields results very close to $\alpha_t = 0.6$.   This again demonstrates that if historical data are not available, $\alpha_t = 1$ – DOD would be a good choice.   Compared with optimizing one train at a time ($\alpha_t = 0$), using rolling horizon operations yield an 8.6% benefit, or approximately 700,000 gallons of fuel savings per year for the IM trains at this route.
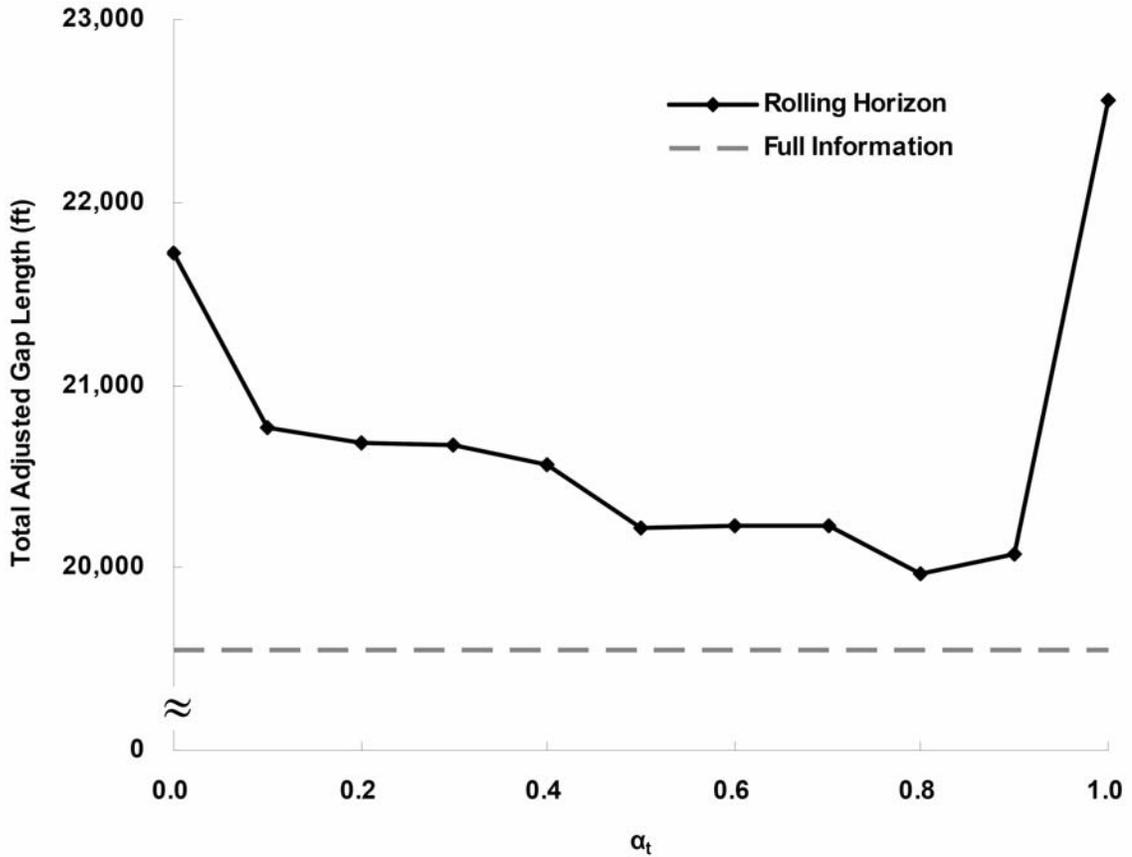
Figure 4.7:   Rolling horizon scheme under non-uniform conditions for different α_t.

4.5 Discussion

Optimizing multiple trains together has the potential to further improve the fuel savings compared to optimizing one train at a time.   The greater the flexibility in placing loads on different trains, the better the aerodynamic efficiency.   However, the marginal benefit drops considerably beyond optimizing four trains at a time (Figure 4.4).   As expected, the marginal benefit is greatest when comparing "no flexibility" (i.e., optimizing one train at a time) to "flexibility in several trains".   Beyond that, there are diminishing returns because there are not enough new types of load choices available to yield additional benefit in train aerodynamics.

Railroad IM terminals usually use computer software (e.g. OASIS) to assist the loading assignment process; therefore, integration of our proposed model into the software currently being used would not require significant institutional or process change. It also should have little if any impact on operating cost because the general process remains the same; the only difference would be terminal managers' decisions on which load should be assigned to which slot to maximize fuel efficiency. We believe implementation of the proposed rolling horizon optimization scheme can automate the terminal managers' tasks regarding the large variety of loads and railcar types, thereby enabling them to load trains in a more aerodynamically efficient manner.

The experiments described above have thus far treated loads with equal importance, ignoring potential constraints on loads' time priorities. It is worthwhile to examine the effect of LOS constraints on the outcome as described in Section 2.4.1. Using the same data from Section 3.3, we now assume that the railroad promises to its customers that all loads arriving before the cutoff time will be loaded onto one of the next three departing trains (i.e., $S=3$). We compared the results in aerodynamic efficiency (total adjusted gap length) and delay (in units of departed trains and hours) of imposing or not imposing LOS constraint.

The optimization results show that imposing an LOS constraint reduces the aerodynamic efficiency of IM trains by 36%. This is expected because LOS constraint reduces the flexibility of assigning loads to different trains. The cumulative delay of

loads (in units of departed trains and hours) with or without the LOS constraints is shown

in Figures 4.8 & 4.9.   With the LOS constraints, the delays are more uniform across

loads, but the total delay of all loads is actually slightly longer.   For example, forcing

one trailer to be placed in an IM well car due to the LOS constraints would probably

cause two containers to be left behind.   The impact on delay would decrease if there was

enough equipment dedicated to trailers (e.g. spine or flat cars); however, this will often

not be the case in practice.   Thus, LOS constraints will generally have a significant

impact on both energy efficiency and operational cost.
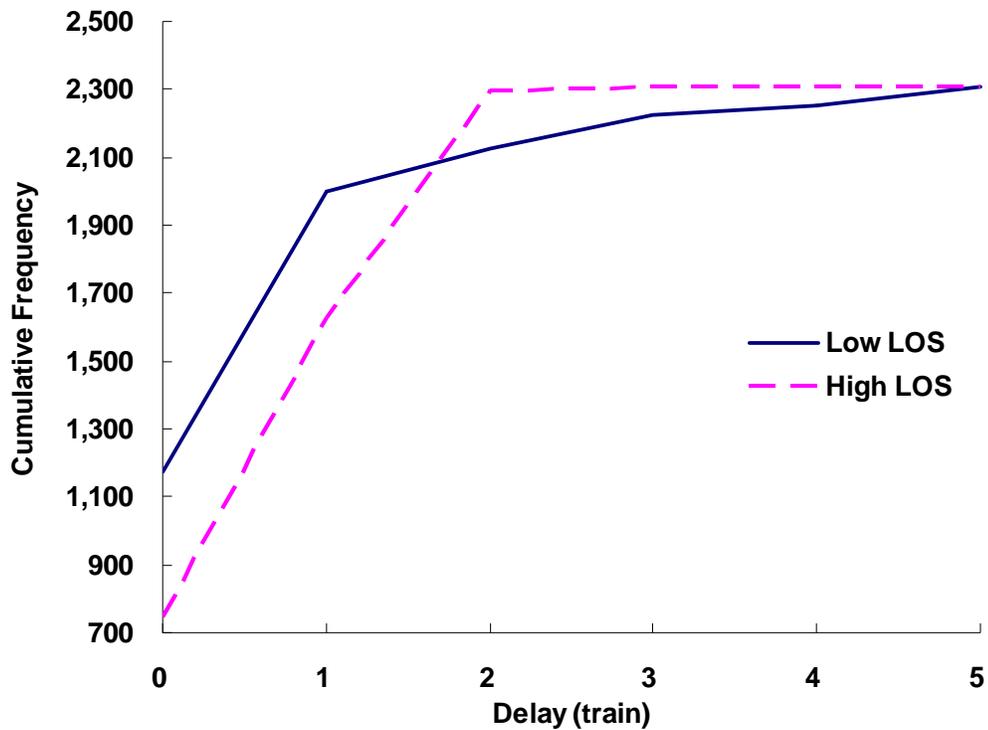
Figure 4.8:   Delay with or without LOS constraints in terms of trains
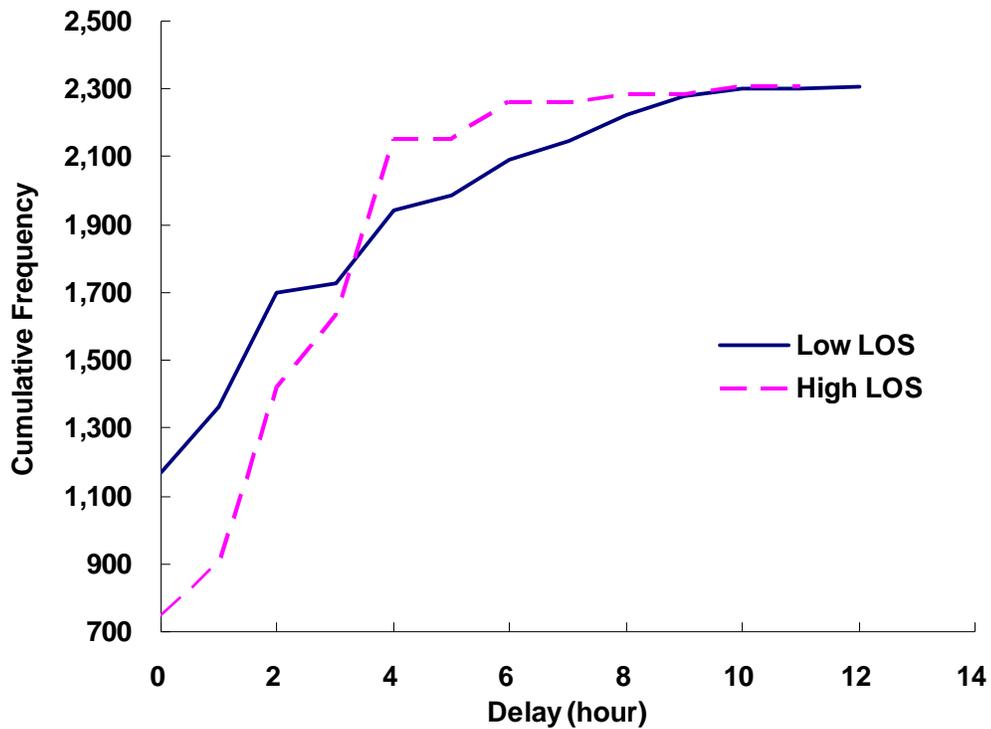
Figure 4.9:   Delay with or without LOS constraints in terms of hours

4.6 Conclusion

This chapter presents static and dynamic aerodynamic efficiency models for the loading of multiple IM trains.   It also develops a rolling horizon scheme for continuous train terminal operations.   For the static case, when full information is available, the system optimum can be reached by optimizing as many trains as possible.   In practice, however, terminals operate in a dynamic environment where not all information on incoming loads and trains is available.   Attempting to optimize the loading of too many trains in this environment will reduce the ability to achieve the most efficient loading configuration.   Therefore, a rolling horizon scheme with decreasing weight assigned to each train is proposed to counterbalance the effect of uncertainty.   Numerical results

show that the realistic rolling horizon scheme significantly reduces the adjusted gap length as compared to the current practice, which leads to a substantial benefit from aerodynamic efficiency of IM trains. Correspondingly larger savings in fuel, emissions and expense are possible if the methodology described in this chapter could be applied to all North American IM trains.

CHAPTER 5

OPTIMIZING RAILWAY CAPACITY PLANNING

"Everything should be made as simple as possible, but no simpler." – Albert Einstein

Railways all over the world are increasingly experiencing capacity constraints. In North America, railway freight traffic has increased nearly 30% over the past 10 years, and this demand is projected to increase another 88% by 2035 (AASHTO, 2007). This raises the question, how can railroads handle this additional traffic on a network that is already experiencing capacity constraints? There is a variety of engineering options that can be used, either singly or in combination, to increase network capacity, such as adding tracks, adding or lengthening sidings, modification of traffic control systems, etc.

To improve capacity using infrastructure upgrades, the North American railroad industry generally relies on experienced personnel and simulation software to identify bottlenecks and propose alternatives to reduce congestion (HDR, 2003; CN, 2005; Vantuono, 2005). Experienced railroaders often identify good solutions, but this does not guarantee that all good alternatives have been evaluated or that the best one has been found. Furthermore, the aging demographics of the railroad industry means that many experienced capacity analysts will soon retire. Simulation can model a section of the network in great detail but it is not suitable for network capacity planning. Instead of

solving the real problem, solutions based on corridor-based simulation analyses may move bottlenecks to other places in the network.

A good decision support tool for railway capacity expansion projects should have the ability to generate and evaluate possible expansion alternatives, and suggest an optimal capacity expansion plan at the network level by minimizing the cost of increasing capacity subject to the estimated future demand.

Several recent studies proposed methods to compare different investment alternatives in transportation systems.   Felipe et al. (1996) developed a multicommodity, multimodal network design model to determine investment priorities for a freight intercity network. Jelaska (1998) proposed a capacity planning support model to evaluate the investment impacts for a range of options.   Fransoo and Bertrand (2000) developed an aggregate capacity estimation model to compare alternatives for investments in infrastructure investment, specifically passing sidings, which can single out the most promising investment alternatives without the time-consuming simulation process.   Petersen and Taylor (2001) presented a method formulated as nested dynamic programming models for determining the optimal timing and economic feasibility of a new railway line in Brazil. Delorme et al. (2001) developed a constraint programming model and a unicost set packing model to evaluate railway infrastructure capacity.   Putallaz and Rivier (2004) presented a methodology and the basis for the development of an effective decision support system; their method deals the planning of investments in capacity and also takes into account the impacts of timetables on maintenance and renewal policies.   Wahlborg (2004) calculates

the capacity consumption based on the UIC 405 model for current and future traffic and infrastructure.

These methodologies are able to compare different proposed alternatives by various means; however, they do not have the ability to create possible alternatives, implement a network flow model to identify the location and method required for upgrade, and evaluate the tradeoff between capital investment and delay. This is the incentive for development of a new decision support tool in this research.

In this chapter, I start by reviewing state-of-art, rail-line-capacity-analysis methodologies, and then develop a decision support framework to help capacity planners determine how to optimally allocate capital for capacity expansion at the network level. A Class 1 railroad's network is usually divided into divisions assigned to different superintendents for operating purposes; and, these divisions are further divided into hundreds of subdivisions or "subs", which represent segments of track ranging from 300-mile mainlines to 10-mile branch-lines. The objective of this research is to develop a framework that can successfully identify the optimal investment plan regarding which subdivisions need to be upgraded and what kind of engineering options should be conducted based on the estimated future demands, available budget, and network properties. Such a decision support framework will help railroads maximize their return from capacity expansion projects and thus be better able to provide reliable service to their customers, and return on shareholder investment.

5.1 Railway Line Capacity Analysis Methodologies

The aim of capacity analysis is to determine the maximum number of trains that would be able to operate on a given infrastructure under a particular set of operational conditions during a specific time interval. Railway line capacity cannot be considered as a static value; instead it is highly dependent on a number of infrastructure and operational factors (Krueger, 1999; Vantuono, 2005), such as:

- Length of subdivision

- Siding spacing and uniformity

- Intermediate signal spacing

- Percentage of single, double, or multiple track

- Peak train counts

- Average and variance speed

- Traffic mix

- Dispatching priorities

- Schedule

Numerous approaches and tools have been developed to determine rail line capacity; however, unlike the highway capacity analysis domain, there is no commonly accepted standard measurement for railway capacity analysis (Abril et. al., 2007). Each model has its strengths and weaknesses and is generally designed for a specific type of analysis (Martland and Hutt, 2005).

In general, railway capacity tools can be categorized into three groups: (1) theoretical (2) detailed simulation, and (3) parametric. Theoretical models are typically the simplest among the three, and can often be computed manually. Although the model is simple, it is useful for a quick evaluation of the line and for example can capture the relative effects of different signal systems and operating practices on capacity. By contrast, simulation is the most sophisticated and computationally intensive, and requires specific, detailed, train and network data. The simulation result is the closest representation of the actual operations. Therefore, there is a tradeoff between use of theoretical versus simulation methods. Parametric models fill the gap between simple theoretical models and detailed simulation by focusing on the key elements of line capacity so as to quickly highlight "bottlenecks" in the system (Krueger, 1999). They are more efficient than simulation and represent the real world better than theoretical models.

5.1.1 Theoretical Capacity Models

The theoretical capacity of a rail line can be considered as the maximum number of trains that can traverse it during a specified time period. This type of capacity model often assumes homogenous train speed and characteristics, few or no service disruptions, directional or alternative running, and few or no meets or passes. The advantage of using theoretical capacity models is that they are simple and require minimum computation; however, the outcome may lack accuracy due to various simplifying assumptions. These simple capacity models are most useful for comparing the relative effects of different alternatives at the strategic planning level, or for computing transit

capacity since the assumptions are more accurate in the simpler transit operating domain (TCRP, 1996).

Theoretical models usually start with calculation of minimum train spacing (a.k.a. headway) based on the signal block length and train speed.   Line capacity, also called "Maximum Throughput", is then computed by dividing headway from the specific time period (AREA, 1947; Conant, 1964).   For example, the Air Brake Association (ABA) (1972) model is developed in this format, and it has two equations for single and bidirectional operation, respectively.   The AREMA (2002) model is also similar to the ABA model, but it further defines "Practical Capacity" as a fraction of theoretical capacity (Krueger, 1999).

In transit capacity analysis, numerous theoretical models have been developed and used.   TCRP Report 13 has a comprehensive review of North American rail transit capacity analysis methodologies (TCRP, 1996).   A major difference between transit and conventional, especially freight railroad capacity is that station dwell time plays an important role in transit systems.   The Canadian Urban Transit Association (1985) model is basically the same as the maximum throughput method except that the dwell time is incorporated into minimum headway calculation.   Alle's (1981) model uses real data from a busy station in New York City, and it produces results that are close to actual experience without applying any of the judgment factors used in many other calculation methods to calibrate theory with practice.   The disadvantage is that only one station was examined, and that real data are not always available, especially for new designs.

In light rail transit, models also need to consider the ratio of traffic light green time to the cycle length because they are running on the street (Levinson, 1994). Vuchic (1981) develops comprehensive mathematical formulae for transit operations; however, according to the review of TCRP Report 13, "the difference between theory and practice is difficult to reconcile or quantify with other mathematical treatments", and Vuchic's model sometimes produces inaccurate results (TCRP, 1996).

A number of more sophisticated models have been developed to allow relaxation some of the unrealistic assumptions in the simplest theoretical models. Petersen (1974) developed a single-track analytical model assuming that the departure times are uniformly distributed over the time period. Petersen and Taylor (1982) presented a methodology that performs probabilistic analyses of dispatch patterns, and it was later extended by Kraft (1988). Chen and Harker (1990) attempted to estimate the delays of given traffic flows in a stochastic single track environment and was extended by Harker and Hong (1990) to include a partially double-track corridor. Pachl (2002) proposed "Blocking Time Calculation", which can accommodate traffic mix but requires detailed information of blocking time overlaps. Ekman (2004) presented an analytic method to estimate the capacity of new infrastructure based on a discrete event model of train paths. Kozen and Burdett (2005) incorporate the average delay of each train on each section into the sectional running time and the standard bottleneck analysis; their model uses directional distribution for bidirectional operations, four stopping protocols for acceleration and deceleration movements, and train distribution for the traffic mix.

They later extended the model to take into account different types of railway characteristics (e.g. traffic mix, signal locations, and dwell times). Improverail Consortium (2003) also proposed a capacity model (UIC 405) that handles traffic mix by calculating the expected number of sequences. UIC 405 model is less complicated than Kozen and Burdett's model, and has been used in the Banverket infrastructure investment plan for 2003 and 2015 (Wahlborg, 2004).

Optimization is another approach to determine railway line capacity analytically. For given infrastructure conditions, optimization models are used to identify the line capacity based on the most compact schedule (create train paths as closely as possible, a.k.a. optimal saturated timetables) (Szpigel, 1972; Assad, 1980; Jovanovic and Harker, 1991; Harker and Hong, 1994; Cai and Goh, 1994; Carey and Lockwood, 1995; Higgins et al., 1996; Oliveira and Smith, 2000; Caprara et al., 2002). A similar strategy, timetable compaction method, is also used in the UIC 406 model, proposed by the International Union of Railways (UIC) (UIC, 2004). The UIC 406 model schedules the existing train paths as closely as possible to each other by modifying the base timetable, so the remaining unused time left in the timetable represents the maximum time available during which additional train services can theoretically be scheduled (Landex et al. , 2006).

As might be expected, some of these enhanced theoretical models or optimization models become too complex to solve manually, which is one of the main attractions of

using theoretical models.   Capacity analysts are responsible for selecting suitable evaluation methods for their particular applications.

5.1.2   Simulation Models

Since the 1960s, computer simulations have been used for rail line capacity analysis (Bronzini and Clarke, 1985).   The power of simulation is that it enables modeling of complex systems with a high degree of reality.   For rail applications, it can be used to evaluate track configurations, signal systems, operating strategies, and various other parameters of interest, either alone or in combination. .

Simulation models generally mimic train dispatcher logic and are used to evaluate infrastructure and/or operational changes (ASSHTO, 2005).   They usually follow a set of fixed rules governing train priorities and a train performance calculator.   By providing track configuration, signal systems, and operating plans as input, an experienced user can evaluate the outputs to determine bottlenecks and conflicts. Adjustments can then be made to the inputs to resolve these conflicts.

There are a number of railroad simulation models with different features and logic. The most popular one at present is Rail Traffic Controller (RTC), developed by Berkeley Simulation Software (Wilson, 2008).   The RTC logic assigns trains through the network according to their priority.   When there are conflicts, the logic seeks alternative routes for the lower priority train.   The results have been validated with hundreds of real-world

networks.   One of the most well-known recent applications is the Chicago CREATE project (Thompson, 2006).

The Route Capacity Model (RCM) was developed by the Canadian National Railroad.   It is a computationally efficient model that can complete 500 runs in 10 seconds and is thus well-suited for system-level capacity analysis.   The user can specify the desired level of statistical confidence (e.g. 95%) and then RCM will automatically run the simulations until the confidence level is achieved.

Other models include RAILSIM by SYSTRA, RAILS 2000 by CANAC, FastTrack II by MultiModal Applied Systems, Rail Dispatch and Capacity Analysis Model (RDCAM), RailSys by Rail Management Consultants, Open Track by Institute for Transport Planning and Systems in Zurich, and etc.

5.1.3 Parametric Models

Simulation is best suited to analysis of local-level problems; however, it becomes computationally difficult when applied at the network level.   On the other hand, theoretical models are sometimes too simple to be valid.   Parametric capacity models are intended to fill the gap between detailed simulation and simple formulae.   They focus on key elements of line capacity to quickly highlight "bottlenecks" in the system (Krueger, 1999).

Prokopy and Rubin (1975) developed the first parametric model for railway line capacity. It utilizes formulae that reflect train delay or capacity as a function of physical plant train operations and control systems. The formula is derived through multi-variable regression analysis of many different simulation runs using the Peat Marwick Mitchell (PMM) model.

Krueger (1999) applied a similar method to establish the CN Parametric Line Capacity Model; however, he developed a totally different parametric capacity model with different parameters. Simulations were conducted with the Route Capacity Model (RCM) instead of using the PMM model. The three most important elements of the CN parametric model that makes it particularly useful are: (1) the ability to calibrate each parameter for each/any scenario (it is not a fixed program); (2) produce a graphical delay versus volume relationship; and (3) "What-if" ability to quantify the sensitivity/significance of individual parameters and in combination. In the CN model, parameters are categorized into plant, traffic, and operating variables. The model can recognize the dynamic nature of capacity and provides a system wide capacity measure of subdivisions in a rail network. This enables comparison of different parts of the network to identify areas of limited or excess capacity (Krueger, 1999).

5.2   A Decision Support Framework for Railway Capacity Planning

The objective of this research is to develop a framework to generate and evaluate possible capacity expansion alternatives and consider the capacity planning problems at

the network level.   This tool will help capacity planners determine how to optimize

allocation of capital for capacity expansion projects in a rail network.

The framework comprises three modules: (1) an "Alternatives Generator (AG)" that

enumerates possible expansion options along with their cost and capacity effects; (2) an

"Investment Selection Model (ISM)" that determines which portions of the network (at the

subdivision level) need to be upgraded with what kind of capacity improvement

alternatives; and (3) an "Impact Analysis Module (IAM)" that evaluates the tradeoff

between capital investment and delay cost (Figure 5.1).   These three components can be

used separately as stand-alone tools, or they can be combined as the decision-support
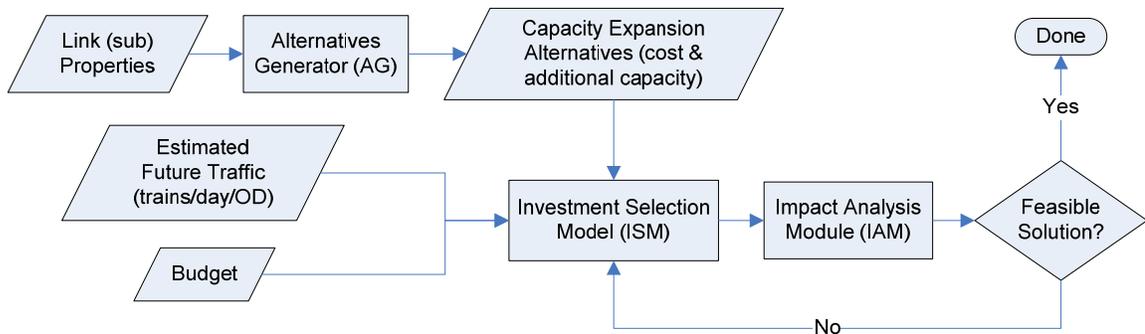
framework.

Figure 5.1:   Decision support framework for railway capacity planning

Table 5.1 shows the inputs and outputs of each module in the decision support

framework.   Based on the link properties (plant, traffic, & operating parameters), the AG

enumerates possible expansion alternatives for each link with the associated costs and

capacity increases.   The ISM then combines this information with the estimated future

demand and available budget to determine the best set of investment options for the network, assuming the level of service remains the same. Finally, The IAM evaluates the tradeoff between capital investment and delay cost to determine if the capital investment is cost-effective. The output will be a set of options that the capacity planner can use to guide decision making. In the following sections, these three modules are demonstrated in more detail.

Table 5.1: Inputs and outputs of each module in the decision support framework

	Input	Output
AG	Plant, Traffic, Operating Parameters	Capacity Expansion Alternatives
ISM	Demand, Budget, Expansion Alternatives	Investment Selection Set
IAM	Budget, Investment Selection Set	Exapnasion Benefit Table

Instead of solving capacity problems based solely on corridor analyses, this decision-support framework is able to conduct network-based analyses to ensure that the system fluidity and interaction between corridors are taken into account. The decision support framework is a brand new process to determine the best strategy of network capacity planning. Within this framework, I developed an enhanced parametric model (AG) to evaluate capacity and generate expansion alternatives; a novel optimization model (ISM) to determine the necessary investment plan; and an original sensitivity analysis procedure (IAM) to evaluate the tradeoff between capital investment and delay cost.

5.3   Alternatives Generator (AG)

In order to explore possible capacity expansion options, we need a tool to evaluate the current state of each subdivision in the network and then generate expansion

alternatives with associated capacity increases and construction costs.   Several capacity

expansion models use simple linear or empirical functions to represent the cost and

increase in capacity of different alternatives (Felipe et al. 1996; Ordonez and Zhao, 2005).

This may work for highway systems because the cost of the infrastructure is proportional

to length and number of lanes, and the capacity of directional traffic movement can be

calculated based on empirical formulae (TRB, 2000).   However, increasing rail line

capacity using a combination of the number of sidings, tracks and signals is more

complicated than a simple linear model (Krueger, 1999).   More accurate model that can

accommodate these aspects is needed to evaluate the capacity increase and the

corresponding cost (AASHTO, 2005).

As discussed in section 5.1, theoretical models are usually too simple to

incorporate all the important line capacity factors, and simulation is too complex for

network-level planning.   Therefore, I chose to adapt the CN parametric line capacity

model as the evaluation tool for rail traffic impact.   In terms of its accuracy and ease of use,

it falls between theoretical formulae and simulation, and incorporates key elements of line

capacity (Krueger, 1999).   Although the CN model can assess capacity based on network

link properties, it is not able to enumerate possible alternatives and compute the

associated construction costs.   Therefore, I enhanced the CN parametric model in this

research to be a complete AG by adding enumeration and cost functions.

Although we chose the CN model for this study, it is not the only tool that can be used

or enhanced as the AG.   Railroads that have their own capacity analysis tools can

substitute these for the CN parametric model as long as they possess similar functionality. Knowledge from industry experts is also an excellent source of capacity expansion options; these empirical alternatives can be incorporated into the final expansion alternatives table in addition to options created by the AG.

5.3.1 Review of CN Parametric Model (Krueger, 1999)

The CN parametric model accounts for the dynamic nature of capacity. It provides a system-wide measure of the subdivisions in a rail network and allows evaluation of the effect of capacity improvement for different alternatives. The model measures the capacity of a subdivision by predicting its relationship between train delay (hours per trip) and traffic volume (trains per day). Generally, the more trains are running on a subdivision for a given time period, the more delay every train experiences over the trip (Prokopy and Rubin, 1975). The CN model achieves this relationship by using those key parameters affecting the traffic handling capability of a subdivision. These key parameters are categorized into plant, traffic, and operating parameters as follows:

Plant Parameters
- Length of Subdivision (SL)
- Meet and Pass Planning Point Spacing (MPPPS):

    MPPPS is the mean spacing of locations used to meet or overtake trains, namely siding spacing. Sidings are crucial for operating bi-directional, mixed priority and different speed trains. MPPPS for a subdivision is computed as:

$$MPPPS = \frac{Length\ of\ Subdivision}{(Number\ of\ MPPP\ +\ 1)} \tag{5.1}$$

- Meet and Pass Planning Point Uniformity (MPPPU):

MPPPU is the measure of uniformity in siding spacing (MPPPS).   With same or similar speed limit across the subdivision, the higher the uniformity of MPPPU, the more the line capacity.   It is a ratio of the standard deviation versus average siding spacing:

$$MPPPU = \frac{Standard\ Deviation\ of\ MPPP\ Spacing}{MPPPS} \tag{5.2}$$

A uniformity value of zero represents a subdivision with equally distributed sidings.

- Intermediate Signal Spacing Ratio (ISSR):

Intermediate signals reduce the required headway between adjacent trains so as to increase line capacity.   This parameter relates the ratio of signal spacing to the siding spacing.   In the CN parametric model, sidings are assumed to have one signal located at the center of the siding.   Below is the parametric expression for ISSR:

$$ISSR = \frac{\left(\dfrac{Length\ of\ Subdivision}{MPPP\ +\ 1\ +\ \#\ of\ Signals}\right)}{MPPPS} \times 100 \tag{5.3}$$

- Percent Double Track (%DT):

Doubling track has a significant impact on line capacity (more than double the capacity of a single track mainline).   Sidings length of 6,000 ft or less were not

used by the original modelers as a useable passing siding but are part of the double track segments.   %DT is calculated as the ratio of double track versus the length of the subdivision:

$$\% \ DT \ = \ \frac{Miles \ of \ Double \ Track}{Length \ of \ Subdivision} \times 100 \qquad (5.4)$$

Note that the CN parametric model can handle %DT up to a limit of 75%; this limit was found to retain the exponential characteristics and fall within the parametric range of most of CN's subdivisions.

Traffic Parameters

- Traffic Peaking Factor (TPF):

  TPF represents the concentration of traffic within a short time frame (4 hours), often called bunching or peaking.   It has a significant impact on capacity, because when the traffic level is greater than the sustainable capacity, it results in a considerable system recovery time.   TPF is calculated as the ratio between the maximum number of trains dispatched in a 4-hour period versus the average number of trains within the same time duration.

$$TPF \ = \ \frac{Maximum \ Tains \ in \ 4 \ hours}{Average \ Trains \ in \ 4 \ hours} \qquad (5.5)$$

- Dispatching Priority Factor (DPF):

  Dispatching priorities for different types of trains dictate which trains will experience delay.   Priority shortens the transportation time of higher priority

trains by penalizing lower priority trains.   Generally the greater the number of

priority classes, the less capacity is available.   DPF is quantified using a

probability function that identifies the chances of a train meeting another train of

a higher priority, which is calculated as:

$$DPF \; = \; \frac{1}{T} \sum_{i=2}^{N} \left(\frac{C_i}{(T-1)} \sum_{j=1}^{i-1} C_j \right) \qquad\qquad (5.6)$$

Where:

N = Number of priority classes (passenger, express, freight, and unit)

T = Daily number of trains

C_i = Number of ith priority class trains

C_j = Number of jth priority class trains

- Speed Ratio (SR):

Besides DPF, speed ratio is another parameter reflecting the traffic mix over the

subdivision.   The difference in speed among trains can significantly increase

delay because of overtakes and/or holding trains in yard.   SR is calculated as the

ratio of the fastest train speed to the slowest train speed:

$$SR \; = \; \frac{Fastest\ Train\ Speed}{Slowest\ Train\ Speed} \qquad\qquad (5.7)$$

- Average Speed (AS):

Average train speed plays a vital role in line capacity because the higher the train

speed the lower the delay and transit time.   AS is measured as the average

minimum run time of all trains in each direction, as obtained from a Train

Performance Calculator (TPC).

$$AS = \frac{\sum\limits_{i=1}^{N} n_i V_i}{\sum\limits_{i=1}^{N} n_i} \qquad (5.8)$$

Where:

V_i = Speed of ith class

n_i = Number of trains in i^{th} class

N = Total number of classes

Operating Parameters

- Track Outage (TO):

Track outages accounts for the planned and unplanned events that take a track out

of service.   TO directly reduces the available service time of a subdivision as

well as line capacity.   Capacity is sensitive to the occurrences and duration of

TO.   This parameter is defined as the number of hours the subdivision is out of

service:

$$TO's = \frac{Total\ Duration\ of\ Outages}{\sum\limits_{i=1}^{N} \dfrac{1}{n_T d_i}} \qquad (5.9)$$

Where:

n_T = Total number of outages per day

d_i = Duration of each outage (hrs)

Temporary Slow Order (TSO):

TSO has a negative impact on line capacity due to: (1) the time loss due to operating at slower than normal speed; and (2) acceleration and deceleration time (V time).   It is often maintenance related and can be applied to a distance or at a single point on the line.   TSO is computed as follows:

$$TSO = V_{time} + Travel\ Time \qquad (5.10)$$

$$V_{time} = \frac{(V_m K - V_{TSO})}{A} + \frac{(V_m K - V_{TSO})}{D} \qquad (5.11)$$

$$Travel\ Time = \left(\frac{L}{V_{TSO}} + \frac{L}{V_m K} \right) \times 60 \qquad (5.12)$$

Where:

V_m = Maximum freight speed (mph)

V_{TSO} = Temporary slow order speed (mph)

K = % of time running at max speed (85%)

A = Acceleration rate (20 mph/min)

D = Deceleration rate (30mph/min)

L = Length of TSO + average train length

The relationships between "delay-volume curve" and "key parameters" were developed based on a series of regression analyses and simulation results from the RCM. The relationship between train delay and traffic volume was found to be best expressed by the following exponential equation:

$$Train\ Delay = A_o e^{B_o V} \tag{5.13}$$

Where:

A_o = Parametric plant, traffic, operating coefficient

B_o = Constant

V = Traffic Volume (trains/day)

Coefficient "Ao" depicts the relationship between train delay and the parametric values. "Ao" is a unique value for each combination of parameters defined by the plant, traffic and operating conditions of a subdivision. A different "Ao" will define a new delay vs. volume curve (Figure 5.2). This parametric model was verified by comparing its output to the RCM output of the CN network, and the results show that the accuracy was on average within 10 %.
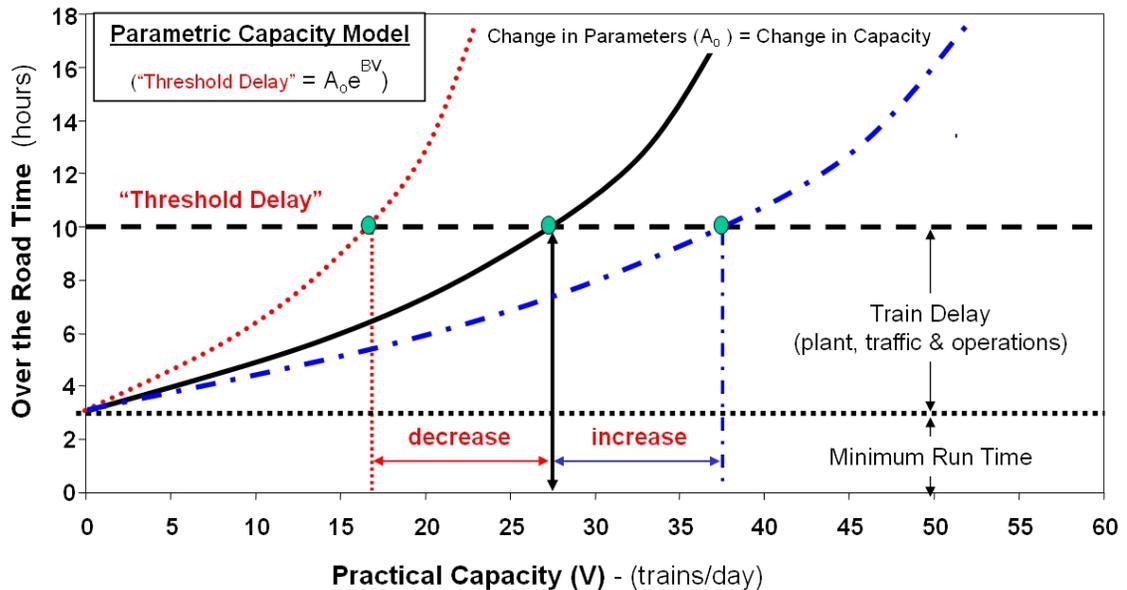
Figure 5.2: Delay versus volume curve

121

There are two versions of the CN parametric model built in different environments: (1) window-based program, and (2) Excel spreadsheet. The spreadsheet version is used in this research, so I can easily incorporate the enumeration and cost functions by using Excel/VBA Macro code.

5.3.2 Enumeration Function

The purpose of the enumeration function is to automatically generate conventional capacity expansion alternatives for each subdivision in the network based on its current properties. Three common types of capacity expansion alternatives are built into this module: adding (1) passing sidings, (2) intermediate signals, and (3) 2nd main track. For the single track scenario, increasing the number of sidings can reduce meet and pass delay, and increasing the number of signals and shortening block length can reduce the headway between trains thereby increasing line capacity. Beyond that, according to Rollin Bredenberg (V.P. Service Design at BNSF Railway), if demand averages 60 trains per day with a peak of 75, double-track must be added to single-track segments (Vantuono, 2005; AAR, 2007).

For each subdivision, the enumeration function will calculate all possible combinations of expansion alternatives until it reaches the limit of minimal siding spacing or maximal number of signals per spacing specified by the user (Figure 5.3). For example, for a 100-mile CTC subdivision with nine existing sidings and no intermediate signals, if the minimum siding spacing is set to eight miles and the

maximum number of signals within each spacing is two, the largest number of sidings

that can be placed in this subdivision is 11 ($\approx 100/8 - 1$), and the largest number of

intermediate signals that can be placed (between two sidings) is two.   Table 5.2 shows

the possible alternatives for this example; the order of enumerated alternatives is based

on ascending construction costs.   Since adding signals are usually less expensive than

adding sidings, adding signals is considered first (up to the limit) before adding another

siding; therefore, the first and second alternatives are to increase the number of

intermediate signal in every spacing by one and by two, respectively.   Since two

intermediate signals is the upper bound for number of signals per spacing, the next (third)

alternative is to increase the number of sidings (by one).
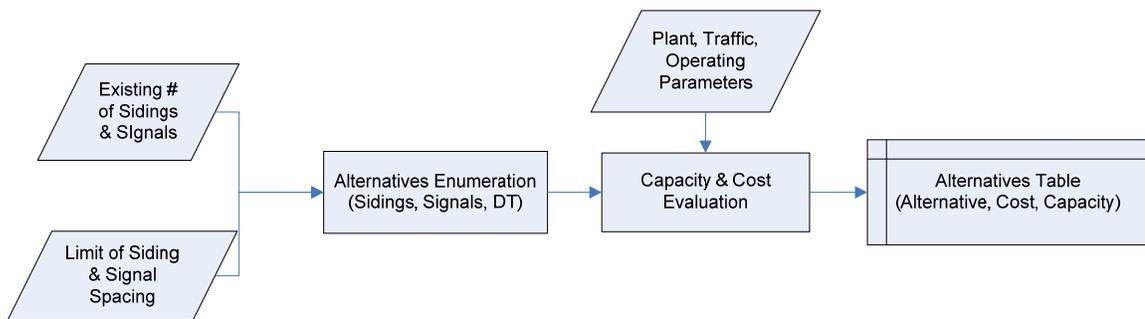
Figure 5.3:   Flowchart of alternatives generator

Table 5.2:   Possible capacity expansion alternatives

Alternatives	Sidings	Signals/Spacing
1	+ 0	+ 0
2	+ 0	+ 1
3	+ 0	+ 2
4	+ 1	+ 0
5	+ 1	+ 1
6	+ 1	+ 2
7	+ 2	+ 0
8	+ 2	+ 1
9	+ 2	+ 2
10	Adding 2nd Main Track	

5.3.3 Line Capacity Evaluation

After the enumeration, the next step is to evaluate the capacity increase and construction cost of each alternative (Figure 5.3).   For each subdivision, AG will first evaluate the current line capacity based on the existing key parameters.   Capacity planners usually have an idea of the current line capacity based on empirical experience.   These empirical values can be used in AG to determine the current LOS by adjusting the acceptable delay to match the capacity values from both AG and empirical experiences.   If empirical values are not available, the maximum trip time of 10 hours will be used to calculate the capacity (Figure 5.2) (Krueger, 1999).   However, users can specify their own suitable limits depending on the context in which it is used.

After obtaining the base case (current condition), AG can then compute the capacity increase of each alternative by changing the plant parameters (e.g. MPPPS & ISR), assuming the traffic and operating parameters remain the same.   The CN parametric model cannot handle subdivisions with %DT more than 75%; consequently, I assigned a capacity of 80 trains per day for a double-track segment according to typical railroad industry freight railroad practices (Vantuono, 2005; AAR, 2007).

5.3.4 Construction Cost Estimation

The unit construction cost of each type of expansion options is needed to compute the cost of expansion alternatives.   Users can specify these values in advance or use the

default cost estimates.   Three required basic unit costs are: the costs of (1) adding a new siding, (2) adding a new intermediate signal, and (3) adding a 2nd main track.

The general default cost estimates are based on information obtained from railroads and consulting companies.   These values serve as the general average case considering the need for new tracks, signals, and bridges, but ignore the cost of acquiring additional land or environment permit.   For a new 12,000-foot passing siding, I assumed as cost of $4,870,000 for track work and civil infrastructure.   For territory with an existing CTC signal system, the cost of signalizing a newly constructed siding within this territory would be $300,000 for each end of the siding or $600,000 in total.   Therefore, the first required unit cost, cost of adding a signalized passing siding, is $5,470,000.   Within existing CTC territory, the cost of a new intermediate signal point (i.e. one signal in each direction) is approximately $100,000 (second required unit cost).   And, the third required unit cost, that of adding the 2nd main track, is $2,250,000 per mile.

5.3.5   Output of Alternatives Generator – Alternatives Table

Table 5.3 lists the alternatives for the subdivision mentioned in 5.3.2.   For our network analysis, a similar table is generated for each subdivision in the network, and these tables together form the input data for the investment selection model.   Ideally, capacity planners would review these alternatives before they are input into ISM. During the process, planners can remove inadequate alternatives or add additional alternatives based on their experience and judgment.

Table 5.3:   Expansion alternatives with capacity increase and construction cost

Alternatives	Sidings	Signals/Spacing	Capacity (trains/day)	Cost
1	+ 0	+ 0	+ 0	$0
2	+ 0	+ 1	+ 3	$1,000,000
3	+ 0	+ 2	+ 4	$2,000,000
4	+ 1	+ 0	+ 3	$5,470,000
5	+ 1	+ 1	+ 6	$6,570,000
6	+ 1	+ 2	+ 7	$7,670,000
7	+ 2	+ 0	+ 6	$10,940,000
8	+ 2	+ 1	+ 9	$12,140,000
9	+ 2	+ 2	+ 10	$13,340,000
10	Adding 2nd Main Track		+ 50	$204,750,000

5.4   Investment Selection Model (ISM)

The ISM is developed to identify which subdivisions need to be upgraded with what type of improvements in the network by using optimization and network analysis techniques.   Trains with different origins and destinations are similar to multiple commodities and they share common line capacity; therefore, we formulate this problem as a mixed integer network design model (Magnanti and Wong, 1984; Minous, 1989; Ahuja et al., 1993).   Based on the estimated future demands of all OD pairs and capacity expansion options, the ISM determines an optimal investment plan for capacity expansion with the premise that "level of service" remains the same as the current conditions.   In other words, there is no difference between delay (hours per train) before expansion with existing traffic, and delay after expansion with the future demand.

5.4.1   The General Investment Selection Model

The following notation is used in the investment selection model: i is an index referring to the starting node of an arc, and j is the ending node of an arc; k corresponds to the kth origin-destination (OD) pairs of nodes (s_1, t_1), (s_2, t_2), …, (s_k, t_k) in which s_k and t_k

denote the origin and destination of the kth OD pair; q represents the index of engineering alternatives. B is the available budget for capital investment; α and γ are the weights to account for the planning horizon; c_{ij} is the cost of running on arc (i, j). d_k is the demand of kth origin-destination pair; h_{ij}^q represents the cost of the qth engineering option on arc (i, j); U_{ij} is the current capacity of arc (i, j); and u_{ij}^q is the increase in capacity of arc (i, j) by the qth engineering option.

There are two sets of decision variables in the ISM. The first variable is denoted by x_{ij}^k, which is the number of trains running on arc (i, j) from the kth OD pair. The second variable is a binary variable, denoted by y_{ij}^q, which determines whether the qth engineering option is used for arc (i, j), namely:

$$y_{ij}^q = \begin{cases} 1, & \text{if the } q^{th} \text{ engineering option is used for arc } (i,\ j) \\ 0, & \text{otherwise} \end{cases}$$

The investment selection model is formulated as follows:

$$min\ \alpha \sum_i \sum_j \sum_q h_{ij}^q y_{ij}^q + \gamma \sum_i \sum_j \sum_k c_{ij} x_{ij}^k \tag{5.14}$$

s.t.

$$\sum_i \sum_j \sum_q h_{ij}^q y_{ij}^q \leq B \tag{5.15}$$

$$\sum_k x_{ij}^k \leq U_{ij} + \sum_q u_{ij}^q y_{ij}^q \qquad \forall\ i,j\ (i \neq j) \tag{5.16}$$

$$\sum_q y_{ij}^q \leq 1 \qquad \forall\ i,j\ (i \neq j) \tag{5.17}$$

$$\sum_j x_{ij}^k - \sum_j x_{ji}^k = \begin{cases} d_k & \text{if } i \in s_k \\ -d_k & \text{if } i \in t_k \\ 0 & \text{otherwise} \end{cases} \qquad \forall\ k \tag{5.18}$$

and

$$x_{ij}^k \geq 0,\ x_{ij}^k \in integer,\ y_{ij}^q \in \{0,1\}$$

The objective function takes into account the capital investment, residual value, and flow cost; it aims to minimize the total cost that includes both the 'net cost' of new infrastructure and the total traffic flow cost over the planning horizon.   The net cost is defined as the difference between capital investment and the residual value of new infrastructure at the end of the horizon.   The residual value is usually a fraction of the initial capital investment, so it is embedded in the first part of the objective function, together with capital investment.   The total flow cost is the summation of transportation cost and maintenance of way (MOW) cost.   Besides the necessary investment to accommodate future demand, additional capital investment might further reduce the total flow cost.   The relative importance of these depends on the planning horizon.   The longer the planning horizon, the more railroads should be willing to invest due to larger reductions in flow cost over time.   Therefore, the appropriate weights $(\alpha \text{ & } \gamma)$ should be determined by using the life cycle cost analysis (LCCA) method (Acharya et al., 1991; Zoeteman and Esveld, 1999; Viet, 2002; Lee, 2002; Ozbay et al., 2004; Ling et al., 2006) according to the implementation circumstances.   For a multiyear capacity planning project, we need to take into account the increase in flow cost over time and the discount factor to compute the net present value; therefore, c_{ij} in this general formulation is a set of discounted flow costs by year within the planning horizon.   The determination of $\alpha, \gamma,$ and the discounted flow cost are further discussed in 5.4.2.

Constraint 5.15 is the budget constraint, which can be removed if the task is to determine how much funding would be required to meet the estimated future demand.

Equation 5.16 is the line capacity restriction ensuring that the total flow on arc (i, j) is less than or equal to the current capacity plus the increased capacity due to upgraded infrastructure. For each arc (i, j), there can be at most one selected engineering option (constraint 5.17). Finally, equation 5.18 is the network flow conservation constraint guaranteeing that the outflow is always equal to the inflow for trans-shipment nodes; otherwise, the difference between them should be equal to the demand of that OD pair.

5.4.2   The Investment Selection Model for a Class 1 Railroad

In this section, I used the network for a North American Class 1 Railroad as an example to further discuss how to determine α, γ, and the flow cost in ISM. Figure 5.4 is a general timeframe for multiyear capacity expansion projects. At year zero (m_0), the decision maker must make the capital investment decision; the new infrastructure will be completed after the construction lead time (L years), and the impact on operational cost will last for N years based on the planning horizon. At the end of the horizon, we need to account for the residual value of the new infrastructure.
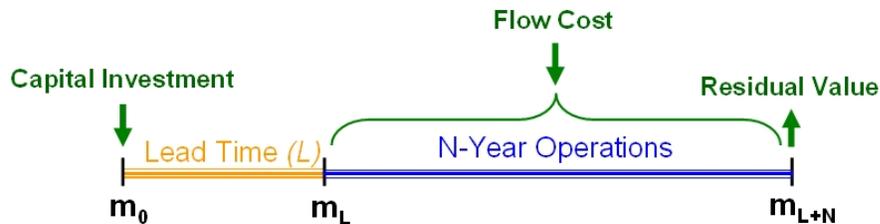
Figure 5.4:   Timeframe of multiyear capacity expansion projects

Determination of Alpha

The first part of the objective function accounts for both capital investment and residual value. Residual value is usually computed as:

$$\frac{Remaining\ Service\ Life}{Total\ Service\ Life} \times Initial\ Capital\ Investment \qquad (5.19)$$

Since net cost is defined as the difference between capital investment and the residual value, it can be computed as:

$$Net\ Cost\ =\ (1 - \frac{Remaining\ Service\ Life}{Total\ Service\ Life}) \times Initial\ Capital\ Investment \qquad (5.20)$$

As a result, α in equation 5.14 is:

$$\alpha\ =\ 1 - \frac{Remaining\ Service\ Life}{Total\ Service\ Life} \qquad (5.21)$$

Determination of Discounted Flow Cost

The unit flow cost is a per train-mile operational cost incurred by railway traffic flow; the total flow cost is the summation of transportation cost and maintenance of way (MOW) cost over the planning horizon. According to Grimes & Barkan (2006), the MOW cost should include both ordinary maintenance expense and renewal expenditure. The transportation cost is the train-operation transportation cost. As a result, the unit flow cost can be computed as:

$$c_{ij} = \frac{Annual\ (MOW\ Cost + Transportatoin\ Cost)}{Total\ Train\ Miles} \qquad (5.22)$$

Where:

c_{ij} = The unit flow cost of running on arc (i, j) at the base year ($/train-mile)

Equation 5.22 can be used to compute the flow cost of a selected base year in the past.   In a multiyear planning project, the unit flow cost in the future should be estimated and discounted to the base year.   Due to recent increases in fuel and steel cost, the annual unit flow cost has been sharply higher from 2003 to 2006 (Figure 5.5).   The average yearly increase in unit flow cost is about 11%.   This information is used to estimate the future unit flow cost of the multiyear plan.

$$c_{mij}^{e} = c_{ij}(1+\pi)^{m} \tag{5.23}$$

Where:

c_{mij}^{e} = The estimated flow cost of running on arc (i, j) at year m

π = Increase rate of unit flow cost

Discounting is another essential element of the overall cost benefit analysis (Ozbay et al., 2004), especially for long-term planning projects.   The present value method (a component of discounted cash flow analysis) was first introduced by Wellington (1914). For each year of an N-year operation (Figure 5.4), a specific discounted flow cost is computed based on the estimated unit flow cost and the real discount rate.   If c_{mij}^{e} is the estimated unit flow cost in year m and f is the real discount rate, the discounted flow cost in the base year would be:

131

$$c_{mij} = c_{mij}^{e} \left(\frac{1}{(1+f)^{m}} \right)$$ (5.24)
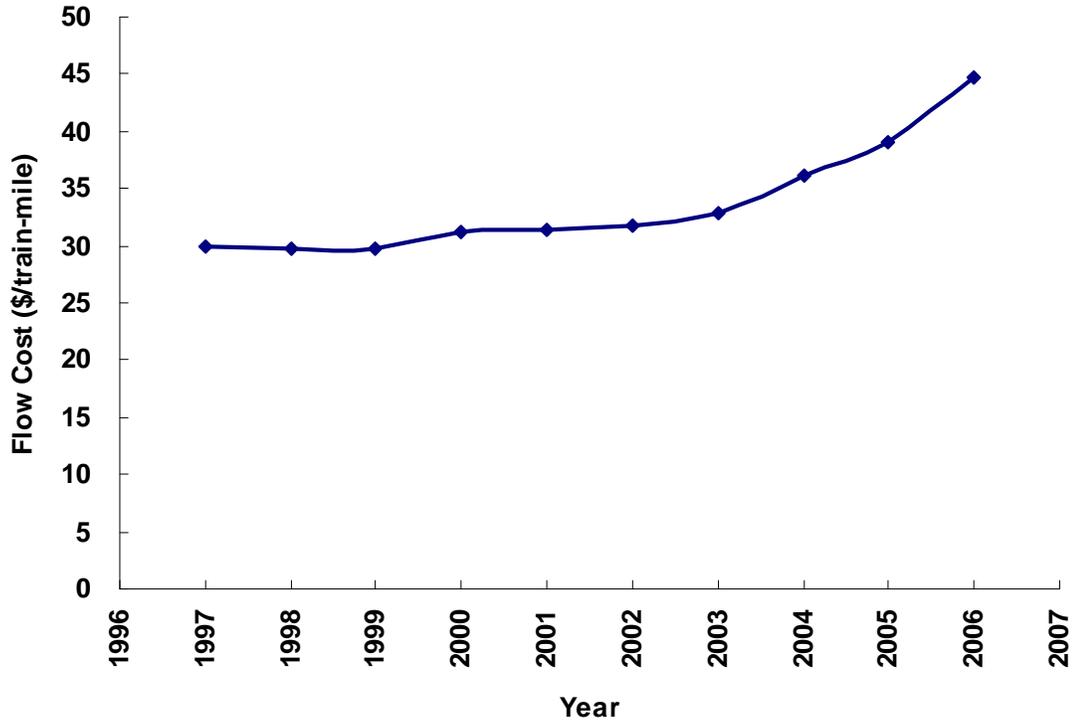
Figure 5.5:   Unit flow cost from 2003~2006

Investment Selection Model for a Class 1 Railroad

North American Class 1 railroads generally operate freight trains according to the base train schedule, which is a guideline of what trains to run on what day (or days) of the week.   This base train schedule can be preprocessed to seven daily traffic flow patterns depicting seven days of the week.   Consequently, the decision variable of the traffic flow in the ISM should be denoted as x_{tij}^{k} depicting number of trains running on arc (i, j) from k^{th} OD pair on day w where w is an index to represent each day of the week (Mon thru Sun).

The following is the ISM for a Class 1 railroad's multiyear planning project. The life of railroad infrastructure is assumed to be approximately 20 years in this example. Hence, if a 5-year operation is considered in this project, the value of $\alpha = 0.25$. γ is determined based on the number of weeks in a year, thus $\gamma = 52$, because the traffic flow pattern is fixed for every week of the year.

$$\min \quad 0.25 \times \sum_i \sum_j \sum_q h_{ij}^q y_{ij}^q + 52 \times \sum_m \sum_t \sum_i \sum_j \sum_k c_{mij} x_{wij}^k \tag{5.25}$$

s.t.

$$\sum_i \sum_j \sum_q h_{ij}^q y_{ij}^q \leq B \tag{5.26}$$

$$\sum_k x_{wij}^k \leq U_{ij} + \sum_q u_{ij}^q y_{ij}^q \quad \forall \; i \,, j \,, w(i \neq j) \tag{5.27}$$

$$\sum_q y_{ij}^q \leq 1 \qquad \qquad \forall \; i \,, j \; (i \neq j) \tag{5.28}$$

$$\sum_j x_{wij}^k - \sum_j x_{wji}^k = \begin{cases} d_{wk} & if \; i \in s_k \\ -d_{wk} & if \; i \in t_k \; , \quad \forall \; w \,, k \\ 0 & otherwise \end{cases} \tag{5.29}$$

and

$$x_{wij}^k \geq 0, \; x_{wij}^k \in integer, \; y_{ij}^q \in \{0,1\}$$

5.4.3 Model Extensions

The scope of this research is to improve network capacity; therefore, node capacity is not considered in the general model. However, if terminal capacity data are available, we can incorporate a node capacity constraint into the original model, which is given by:

$$\sum_k \sum_i x_{ij}^k \leq G_i \qquad \forall \; j \; (j \notin s_k, t_k) \tag{5.30}$$

Where

$$G_i \quad = \text{Capacity of node } i$$

The general model in 5.4.1 would be infeasible if there is not enough budget for upgrading the system to accommodate the future demand.   Two strategies can be applied to cases in which the budget is insufficient.   The first strategy is to assume an unlimited budget while solving the ISM, and then based on the available budget, select the most important links to upgrade as identified using the impact analysis module (section 5.5). The second strategy is to extend the original ISM formulation to incorporate the possibility of having unfulfilled demand (supply < demand), a.k.a. "deficit".   This enables the possibility of rejecting demand if there is not enough budget to accommodate all the forecasted traffic.   The extended model is formulated as follows:

$$min \; \alpha \sum_i \sum_j \sum_q h_{ij}{}^q y_{ij}{}^q + \gamma \sum_i \sum_j \sum_k c_{ij} x_{ij}{}^k + \phi \sum_k \sum_i (d_i^k - a_i^k) \tag{5.31}$$

s.t.

$$\sum_i \sum_j \sum_q h_{ij}{}^q y_{ij}{}^q \le B \tag{5.32}$$

$$\sum_k x_{ij}{}^k \le U_{ij} + \sum_q u_{ij}{}^q y_{ij}{}^q \quad \forall \; i,j \; (i \ne j) \tag{5.33}$$

$$\sum_q y_{ij}{}^q \le 1 \qquad\qquad \forall \; i,j \; (i \ne j) \tag{5.34}$$

$$\sum_j x_{ij}{}^k - \sum_j x_{ji}{}^k = \begin{cases} a_i^k & i \in s_k \; or \; t_k \\ 0 & otherwise \end{cases} \quad \forall \; i \tag{5.35}$$

$$\begin{cases} a_i^k \le d_i^k & if \; d_i^k > 0 \\ a_i^k \ge d_i^k & if \; d_i^k < 0 \\ \sum_i s_i^k = 0 & otherwise \end{cases} \quad \forall \; i \tag{5.36}$$

and

$$x_{ij}{}^k \ge 0, \; x_{ij}{}^k \in integer, \; y_{ij}{}^q \in \{0,1\}$$

The new symbols introduced in this modified ISM are: ϕ is the weight for deficit; d_i^k is the demand of the kth OD pair on node i, where d_i^k is a positive value if i is an origin and a negative value if i is a destination; and a_i^k is the supply of the kth OD pair on node i, so it has to be less than or equal to d_i^k if i is an origin, and greater than or equal to d_i^k if i is a destination.

The objective function is to minimize the sum of net cost, flow cost and deficit. α, γ, and ϕ denote the relative importance of each cost component. For our purpose, ϕ is considerably greater than α *and* γ because it is undesirable to have a deficit if the budget is sufficient to fulfill demand. On the other hand, even if the budget is insufficient, the best scenario is to have the least unfulfilled demand.

Constraints 5.31, 5.32, and 5.33 are the same as 5.16, 5.17, and 5.18 in the general ISM. Constraint 5.19 in the original model is modified into 5.34 and 5.35 in the modified ISM; constraint 5.34 is the flow conservation equation subject to supply (as opposed to demand); constraint 5.35 guarantees that supply is less than or equal to demand. Combining equations (5.34 and 5.35) enables the modified ISM to handle cases without sufficient budget.

5.5   Impact Analysis Module (IAM)

As mentioned above, ISM determines the best set of capacity improvement alternatives with the premise that "LOS remains the same". For example, in Figure 5.6a, the solid exponential curve represents the general delay-volume relationship for the

existing infrastructure, whereas the dashed curve depicts the delay-volume relationship

with upgraded infrastructure. With the same LOS, the upgraded infrastructure can

provide more capacity than the existing track. However, it is also possible to gain

additional capacity by reducing the LOS (increasing delay) (Figure 5.6b). Line capacity is

increased by increasing delay along the delay-volume curve of the existing infrastructure.
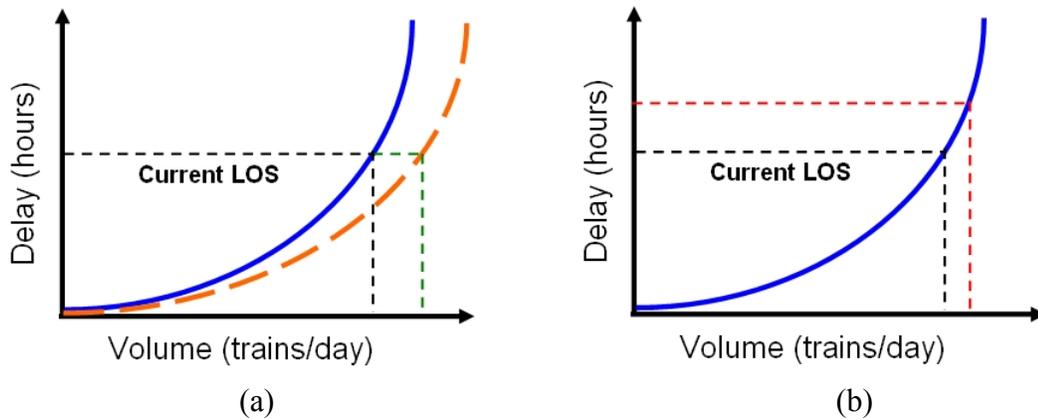
Figure 5.6: Increase volume by (a) upgrading infrastructure (b) lowering LOS

The impact analysis module evaluates whether the capital investment is cost-effective

by comparing the "required capital investment" to the "delay cost". The "required capital

investment" of each link is the output of the ISM. The "delay cost" depends on the impact

of adding additional demand to the existing track layout without upgrading the

infrastructure. According to the new demand for each link obtained from the ISM, the

increase in delay can be determined using the delay-volume curve (Figure 5.6b). We can

then compute the delay cost as the product of total delay hours, and unit delay cost per hour.

From an operational point of view, the unit delay cost can be calculated by summing four

components: (1) unproductive locomotive cost; (2) idling fuel cost; (3) car/equipment cost;

and (4) crew cost.   A recent estimate for one Class 1 railroad is approximately $261 per train-hour in 2007.

After obtaining the delay cost, we rank the importance of each link based on the benefit as defined by delay cost divided by net cost.   A benefit value less than 1 means the investment is not cost effective because the return on investment is negative.   The output of the IAM is a table showing the net cost, delay cost, and benefit for each link subject to capacity expansion.   This expansion benefit table can be provided to the capacity planners for use in their decision-making.

Finally, the problem can be formulated as a "knapsack" problem in which investment decisions are made with a limited budget.   The objective of this model would be to minimize the delay cost subject to a budget constraint.   With a specific budget level (B), the optimal set of investments can be determined by solving the following optimization model:

$$\min \ \text{total delay cost} - \sum_i \sum_j d_{ij} y_{ij} \tag{5.37}$$

$s.t.$

$$\sum_i \sum_j h_{ij} y_{ij} \leq B \tag{5.38}$$

Where d_{ij} is the delay cost due to the increase in future demand on arc (i, j) without upgrading the subdivision, and y_{ij} is the binary decision variable determining whether arc (i, j) is upgraded ($y_{ij} = 1$) or not ($y_{ij} = 0$).

5.6 Case Studies

To demonstrate the potential use of the decision support framework, two case studies were selected, analyzed and are presented in 5.6.1 and 5.6.2. Case study I has a network with 15 nodes, 22 links, and 14 train OD pairs. Among the 22 links, three of them are secondary lines currently serving limited traffic (close to zero). Converting a secondary line into a mainline is costly but, at the same time, it introduces additional routes into the network that may reduce the flow cost of certain trains. In this example, I evaluate the tradeoff between capital investment and flow cost by comparing the results of including or ignoring secondary lines.

Case study II is based on an actual Class 1 railroad network and its traffic data including 39 nodes, 42 links, and over a thousand train OD pairs. It is intended as a feasibility study to investigate computational efficiency of the decision support framework when solving a large scale network problem. Also, the IAM is used to evaluate the tradeoff between capital investment and delay cost assuming there will be a 50% increase in traffic demand. In both case studies, the operational horizon is assumed to be five years, and the increased rate of flow cost is the same as the discount rate, namely the unit flow cost is a constant regardless of the year.

5.6.1 Case Study I

Figure 5.7 shows the selected network in which the nodes represent junctions, and the arcs represent the connecting rail lines. There are two types of links in this network, mainline (denoted by solid lines) and secondary line (dotted lines). As mentioned

before, upgrading secondary lines is costly in terms of capital investment but it may reduce the total flow cost.   In this application, I use both the AG and ISM to determine the optimal capital allocation plan and study the tradeoff between capital investment and flow cost.

In order to use the AG to determine the current line capacity and expansion alternatives, we first need to compute the traffic, plant, and operating parameters.   Table 5.4 shows the values of the key parameters in each subdivision except (14,15), which is already a double-track line.   The AG then uses these parameters to determine the current line capacity based on the link properties (Figure 5.8).
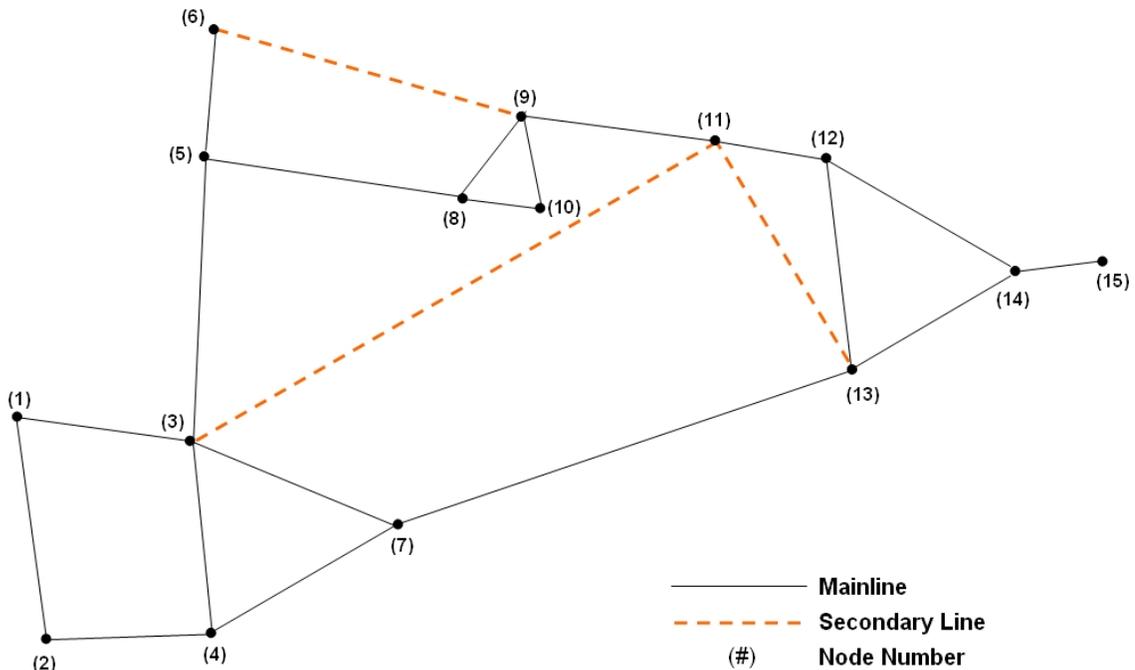
Figure 5.7:   Case study I network

Table 5.4: Key parameters of case study I network of link (i, j)

	(1,2)	(1,3)	(2,4)	(3,4)	(3,5)	(3,7)	(3,11)	(4,7)	(5,6)	(5,8)	(6,9)
SL	100	100	100	100	150	100	300	100	100	150	200
MPPPS	20	14.286	20	10	12.5	12.5	10	16.667	20	21.429	13.333
MPPPU	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.5	0.5
ISS	1	1	1	1	1	1	1	1	1	1	1
%DT	0.04	0.06	0.04	0.09	0.0733	0.07	0.0967	0.05	0.04	0.04	0.07
TPF	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62
DPF	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344
SR	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44
AS	30	32	30	36	36	36	36	36	30	32	31
TO	0	0	0	0	0	0	0	0	0	0	0
TSO	0	0	0	0	0	0	0	0	0	0	0

	(7,13)	(8,9)	(8,10)	(9,10)	(9,11)	(11,12)	(11,13)	(12,13)	(12,14)	(13,14)
SL	200	50	50	50	100	50	100	100	100	100
MPPPS	10	16.667	16.667	16.667	14.286	16.667	10	14.286	10	10
MPPPU	0.5	0.6	0.6	0.6	0.4	0.6	0.5	0.5	0.5	0.5
ISS	1	1	1	1	1	1	1	1	1	1
%DT	0.095	0.04	0.04	0.04	0.06	0.04	0.09	0.06	0.09	0.09
TPF	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62	1.62
DPF	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344
SR	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44	1.44
AS	30	28	28	28	32	28	30	32	30	36
TO	0	0	0	0	0	0	0	0	0	0
TSO	0	0	0	0	0	0	0	0	0	0

SL   : Length of Subdivision

MPPPS   : Meet and Pass Planning Point Spacing

MPPPU   : Meet and Pass Planning Point Uniformity

ISS   : Intermediate Signal Spacing

%DT   : Percent Double Track

TPF   : Traffic Peaking Factor

DPF   : Dispatching Priority Factor

SR   : Speed Ratio

AS   : Average Speed

TO   : Track Outage

TSO   : Temporary Slow Order
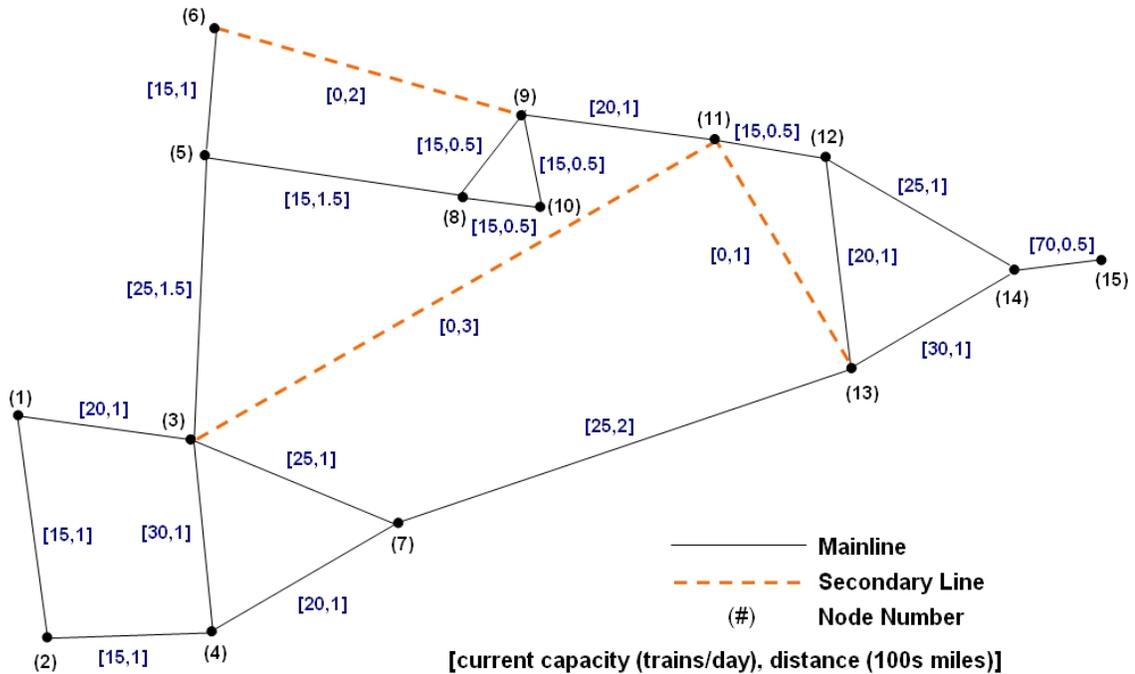
Figure 5.8: Current line capacity and distance of each subdivision

Similar to the process demonstrated in section 5.3, two strategies are considered in this application to increase line capacity:

- Add sidings: one, two, three, etc. (until the spacing is reduced to 8-miles).

- Add intermediate signals: none, one per spacing, two per spacing (at most two signals per spacing on average).

Based on these options, Table 5.5 shows the possible capacity improvement options enumerated by AG for link (1, 3). Similarly, the rest of the single-track mainlines have similar patterns. On the other hand, the cost of upgrading the secondary lines was assumed to be $1 million per mile. Hence, it requires $100, 200, and 300 million to upgrade link (11, 13), (6, 9), and (3, 11) to accommodate 15 more trains per day, respectively.

Besides the expansion options, another set of inputs for the ISM is the estimated future demand.   The demand for each OD pair is expressed as the number of daily trains projected to be needed from origin i to destination j (Table 5.6).   In this example, it is assumed that there is only one type of daily traffic pattern so all days are the same in the study period.

Once the above inputs were defined, the capacity expansion problem can be solved by the ISM.   Two scenarios were considered in this analysis, with or without considering secondary lines as improvement options.   Both cases are coded in GAMS and solved by CPLEX.   Figure 5.9a is the optimal solution for the scenario without secondary lines whereas Figure 5.9b is the scenario with secondary lines.   In Figure 5.9, links with bold numbers are those that require a capacity upgrade, and the number represents the amount of additional capacity needed.   Table 5.7 shows the selected expansion option for each link requiring upgrade.   The net cost of scenario 1 is lower than that in scenario 2; conversely, the flow cost over a 5-year span is higher in scenario 1 than in scenario 2.   This is what we expected as the tradeoff between capital investment and flow cost.   In this case, it is more beneficial to upgrade secondary lines because the total overall cost is lower than the scenario with only mainline upgrades.

Figure 5.10 shows the final capacity usage for each arc of the network.   The dashed lines represent links currently at capacity whereas the dotted lines represent unused links. Hence, the ISM output helps identify not only the important links, but also the

unimportant ones.   Consequently, capacity planners could also use these results to

consider downgrading unimportant links to reduce cost and shift resources to links that

provide higher benefit.

Table 5.5:   Capacity improvement options for link (1,3)

Sidings #	Signals #	Capacity Increase trains/day	Cost $
6	0	0	0
6	7	2	700,000
6	14	3	1,400,000
7	0	2	5,470,000
7	8	5	6,270,000
7	16	6	7,070,000
8	0	5	10,940,000
8	9	7	11,840,000
8	18	8	12,740,000
9	0	7	16,410,000
9	10	9	17,410,000
9	20	11	18,410,000
10	0	9	21,880,000
10	11	12	22,980,000
10	22	13	24,080,000
11	0	12	27,350,000
11	12	15	28,550,000
11	24	16	29,750,000
12	0	16	32,820,000
12	13	18	34,120,000
12	26	20	35,420,000

Table 5.6: The estimated future demand of link (i, j)

i	j	Demand (trains/day)
1	9	8
3	6	6
3	11	9
3	15	8
4	6	8
4	13	7
5	15	9
6	3	6
9	13	5
10	3	2
13	3	8
13	9	5
15	6	8
15	3	5

Table 5.7: Optimal capacity expansion option for each link (i, j)

(a) without (b) with secondary lines

(a)

i	j	Alternatives	Cost ($millions)
3	7	Add signals (3)	1.60
3	5	Add 2 sidings & signals (2)	13.74
8	9	Add 3 sidings	16.41
9	11	Add 3 sidings & signals (1)	17.41
11	12	Add 3 sidings & signals (1)	17.01
5	6	Add 5 sidings & signals (2)	29.35
7	13	Add 5 sidings & signals (2)	32.35
5	8	Add 8 sidings & signals (1)	45.26

(b)

i	j	Alternatives	Cost ($millions)
3	11	Upgrade secondary line	300.00
6	9	Upgrade secondary line	200.00
11	13	Upgrade secondary line	100.00
8	9	Add signals (1)	0.30
11	12	Add signals (1)	0.30
7	13	Add signals (2)	4.00
3	5	Add 1 sidings & signals (2)	8.07
5	8	Add 1 sidings & signals (3)	7.07
5	6	Add 2 sidings & signals (1)	11.64
9	11	Add 2 sidings & signals (2)	11.84
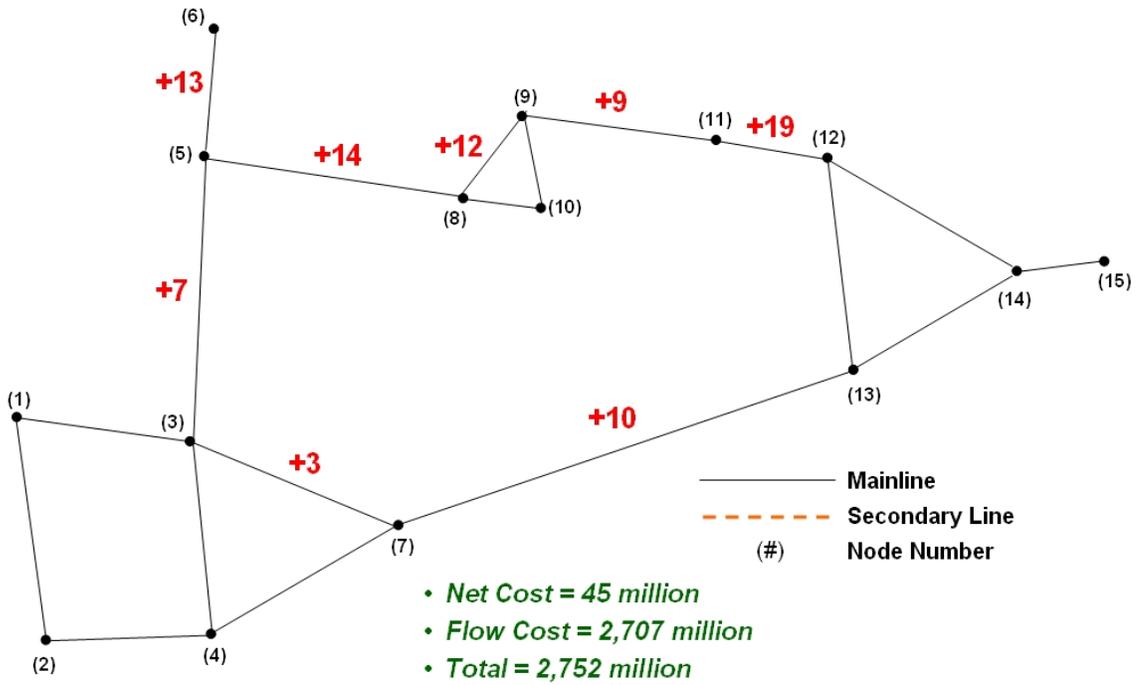
(a)
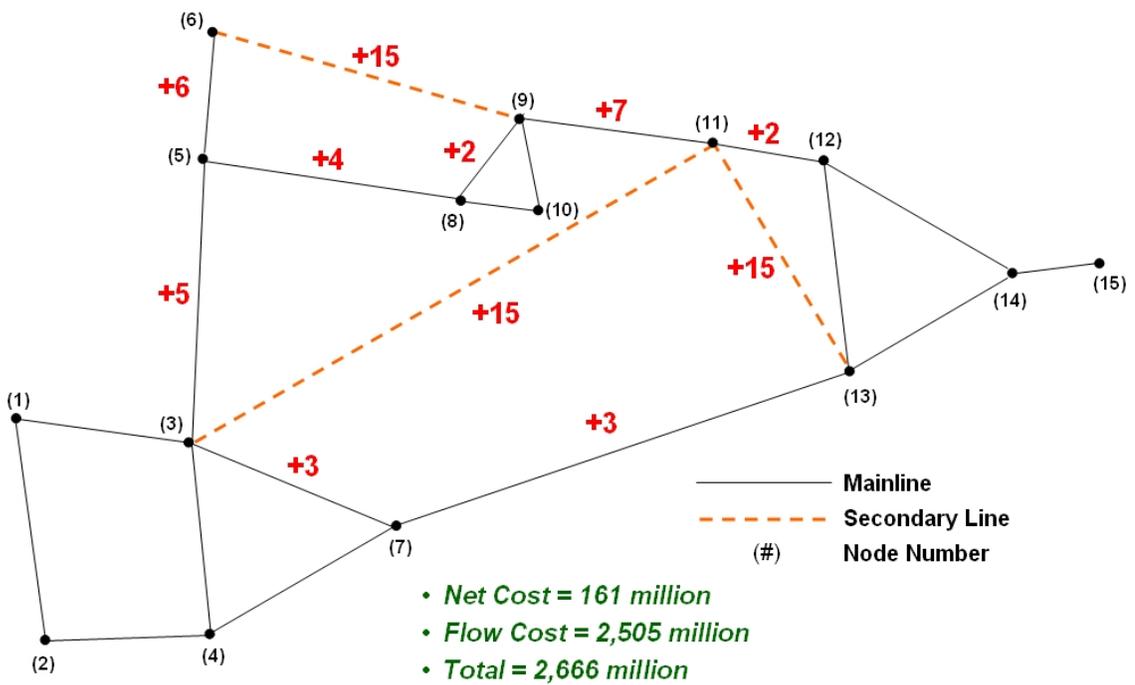
(b)

Figure 5.9: Optimal solution of case study I (a) without (b) with secondary lines
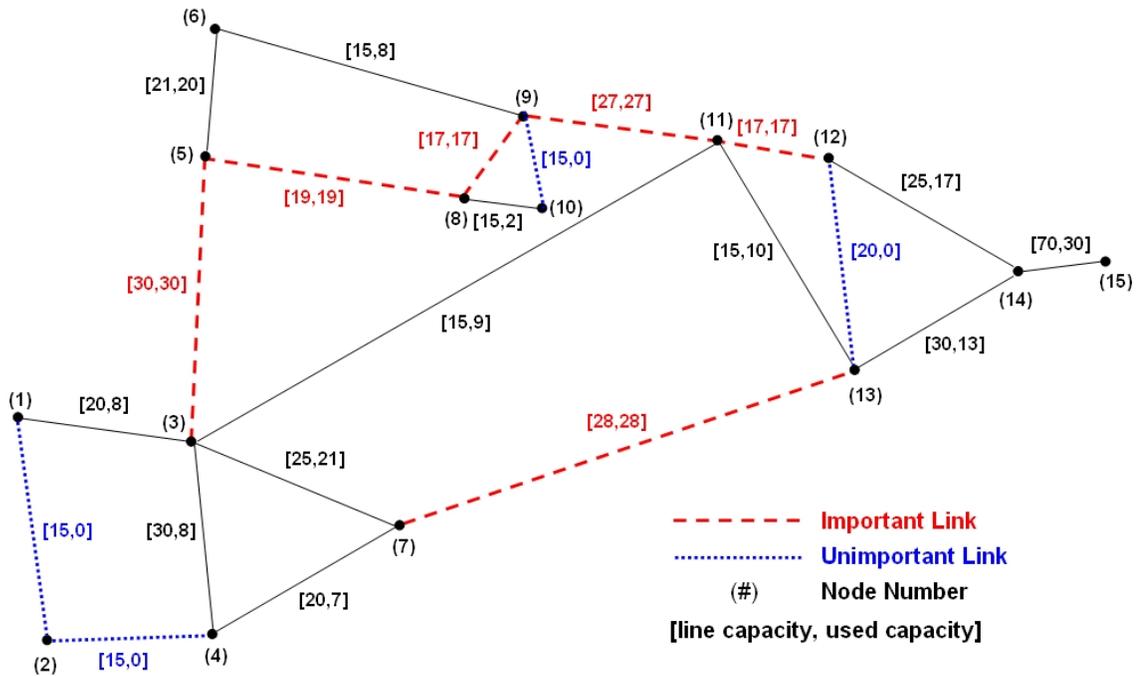
Figure 5.10:   The capacity usage of each link

5.6.2   Case Study II

In case study II, the decision support framework is implemented to establish the multiyear capacity expansion plan based on a real Class 1 railroad network and base train schedule (Figure 5.11).   In this problem, there are 39 nodes, 42 links, and over 1,000 train OD pairs.   It is intended to show the feasibility of using the proposed framework to solve large scale problems.   Also, I use the impact analysis module to show the tradeoff between capital investment and train delay cost.
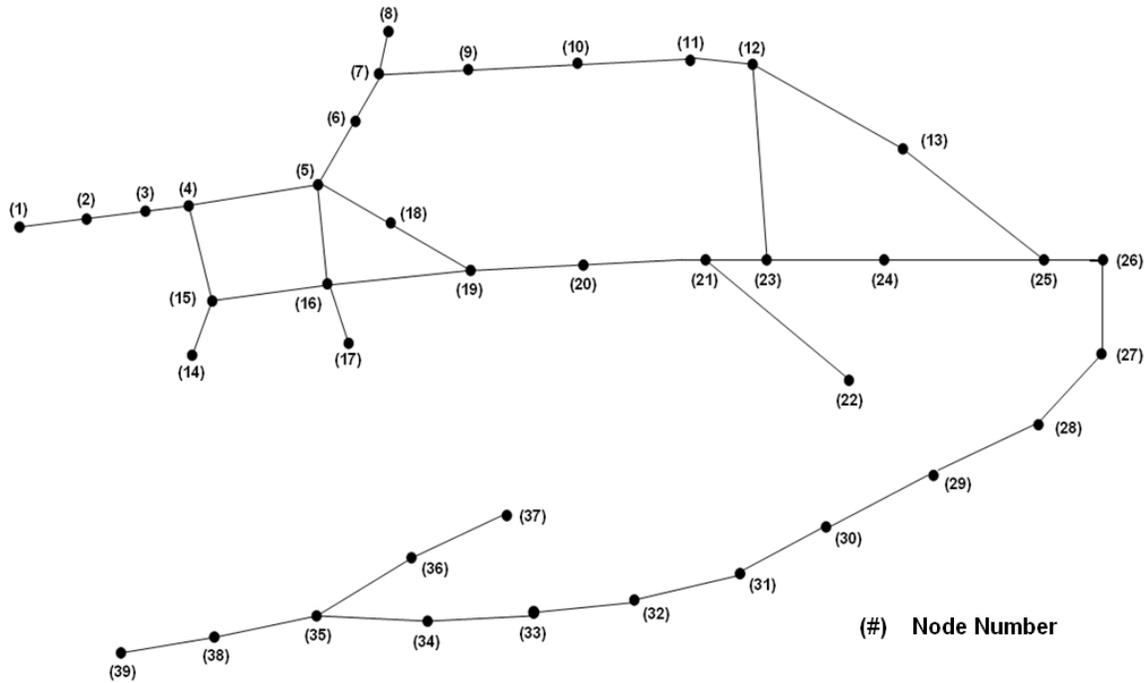
Figure 5.11:   Case study II network

According to the "base train schedule" obtained from the railroad, trains are scheduled to run on a certain day (or days) each week.   Therefore, the base train schedule can be converted into a weekly traffic flow pattern in which there are seven individual daily demand patterns (for each day of the week).   This results in over 1,000 train OD pairs within a week.

The estimated future demand in this case study was assumed to be 50% more than the current traffic flow, and three types of expansion strategies are considered in this application:

- Add sidings: one, two, three, etc. (until the spacing is reduced to 8-miles).

147

- Add intermediate signals: none, one, two, three per spacing (at most three signals per spacing on average).

- Add a 2nd main track

As in case study I, I used the AG to evaluate the current line capacity based on the current network characteristics, and generate possible capacity expansion options from the strategies listed above.   Note that the general enumeration scheme presented in section 5.3.2 is more suitable for a single track mainline with Centralized Traffic Control (CTC) or Automatic Block Signal (ABS) control, as opposed to lines with Track Warrant Control (TWC) (sometimes referred as "dark territory").   Routes with CTC and ABS signal systems have existing track circuit and power infrastructure, so adding signals on these routes would not be as costly as adding signals on TWC trackage which require installation of infrastructure.   Consequently, in this example, I omitted the option of signalizing the TWC routes since they have light traffic and the future demand is also relatively low.

Based on available options and future demand, the ISM gives the optimal solution shown in Figure 5.12.   In this problem, the model includes 89,712 variables and 41,015 equations; and the solution time was only 3.5 seconds.   Although this is just one instance and more empirical evidence is needed to generalize this finding, it shows the feasibility and computational convenience of using the decision support framework to solve large scale network problems.
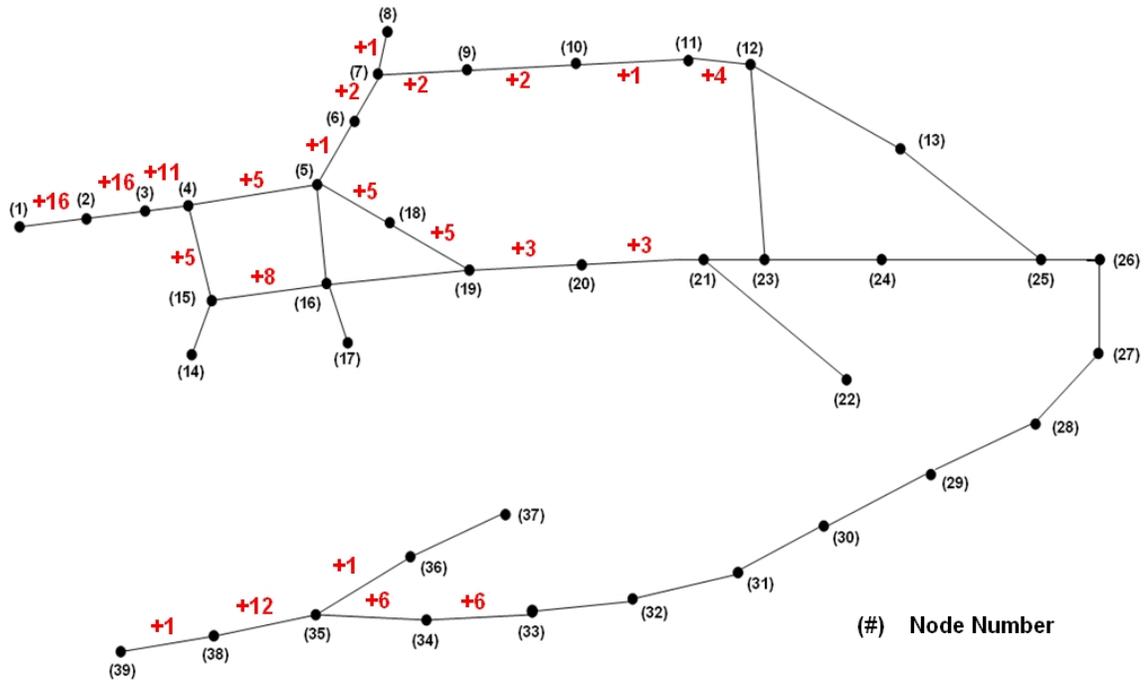
Figure 5.12:   Optimal solution of case study II

As mentioned before, the ISM determines the required upgrade with the premise "LOS is unchanged", but it is possible to gain capacity by increasing delay (reduce LOS) of the subdivision.   This is particularly important for routes that require only a small amount of additional capacity.   For example, from table 5.8, link (35, 36) requires only one train per day additional capacity for Tuesdays and Thursdays; instead of investing millions of dollars to upgrade the infrastructure, it may be more beneficial to reduce the LOS of this link to incorporate additional trains on those two days.   On the contrary, link (1, 2) requires at least 15 trains per day additional capacity; therefore, the return on investment in infrastructure of this link is more likely to be justifiable.   Consequently, the tradeoff between capital investment and delay cost should be taken into account for the final decision.

Table 5.8:   The required additional capacity for each link and day of the week

i	j	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	16	15	16	16	16	16	16
2	3	14	15	15	16	15	16	13
3	4	9	10	10	11	10	11	8
4	5	4	4	5	4	5	4	5
5	6	1	0	0	0	0	0	1
6	7	1	2	2	1	2	0	2
7	8	1	0	1	1	1	0	0
7	9	1	2	0	1	0	1	1
9	10	1	2	0	1	0	1	1
10	11	0	1	0	1	0	1	1
11	12	2	0	4	0	4	0	3
4	15	3	5	5	5	5	5	4
15	16	7	7	6	7	6	7	8
5	18	4	5	5	5	5	5	5
18	19	4	5	5	5	5	5	5
19	20	1	3	3	3	3	3	3
20	21	1	3	3	3	3	3	3
33	34	6	4	5	5	6	5	6
34	35	5	4	2	6	5	6	6
35	36	0	0	1	0	1	0	0
35	38	2	7	9	11	12	12	10
38	39	0	0	0	1	0	1	0

The impact analysis module determines if the capital investment is cost-effective by comparing the investment to delay cost.   According to the required additional capacity of each link (output of ISM), we can use the delay-volume relationship based on the current link properties to compute the increase in delay for each link due to additional traffic without upgrades (Table 5.9).

Table 5.9:   Additional delay due to the increase in future traffic demand

Link		Delay (train-hours)						
i	j	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	2	280.50	250.00	280.50	280.50	280.50	280.50	280.50
2	3	235.20	290.00	290.00	321.30	290.00	321.30	206.40
3	4	245.00	300.00	300.00	331.50	300.00	331.50	192.00
4	5	15.50	15.50	19.20	15.50	19.20	15.50	19.20
5	6	1.60	0.00	0.00	0.00	0.00	0.00	1.60
6	7	1.60	2.55	2.55	1.60	2.55	0.00	1.70
7	8	0.80	0.00	0.80	0.80	0.80	0.00	0.00
7	9	0.30	1.05	0.00	0.30	0.00	0.30	0.30
9	10	0.60	1.40	0.00	0.60	0.00	0.60	0.60
10	11	0.00	0.35	0.00	0.35	0.00	0.35	0.35
11	12	0.80	0.00	3.00	0.00	3.00	0.00	1.80
4	15	4.75	9.45	9.45	9.45	9.45	9.45	7.00
15	16	16.80	16.80	14.00	16.80	14.00	16.80	22.00
5	18	34.20	39.00	39.00	39.00	39.00	39.00	39.00
18	19	22.80	27.30	27.30	27.30	27.30	27.30	27.30
19	20	3.70	11.70	11.70	11.70	11.70	11.70	11.70
20	21	3.70	11.70	11.70	11.70	11.70	11.70	11.70
33	34	15.60	9.60	12.50	12.50	15.60	12.50	15.60
34	35	12.50	9.60	4.40	15.60	12.50	15.60	15.60
35	36	0.00	0.00	1.65	0.00	1.65	0.00	0.00
35	38	5.20	40.30	59.40	80.50	100.80	100.80	71.40
38	39	0.00	0.00	0.00	2.40	0.00	2.40	0.00

The total delay cost of a subdivision is then computed as the product of total delay

hours and unit delay cost per hour ($261 per train-hour).   Table 5.10 shows both the train

delay cost and net cost for each link.   The links are ranked by their benefit, which is the

result of delay cost divided by capital investment.   The first 12 links have benefit value

greater than 1 meaning the return is greater than the investment; however, the other 10 links

have negative return on investment.   This table would be provided to the capacity planner,

as an aid to their final decision making based on the available budget.

Table 5.10:   The benefit of upgrading track

| Link | | Capacity | | Cost ($,k) | | Difference ($,k) | Benefit | Cumulative |
i	j	Current	Max	Train Delay	Net Cost	Delay - Net Cost		Benefit
35	38	24	36	31,107	2,289	28,818	13.59	14
5	18	34	39	18,200	1,643	16,558	11.08	25
3	4	40	51	135,720	13,169	122,551	10.31	35
18	19	34	39	12,663	2,735	9,928	4.63	40
20	21	36	39	5,015	1,393	3,622	3.60	43
15	16	14	22	7,953	2,435	5,518	3.27	46
2	3	35	51	132,612	42,188	90,425	3.14	50
19	20	36	39	5,015	1,693	3,322	2.96	53
4	5	27	32	8,116	4,278	3,839	1.90	54
1	2	35	51	131,173	84,813	46,361	1.55	56
33	34	20	26	6,372	4,870	1,502	1.31	57
34	35	20	26	5,822	4,870	952	1.20	59
4	15	16	21	4,004	4,870	(866)	0.82	59
6	7	15	17	852	1,218	(366)	0.70	60
38	39	23	24	326	1,218	(892)	0.27	60
11	12	6	10	584	2,435	(1,851)	0.24	61
35	36	32	33	224	1,218	(994)	0.18	61
5	6	15	16	217	1,218	(1,000)	0.18	61
7	8	15	16	217	1,218	(1,000)	0.18	61
9	10	5	7	258	2,435	(2,177)	0.11	61
10	11	6	7	95	1,218	(1,122)	0.08	61
7	9	5	7	153	2,435	(2,282)	0.06	61
Sum				506,697	185,852	320,845		

With limited budget (B), the optimal investment set can be determined by solving the

knapsack formulation presented in section 5.5:

$$\text{min total delay cost} - \sum_i \sum_j d_{ij} y_{ij} = 506,697 - \sum_i \sum_j d_{ij} y_{ij}$$

s.t.

$$\sum_i \sum_j h_{ij} y_{ij} \leq B$$

Where d_{ij} is the delay cost of arc (i, j) in the fifth column in table 5.10; h_{ij} is the

construction costs to upgrade arc (i, j); and y_{ij} is the binary decision variable determining

whether arc (i, j) is upgraded ($y_{ij} =1$) or not ($y_{ij} =0$).   For example, if the available budget

is 70 million dollars, the optimal investment set will be to upgrade links (35, 38), (5, 18),

and (3, 4) because they can provide the most reduction in total delay within the budget constraint.

5.7 Conclusion and Future Work

Many railroad lines are approaching the limits of practical capacity given their current infrastructure.   I have developed a decision support framework to help capacity planners determine how to optimally allocate funds for railway capacity expansion projects at the network level.   The framework is comprised of three components: (1) an "Alternatives Generator (AG)" that enumerates possible expansion options along with their cost and capacity effects; (2) an "Investment Selection Model (ISM)" that determines which portions of the network (at the subdivision level) need to be upgraded with what kind of capacity improvement alternatives; and (3) an "Impact Analysis Module (IAM)" that evaluates the tradeoff between capital investment and delay cost.   These components can be used separately as stand-alone tools, or they can be combined as a consolidated decision-support framework.

Based on network characteristics, estimated future demand, and available budget, the proposed decision support framework can successfully determine the optimal solution regarding which subdivisions need to be upgraded and what kind of engineering options should be conducted.   Such a tool will help railroads maximize their return from capacity expansion projects and thus be better able to provide reliable service to their customers, and return on shareholder investment.   Such a decision support framework is highly

beneficial for optimizing the budget planning for expanding the capacity of North American railroads.

The investment selection model I have developed here is a deterministic, one-time investment model that does not account for stochastic future demand, multi-period decision making, or global optimization by railroads. The optimal investment plan may be different if funding is constrained for each railroad considered in each year of the planning horizon. Therefore, an interesting extension may be to use a dynamic optimization model to identify the optimal sequence of upgrades for all the railroads considered in the capacity expansion projects. Also, because the demands of all commodities are assumed to be fixed, another interesting extension would be to incorporate seasonal demand patterns into the model, and again try to determine what the best set of investment options is. A multi-period stochastic investment selection model such as this could help capacity planners determine how to optimally allocate the budget in different decision time(s) for capacity expansion.

CHAPTER 6

SUMMARY AND FUTURE RESEARCH

The focus of my dissertation is to increase railway efficiency and capacity through improved operations, control and planning. Various analytical approaches were developed and carried out using operations research techniques, and capacity analysis methodologies for optimization of the aerodynamic efficiency of intermodal freight trains (chapter 2, 3, & 4), and railway capacity planning (chapter 5).

6.1 Summary

In chapter 2, I investigate and evaluate possible options to improve the energy efficiency of intermodal trains. I found that maximizing slot efficiency by matching intermodal loads with cars (of an appropriate length) reduces the gap length between loads, thereby improving airflow and reducing drag, which in turn reduces fuel consumption and operating cost. Filling empty slots with empty containers or trailers also reduces aerodynamic resistance, further improving energy efficiency. The analytical results show that train resistance can be reduced by as much as 27% and fuel savings by 1 gal/mile per train.

I then develop the Aerodynamic Loading Assignment Model (ALAM), an integer programming (IP) model that incorporates aerodynamic characteristics of loads and railcar

combinations to enable optimization of loading patterns to maximize fuel efficiency

(chapter 3). ALAM can be used to help terminal managers make up more fuel-efficient

trains. This is the first use of optimization modeling with the objective of improving the

aerodynamics and consequent energy efficiency of intermodal freight trains. The model

developed here can be adapted to a variety of other intermodal train loading assignment

problems through modification of the objective function. Finally, several policy

recommendations regarding railway intermodal operations are developed based on a series

of scenario analyses. The potential annual savings in fuel consumption through use of

ALAM by one large railroad on one of its major intermodal routes is estimated to be

approximately 15 million gallons with a corresponding value in 2007 of 29 million dollars.

ALAM was further improved to allow optimization of multiple trains simultaneously

if advance information about outgoing trains and loads is available. Chapter 4 presents

static and dynamic aerodynamic efficiency models for the loading of multiple IM trains.

It also develops a rolling horizon scheme for continuous train terminal operations. For the

static case, when full information is available, the system optimum can be reached by

optimizing as many trains as possible. In practice, however, terminals operate in a

dynamic environment where not all information on incoming loads and trains is available.

Attempting to optimize the loading of too many trains in this environment will reduce the

ability to achieve the most efficient loading configuration. Therefore, a rolling horizon

scheme with decreasing weight assigned to each train is proposed to counterbalance the

effect of uncertainty. Numerical results show that the rolling horizon scheme

significantly reduces the adjusted gap length compared to current practice, thereby

leading to further improvement in the aerodynamic efficiency of IM trains. Correspondingly greater savings in fuel, emissions and expense are possible if this methodology is applied to all North American IM trains.

In chapter 5, I demonstrate a new decision support framework to help railroad capacity planners determine how to optimize the allocation of capital investment for capacity expansion projects. This framework has three stand-alone tools: (1) an "Alternatives Generator" that enumerates possible expansion options along with their cost and capacity effects; (2) an "Investment Selection Model" that determines which portions of the network (at the subdivision level) need to be upgraded with what kind of capacity improvement options; and (3) an "Impact Analysis Module" that evaluates the tradeoff between capital investment and delay cost. Based on network characteristics, estimated future demand, and available budget, the proposed decision support framework can successfully determine the optimal solution regarding which subdivisions need to be upgraded and what kind of engineering options should be conducted. This will help railroads maximize their return from capacity expansion projects and thus be better able to provide reliable service to their customers, and return on shareholder investment. Such a decision support framework can be used to optimize the efficiency and effectiveness of railroad capacity expansion programs.

6.2 Future Research

In this section I discuss some of the possible future research needs and directions in the area of intermodal train efficiency and railroad capacity planning.

Improving the Design of Intermodal Equipments to Improve Efficiency

In Chapter 2, I pointed out that intermodal trains are the least fuel efficient trains due to the physical constraints imposed by the combination of loads and the railcar design (Engdahl et al., 1986).   Research on practical methods to improve the aerodynamic design of intermodal equipment and/or railcars to improve their energy efficiency would be worthwhile.   This research would focus on ways to reduce the aerodynamic turbulence and consequent drag caused by the gaps between intermodal loads since these contribute the majority of resistance at high speed.   In addition to the aerodynamic design analyses, the research should include a thorough analysis of fuel savings, additional investment, and effects on operating costs to determine the cost effectiveness of whatever designs are evaluated.

A Multi-Period Stochastic Investment Selection Model for Railway Capacity Expansion

The investment selection model I have developed in the decision support framework is a deterministic, one-time investment model that does not account for stochastic future demand and multi-period decision making.   I plan to develop a multi-period stochastic investment selection model by using approximate dynamic programming (ADP) (Powell, 2007).   ADP is a relatively new solution strategy for large-scale resource allocation problems that take place over multiple time periods under uncertainty.   The multi-period stochastic investment selection model can help capacity planners determine how to optimally allocate the budget in different decision time(s) for capacity expansion.

Impact of Advanced Railway Technologies On Line and Network Capacity

New railway operating technologies (e.g. PTC, ECP brakes, switch position indication, intelligent bearings, etc.) have been developed to improve safety, reliability and productivity.   Two of the most promising are PTC and ECP brakes, which have both shown a variety of potential benefits.   However, there remain practical, economical and institutional barriers to their implementation, in part because of insufficient understanding of the benefits they offer railroads in terms of enhanced capacity.   The impact on network capacity from these new technologies is not clearly understood and quantified.   Hence, there is a need to identify advanced operating technologies with the potential to positively affect rail line and network capacity, and then develop analytical tools and conduct analyses to quantify the capacity benefits of various possible scenarios.   This will provide both public and private rail agencies with a method to evaluate and quantify the potential benefits and help them determine the most cost effective plan for adopting these new technologies to meet future demand.

Improving Line Capacity through Mitigating the Heterogeneity in Train Operations

The greater the heterogeneity in train speeds on a line, the more potential conflicts, and the lower the line capacity.   Railway capacity can generally be increased by two options: (1) reducing the heterogeneity among different types of trains, and (2) upgrading railway infrastructure.   The former increases capacity by adjusting the horse power to ton (HPT) ratio to reduce the heterogeneity in train speeds and the number of conflicts.   The latter improves physical capacity but usually requires long lead time and substantial capital investment.   Furthermore, because such investments are permanent, it is less flexible and

fungible than adjusting the HPT so the risk is greater. There is a need to understand the relative benefits and interaction between operational and engineering options for capacity expansion. The objective of this study is to compare the impact on capacity through adjusting the HPT or upgrading infrastructure, and then develop criteria to better inform capacity planning decisions.

Development of a Parametric Capacity Model for Multiple Track Scenario

As shown in chapter 5, the CN parametric model recognizes the dynamic nature of capacity and provides a system-wide capacity measure of subdivisions in a rail network. However, the current version is designed for a single track network that does not take into account multiple track scenarios (e.g. crossovers), and/or other different operational practices (e.g. directional running). In addition to identifying areas of limited or excess capacity, capacity tools serve as the baseline evaluation instrument for many other complicated optimization models, such as the decision support framework presented in chapter 5, and railway scheduling optimization tools for solving train, crew, and locomotive scheduling problems. The better the user can assign the right capacity value, the better the optimal plan can be created from those tools. Besides private sector railroads, this capacity evaluation tool will also be useful to public agencies to help them set national transportation priorities and investment plans. Therefore, there is a substantial need to develop a standard, comprehensive railway parametric capacity model analogous to the Highway Capacity Manual (TRB, 2000). Such a model could assist public and private financing of rail capacity investments by determining the magnitude, cost, and type of capacity improvements needed for the desired service(s).

REFERENCES

Abril, M., Barber F., Ingolotti L., Salido M.A., Tormos P., and Lova A. (2007), An Assessment of Railway Capacity, *Transportation Research Part E*, doi:10.1016/j.tre.2007.04.001

Acharya, D., Mishalani, R., Martland, C., and Esheby, J. (1991)   REPOMAN: An Overall Computer-Aided Decision Support System for Planning Rail Replacements, *International Heavy Haul Railway Conference*, Vancouver, BC Canada, 113–123.

Ahuja, R.K., Magnanti, T.L., and Orlin, J.B. (1993)   *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, New Jersey.

Air Brake Association (ABA) (1972)   Management of Train Operation and Train Handling, *The Air Brake Association*, Chicago, IL.

Airflow Science Corporation (2006)   *Aerodynamic Subroutine Version 4.0*, Association of American Railroads, Washington, DC.

Alle, P. (1981)   Improving Rail Transit Line Capacity Using Computer Graphics, *Logistics and Transportation Review*, Volume 17, Number 4, University of British Columbia, Faculty of Commerce.

American Association of State Highway and Transportation Officials (AASHTO) (2005) *Return on Investment on Freight Rail Capacity Improvement*, National Cooperative Highway Research Program, NCHRP 8-36, Washington, DC.

American Association of State Highway and Transportation Officials (AASHTO) (2007) *Transportation - Invest in Our Future: America's Freight Challenge*, AASHTO, Washington, DC.

American Railway Engineering and Maintenance-of-Way Association (AREMA) (2001) Train Performance, *Manual for Railway Engineering*, Chapter 16, Part 2, AREMA, Landover, MD.

American Railway Engineering Association (AREA) (1947) *Train Capacity and Train Performance*, AREA, Bulletin 452, 125-144.

Anderson, F. (2006) Personal Communication, BNSF Railway Company, Fort Worth, TX.

Armstrong, J. (1998) *The Railroad: What It Is, What It Does.*, Fourth Ed., Simmons-Boardman Books, Inc., Omaha, NE.

Assad, A. (1980) Models for Rail Transportation. *Transportation Research Part A* , Vol 14, 205–220.

Association of American Railroads (AAR) (2005) *UMLER Data Specification Manual*, Railinc Corporation, Cary, NC.

Association of American Railroads (AAR) (1981) *Proceedings Railroad Energy Technology: The Alternatives,* Railway Systems & Management Association, North Field, NJ.

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

Association of American Railroads (AAR) (2004-2007) *Railroad Facts*, AAR, Washington, DC.

Association of American Railroads (AAR) (2007) *National Rail Freight Infrastructure Capacity and Investment Study*, AAR, Washington, DC.

Avriel, M., Penn, M., Shpirer, N., and Witteboon, S. (1998) Stowage Planning for Container Ships to Reduce the Number of Shifts, *Annals of Operations Research*, Vol 76, 55-71.

Balakrishnan A., Magnanti T., and Wong R. (1995)   A Decomposititon Algorithm for Local Access Telecommunications Network Expansion Planning, *Operations Research*, Vol 43(1), 58–76.

BNSF Railway Company (2004)   Every Drop Counts in Goal to Improve Fuel Efficiency, *BNSF News*, BNSF Railway Company, Fort Worth, TX.

BNSF Railway Company (2004)   *Intermodal Loading Guide*, BNSF Railway Company, Fort Worth, TX.

BNSF Railway Company (2005)   *Train Documents*, BNSF Railway Company, Fort Worth, TX..

Barkan, C.P.L. (2004)   Cost Effectiveness of Railroad Fuel Spill Prevention Using a New Locomotive Refueling System, *Transportation Research Part D,* Vol 9, 251-262.

Bronzini, M.S., and Clarke, D.B. (1985)   Estimating Rail Line Capacity and Delay by Computer Simulation, *Transportation Forum,* Vol 2, 5-11.

Brooke, A., Kendrick, D., Meeraus, A., and Raman, R. (1998)   *GAMS – A User's Guide*, GAMS Development Corporation, Washington, DC.

Burriss, C. (2003)   Eastbound Statistical Analysis: Slot Utilization & Mixed Cars, *OASIS Users' Conference*, Austin, TX.

Cai, X., and Goh, C. (1994)   A Fast Heuristic for the Train Scheduling Problem. *Computers and Operations Research*, Vol 21, 499–510.

Canadian National Railway (CN) (2005)   *BCNL Long Siding Requirements*, CN, Edmonton, AB.

Canadian Urban Transit Association (1985)   *Capacity, Canadian Transit Handbook*, Second Edition, Canadian Urban Transit Association, and the Roads and Transportation Association of Canada, Toronto, ON.

Caprara, A., Fischetti, M., and Toth, P. (2002)　Modeling and Solving the Train Timetabling Problem, *Operations Research*, Vol 50(5), 851-861.

Carey, M., and Lockwood, D. (1995). A Model, Algorithms and Strategy for Train Pathing, *Journal of the Operational Research Society*, Vol 46, 988–1005.

Chen, B., and Harker, P. (1990)　Two Moments Estimation of Delay on Single-Track Rail Lines with Scheduled Traffic, *Transportation Science*, Vol 24, 261–275.

Conant, M. (1964)　*Route Capacity, Excess Capacity, and Overinvestment, Railroad Mergers and Abandoments*, University of California Press, Los Angeles, CA.

Corry, P.G., and Kozan, E. (2006)　An Assignment Model for Dynamic Load Planning of Intermodal Trains, *Computers and Operations Research*, Vol 33, 1–17.

Delorme, X., Rodriguez, J., and Gandibleux, X. (2001)　Heuristics for Railway Infrastructure Saturation, *Theoretical Computer Science*, Vol 50(1), 39-53.

Department of Transport (2001)　*A Strategy for Regeneration of Rail Research in Great Britain*, Department of Transport, London, UK.

Drish, W.F. (1992)　*Train Energy Model Version 2.0 Technical Manual*, SD-485, AAR, Chicago, IL.

Ekman, J. (2004)　Capacity Estimation of New Infrastructure Based on a Discrete Event Model of Train Paths, *Computers in Railways IX*, 539-548.

Engdahl, R. (1987)　*Full-Scale Rail Car Testing to Determine the Effect of Position-in-Train on Aerodynamic Resistance*, SD-705, AAR, Washington, DC.

Engdahl, R., Gielow, R.L., and Paul, J.C. (1987)　Train Resistance – Aerodynamics Volume I of II Intermodal Car Application, *Proceedings of Railroad Energy Technology Conference II*, AAR, Chicago, IL.

Engdahl, R., Gielow, R.L., and Paul, J.C. (1987)   Train Resistance – Aerodynamics Volume II of II Open Top Car Application, *Proceedings of Railroad Energy Technology Conference II*, AAR, Chicago, IL.

Eppen G., Martin R., and Schrage L. (1989)   A Scenario Approach to Capacity Planning, *Operations Research*, Vol 37(4), 517–527.

Felipe, G., Loureiro, C., and Ralston, B. (1996)   Investment Selection Model for Multicommodity Multimodal Transportation Networks, *Transportation Research Record*, Vol 1522, 38-46.

Feo, T.A., and Gonzalez-Velrade, J.L. (1995)   The Intermodal Trailer Assignment Problem, *Transportation Science*, Vol 29, 330–341.

Fransoo, J.C., and Bertranda, J. (2000)   Aggregate Capacity Estimation Model for the Evaluation of Railroad Passing Constructions, *Transportation Research, Part A*, Vol 34(1), 35-49.

Furlong, C.F. (1988)   *Aerodynamic Subroutine Users Guide*, SD-683, AAR, Washington, DC.

Gallamore, R.E. (1998)   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, VA.

Gielow, M.A., and Furlong, C.F. (1988)   *Results of Wind Tunnel and Full-Scale Tests Conducted from 1983 to 1987 in Support of the Association of American Railroads' Train Energy Program*, SD-685, AAR, Washington, DC.

Giemsch, P., and Jellinghaus, A. (2004)   Optimization Models for the Containership Stowage Problem, *Operations Research Proceedings 2003*, Springer, Berlin.

Grimes, G.A., and Barkan, C.P.L. (2006)   Cost-Effectiveness of Railrway Infrastructure Renewal Maintenance, *Journal of Transportation Engineering*, Vol 132(8), 601-608

Harker, P.T., and Hong, S. (1990) Two Moments Estimation of the Delay on a Partially Double-Track Rail Line with Scheduled Traffic, *Journal of the Transportation Research Forum*, Vol 31(1), 38–49.

Harker, P.T., and Hong, S. (1994) Pricing of Track Time in Railroad Operations: an Internal Market Approach, *Transportation Research Part B*, Vol 28(3), 197–212.

Hay W.W. (1977) *An Introduction to Transportation Engineering*, Second Ed., John Wiley & Sons, Inc., New York, NY.

Hay, W.W. (1982) *Railroad Engineering*, Second Ed., John Wiley & Sons, Inc., New York, NY.

HDR (2003) *I-5 Rail Capacity Study*, HDR, Portland, OR.

Higgins, A., Kozan, E., and Ferreira, L. (1996) Optimal Scheduling of Trains on a Single Line Track, *Transportation Research Part B*, Vol 30, 147–161.

Hsu V.N. (2002) Dynamic Capacity Expansion Problem with Deferred Expansion and Age-Dependent Shortage Cost, *Manufacturing & Service Operations Management*, Vol 4(1), 44–54.

Improverail Consortium (2003) *Methods for Capacity and Resources Management,* European Project, Deliverable D6, Improverail Consortium.

International Union of Railways (UIC) (2003) *Energy Efficiency Technologies for Railways* [online]. [Accessed May 20, 2006]. Available from World Wide Web: < http://www.railway-energy.org/tfee/index.php?ID=105>

International Union of Railways (UIC) (2004) *Leaflet 406: Capacity*, UIC, Paris, France

International Union of Railways (UIC) (2004) *Proceedings 2nd UIC Railway Energy Efficiency Conference*, UIC, Paris, France.

Jelaska, M. (1998)   Railway Line Capacity Planning Support Model, *Computers in Railway VI,* 1120-1131.

Jovanovic, D., and Harker, P. (1991)   Tactical Scheduling of Rail Operations: the SCAN I system, *Transportation Science,* Vol 25, 46–64.

Kozan, E., and Burdett, R. (2005)   A Railway Capacity Determination Model and Rail Access Charging Methodologies, *Transportation Planning and Technology*, Vol. 28(1), 27-45.

Kraft, E.R. (1988)   Analytical Models for Rail Line Capacity Analysis, *Journal of Transportation Research Forum*, Vol 29(1), 153–162.

Krueger, H. (1999)   Parametric Modeling in Rail Capacity Planning, *Proceedings of Winter Simulation Conference*, Phoenix, AZ.

Lai, Y.C., and Barkan, C.P.L. (2005)   Options for Improving the Energy Efficiency of Intermodal Freight Trains, *Transportation Research Record*, Vol 1916, 47–55.

Lai Y.C., Barkan CPL, and Önal H. (2007)   Optimizing the Aerodynamic Efficiency of Intermodal Freight Trains, *Transportation Research Part E*, doi:10.1016/j.tre.2007.05.011.

Lai, Y.C., Ahuja, N., Barkan, C.P.L., Drapa, J., Hart, J.M., Jawahar, C.V., Kumar, A., Milhon, L., Narayanan, P.J., and Stehly, M. (2007)   Machine Vision Analysis of the Energy Efficiency of Intermodal Freight Trains, *Journal of Rail and Rapid Transit*, Vol 221, 353–365.

Lai, Y.C., Ouyang, Y., and Barkan, C.P.L. (2008)   A Rolling Horizon Model to Optimize Aerodynamic Efficiency of Intermodal Freight Trains with Uncertainty, *Transportation Science*, Accepted.

Landex, A., Kaas, A.H., Schittenhelm, B., and Schneider-Tilli, J. (2006)   Practical Use of the UIC 406 Capacity Leaflet by Including Timetable Tools in the Investigations, *Computers in Railways X*, 643–652.

Larsen, A. (2001)   *The Dynamic Vehicle Routing Problem*, Ph.D. Thesis, Institute of Mathematical Modeling, Technical University of Denmark, Lyngby.

Lee, D.B., Jr. (2002)   Fundamentals of Life-Cycle Cost Analysis, *Transportation Research Record*, Vol 1812, 203–210.

Levinson, H.S., and Hoey, W.E. (1991)   Some Reflections on Transit Capacity, *Proceedings of the International Symposium on Highway Capacity*, Karlsruhe, Germany.

Ling, D.J., Roy, R., Shehab, E., Jaiswal, J., and Stretch, J. (2006)   Modelling the Cost of Railway Asset Renewal Projects using Pairwise Comparisons, *Journal of Rail and Rapid Transit*, Vol. 220, 331–346.

Lund, K., Madsen, O.B.G., and Rygaard, J.M. (1998)   *Vehicle Routing Problems with Varying Degrees of Dynamism*, Technical Report, Institute of Mathematical Modeling, Technical University of Denmark, Lyngby, Denmark.

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

Martland C.D., and Hutt G. (2005)   *Analysis of Potential Approaches to Interline Capacity Flow Management*, RaiLinc, Cary, NC.

Milhon L.R. (2004)   Personal Communication, BNSF Railway Company, Topeka, KS.

Minoux M. (1989)   Network Synthesis and Optimum Network Design Problems: Models, Solution Methods, and Applications, *Networks*, Vol 19, 313–360.

Muller, G. (1999)   *Intermodal Freight Transportation*, Fourth Ed., Eno Transportation Foundation and Intermodal Association of North America, Washington, DC.

Murphy F., and Weiss. H. (1990)   An Approach to Modeling Electric Utility Capacity Expansion Planning, *Naval Research Logistics*, Vol 37, 827–845.

Oliveira, E., and Smith, B.M. (2000)   *A Job-Shop Scheduling Model for the Single-Track Railway Scheduling Problem*, Research Report 2000.21, University of Leeds, Leeds, UK.

Optimization Alternatives Ltd. Inc. *Optimization Alternatives' Strategic Intermodal Scheduler (OASIS)*. <http://www.oax.com> Accessed 06.01.07.

Ordonez, F., and Zhao, J. (2005)   *Robust Capacity Expansion of Network Flows*, University of Southern California, Los Angeles, CA.

Ozbay, K., Jawad, D., Parker, H., and Hussein, S. (2004)   Life Cycle Cost Analysis: State-of-the-Practice vs. State-of-the-Art, *Transportation Research Record*, Vol 1864, 62-70

Pachl, J. (2002)   *Railway Operation and Control*, VTD Rail Publishing, Mountlake Terrace, WA.

Petersen, E.R. (1974)   Over-the-Road Transit Time for a Single Track Railway, *Transportation Science*, Vol 8(1), 65–74.

Petersen, E.R., and Taylor, A.J. (1982)   A Structured Model for Rail Line Simulation and Optimization, *Transportation Science*, Vol 16, 192–206.

Petersen, E.R., and Taylor, A.J. (2001)   An Investment Planning Model for a New North-Central Railway in Brazil,   *Transportation Research Part A*, Vol 35(9), 847-862.

Poole E. C. (1962)   *Costs – a Tool for Railroad Management*, Simmons-Boardman Publishing Corporation, New York, NY.

Powell, W.B., and Carvalho, T.A. (1998)   Real-time Optimization of Containers and Flatcars for Intermodal Operations, *Transportation Science*, Vol 32, 110–126.

Powell W.B. (2007)   *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, Wiley-Interscience, New York, NY.

Prokopy, J. C., and Rubin, R.B. (1975) *Parametric Analysis of Railway Line Capacity*, DOT-FR-5014-2, Federal Railroad Association, U.S. Department of Transportation, Washington, DC.

Putallaz, Y., and Rivier, R. (2004) Strategic Evolution of Railway Corridor Infrastructure: Dual Approach for Assessing Capacity Investments and M and R Strategies, *Computers in Railways IX*, 61-72.

Riis M., and Andersen K.A. (2004) Multiperiod Capacity Expansion of a Telecommunications Connection with Uncertain Demand, *Computers & Operations Research*, Vol 31(9), 1427–1436.

Smith, M.E. (1987) Economics of Reducing Train Resistance, *Proceedings of Railroad Energy Technology Conference II*, AAR, Chicago, IL.

Smith, R.A. (2003) How they may contribute to a sustainable future? *Journal of Rail and Rapid Transit*, Vol 217(4), 243-248.

Stodolski, F. (2002) *Railroad and Locomotive Technology Roadmap*, ANL/ESD/02-6. Argonne National Laboratory, Argonne, IL.
<http://www.climatevision.gov/sectors/railroads/pdfs/roadmap.pdf > Accessed 08.26.07.

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

Szpigel, B. (1972) Optimal Train Scheduling on a Single Track Railway, *Operations Research*, Vol 72, 343–352.

Thompson, W. (2006) CREATE Update, *Proceedings of AREMA 2006 Annual Conference*, AREMA, Louivvile, KY

Transit Cooperative Research Program (TCRP) (1996) *Rail Transit Capacity*, TCRP Report 13, TRB, Washington, DC.

Transportation Research Board (TRB) (2000)   *Highway Capacity Manual 2000*, TRB, Washington, DC.

Transportation Research Board (TRB) (2006)   Research to Enhance Rail Network Performance: A Workshop [online].   [Accessed May 20, 2006].   Available from World Wide Web: <http://trb.org/conferences/railworkshop/>

TTX Company (1999)   *Equipment Guide*, TTX Company, Chicago, IL

Union Pacific Railroad Corporation (UP) (2004)   *Intermodal Loading Guid*e, <http://www.uprr.com/customers/dam-prev/attachments/intgenloadguide.pdf> Accessed 06.01.07.

Utterback, M. (2006)   Personal Communication. BNSF Railway Company, Chicago, IL.

Vantuono, W. C. (2005)   Capacity is Where You Find It: How BNSF Balances Infrastructure and Operations, *Railway Age*, Feb-01-2005.

Viet, P. (2002)   Track Maintenance based on Life-Cycle Cost Calculations, *Innovations for a Cost Effective Railway Track*, 6–13.

Vuchic, V. (1981)   *Urban Public Transportation*, Prentice-Hall, New Jersey.

Wahlborg, M. (2004)   Banverket Experience of Capacity Calculations according to the UIC Capacity Leaflet, Advances in Transport, v 15, Computers in Railways IX, 665-673.

Wellington, A.M. (1914)   *The Economic Theory of the Location of Railways*, Sixth Ed., Wiley, New York, NY.

Wierderkehr, P. (2004)   Environmentally Sustainable Transport (EST) - Driving the Economic Development? *Proceedings 2nd UIC Railway Energy Efficiency Conference*, International Union of Railways (UIC), Paris, France.

Wilson, E. (2008)   *Rail Traffic Controller (RTC) Brochure,* Berkeley Simulation Software, Berkeley, CA.

Zoeteman, A., and Esveld, C. (1999)   Evaluating Track Structures: Life Cycle Cost Analysis as a Structured Approach, *Proceedings of World Congress on Railway Research*, Tokyo, Japan.

AUTHOR'S BIOGRAPHY

Yung-Cheng Lai was born in Taipei, Taiwan, on December 28th, 1979.   He received his B.S. in Civil Engineering at National Taiwan University in 2002 and M.S. in Civil and Environmental Engineering at University of Illinois in 2004.   His main research interests include railroad traffic operations and control, application of mathematical programming, and energy efficiency in transportation systems.   His dissertation research was being sponsored by the BNSF Railway and he was awarded a CN Research Fellowship in Railway Engineering in 2004.