CAPACITY EVALUATION AND INFRASTRUCTURE PLANNING TECHNIQUES FOR HETEROGENEOUS RAILWAY TRAFFIC UNDER STRUCTURED, MIXED, AND FLEXIBLE OPERATION

BY

MEI-CHENG SHIH

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, 2017

Urbana, Illinois

Doctoral Committee:

Professor Christopher P. L. Barkan, Chair
Professor Yanfeng Ouyang
Associate Professor Yung-Cheng Lai, National Taiwan University
Assistant Professor Daniel B. Work
C. Tyler Dick, P.E.

ABSTRACT

North American railroads have a strong business incentive to match rail line capacity to traffic demand. Since insufficient capacity reduces level of service and excess capacity represents inefficient use of capital, either one of these situations is undesirable. Various processes, models, and tools have been developed to assist the railroads in determining appropriate infrastructure projects and operational plans to balance network capacity. In North America, these approaches have typically been tailored to operating conditions on rail corridors that are dominated by freight trains that do not run according to a precise schedule. Changes in the composition of rail traffic have resulted in new operating conditions that require new approaches to rail capacity evaluation.

The long-term growth of freight rail traffic (with particular increases in premium intermodal traffic) and recent interest in the expansion of passenger service on freight corridors have increased rail traffic volume and heterogeneity, while altering the level of randomness involved in train departure and trip times. The single-track lines that comprise the majority of the North American rail network have limited capacity and can frequently become congested under these new rail traffic demands. The combined impact of traffic volume, heterogeneity, and level of randomness in train plans has not always been fully considered by previous approaches to the study of rail line capacity. This dissertation develops new capacity evaluation and infrastructure planning techniques for single-track lines that consider the impact of relationships between infrastructure layout, train operating plans including train-specific levels of service, and train characteristics on line capacity.

In this study, the randomness involved in a train operating plan is described by "schedule flexibility" and "operating style". In chapter 1, the concepts of operating style and schedule flexibility are proposed and defined. In chapters 2 and 3, a capacity evaluation and alternative comparison process are proposed to assist the capacity evaluation and planning of single-track lines under mixed or flexible operation. In chapter 4, an optimization model is developed to determine the optimal number and locations of passing sidings for single-track lines under structured operation. In chapter 5, the concept of traffic conflict analysis is introduced as a research direction to address rail infrastructure and operational planning problems.

The methods developed in this dissertation can help to better assess mainline capacity under current operating conditions and determine more effective infrastructure expansion projects or changes in operational strategy for railroads and passenger rail agencies in North America. Use of these methods can help railroads improve their service quality and maximize returns to their stakeholders.

ACKNOWLEDGEMENTS

This dissertation is dedicated to my family, including my parents, Yew-Sheng Shih, Jin-Shiang Wu, and my brother Mei-Ji Shih. They have given me great support during my graduate study.

I would like to express my most sincere gratitude to my advisors, Dr. Christopher P.L. Barkan and C. Tyler Dick, P.E. for the support and encouragement they provided. I am grateful for the opportunity to work under them.

I would also like to thank my Ph.D. committee members, Dr. Christopher P.L. Barkan, Dr. Yanfeng Ouyang, Dr. Yung-Cheng Lai, Dr. Daniel B. Work, and C.Tyler Dick, P.E., for their reviews and comments on my research.

I wish to thank all my friends, colleagues, and classmates who I met during my graduate study life. Without you I cannot complete my study and become the person I am today.

During my graduate study, my research work was supported by grants from the Association of American Railroads, the National University Rail (NURail) Center (a US Department of Transportation Tier-1 University Transportation Center) and a CN Railway Research Fellowship in Railroad Engineering. I am sincerely thankful to these organizations for their support. I also want to thank Eric Wilson of Berkeley Simulation Software for academic use of Rail Traffic Controller simulation software.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

1.1. Purpose of the Study 1

1.2. Background and Current Problem 1

1.3. Objective and Scope of the Study 13

1.4. Structure of Dissertation 14

1.5. Summary of Dissertation Contributions 15

CHAPTER 2: EVALUATING THE IMPACT OF TRAFFIC HETEROGENEITY AND LEVEL OF SERVICE ON CAPACITY 20

2.1. Overview of the Current Status 21

2.2. Methodology 23

2.3. Case Study 32

2.4. Discussion and Conclusion 48

CHAPTER 3: COMPARING CAPACITY EXPANSION STRATEGIES FOR SINGLE-TRACK LINES UNDER MIXED OR FLEXIBLE OPERATION 51

3.1. Overview of the Current Status 51

3.2. Approach to Evaluating Performance of Alternative Expansion Strategies 53

3.3. Case Study 59

3.4. Extension of Case Study to the Incremental Benefit of Second Main Track 81

3.5. Discussion and Conclusion 84

CHAPTER 4: OPTIMIZATION OF SIDING LOCATION FOR SINGLE-TRACK LINES UNDER STRUCTURED OPERATIONS 87

4.1. Overview of the Current Status 87

4.2. Methodology 89

4.3. Case Study 102

4.4. Discussion and Conclusion 108

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH 112

5.1. Conclusion 112

5.2. Future Study 115

REFERENCES 121

CHAPTER 1: INTRODUCTION

1.1. Purpose of the Study

The objective of this dissertation is to develop new capacity evaluation and infrastructure planning techniques that consider the impact of relationships between train operating plans, train characteristics and train-specific levels of service on line capacity.

1.2. Background and Current Problem

Railroad line capacity can be defined as the maximum allowable flow rate of trains passing a point per unit time while maintaining a required level of service (Abril et al., 2008). North American railroads have an ongoing business incentive to properly match railway line capacity to traffic demand. While insufficient capacity reduces the level of service to railway customers, excess capacity, or poorly located capacity expansion projects, represents an inefficient use of railroad capital. The forecast increase in future rail traffic, and corresponding changes in demand for railway capacity, will require railroads to make strategic decisions regarding the infrastructure and operational changes required to meet this demand.

North American railroad operations and infrastructure planning are typically conducted based on practitioner experience combined with detailed simulations of train operations (Bronzini and Clarke, 1985). Recent trends in rail traffic have resulted in operating conditions that fall outside the realm of historical experience and may lend themselves to different types of analytical capacity evaluation and optimization tools. These trends include growth of freight rail traffic and expansion of passenger service on freight corridors, with a resulting increase in rail traffic heterogeneity and the need for more precisely scheduled operations. These trends are compounded by the limited capacity of single-track lines that comprise the majority of the North

American rail network. Since these issues are highly relevant to the analytical capacity evaluation and optimization techniques developed through this research, they will be introduced in more detail in the following subsections.

1.2.1. Long-Term Growth of Traffic Volume

Although rail traffic volumes may fluctuate over the short-term, long-term demand is expected to increase (HDR and Transit Safety Management, 2006; AAR, 2007; AAR 2015). In the US, freight rail traffic volumes steadily increased from 1990 to a peak in 2006 before declining during a period of economic recession (Figure 1.1). Since 2009, the economic recovery has again spurred increases in freight transportation demand and certain traffic metrics have reached new all-time highs (AAR, 2015). Although traffic has increased, the track mileage owned by Class I railroads has been decreasing since 1990. The combined effect of these trends is illustrated by an increase in daily average freight train-miles per track-mile owned. Train-miles per track-mile provides a more direct measure of train density across the rail network than other

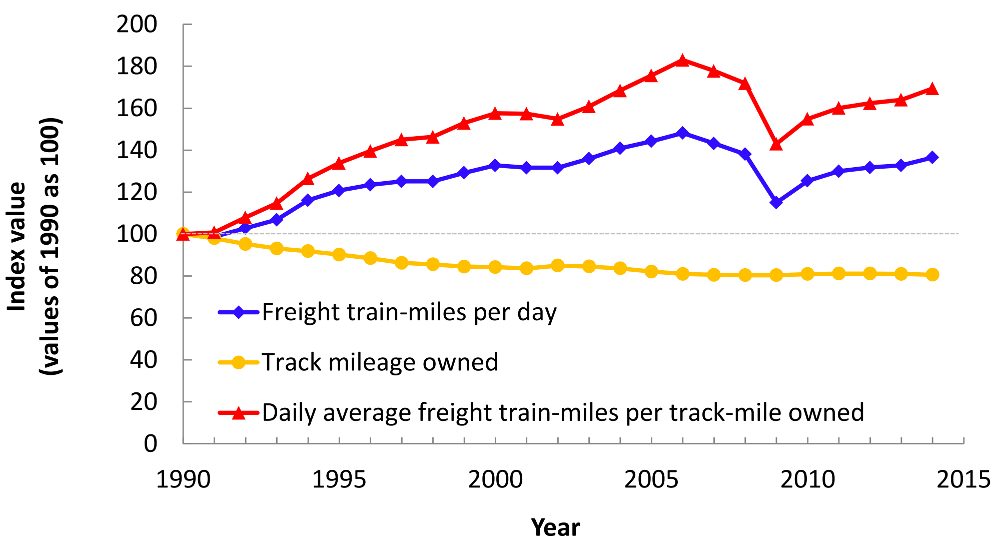


Figure 1.1. US Class I railroad traffic volume, train density, and track mileage owned from 1990-2014 (AAR, 2015)

volume metrics such as carloads or ton-miles. The near 60-percent increase in train density on the network of Class I railroads between 1990 and 2014 (Figure 1.1) suggests that railroads are facing unprecedented demand for railway line capacity. The increase in traffic density is due to both the long-term growth of traffic volume and the reduction in total length of track in the Class I rail network.

1.2.2. Traffic Heterogeneity

Dingler et al. (2009) defined the difference between the speed, priority, acceleration and braking characteristics of trains that serve the domestic intermodal, bulk freight and passenger rail markets as "traffic heterogeneity". They also used simulation to show that with the same number of trains per day, heterogeneous train types consume more capacity than operation of homogeneous train types, resulting in a lower level of service. There continues to be interest in expanding intercity passenger and commuter rail services, including increasing both train frequency and speed on existing freight corridors (Bing et al., 2010; Martland, 2010). Introducing additional passenger service to a freight corridor increases the heterogeneity. This reduces the available time and space for operation of freight trains (Sogin et al., 2013a; Shih et al., 2015a) and overall mainline capacity (Sogin, 2013; Sogin et al., 2013b; Shih et al., 2015a).

Quantifying the impact of heterogeneity on railway capacity has been a focus of many railway operations researchers. European and North American approaches to the subject differ however, reflecting the different characteristics of traffic heterogeneity in their respective rail systems. In Europe, most rail capacity studies or tools are related to schedule-based analysis. The variability of headways in a train schedule is an important concept used to quantify traffic heterogeneity. Carey (1999) proposed several headway-related indices to measure traffic heterogeneity at a single location on a network. Based on a similar concept proposed by Carey (1999) and the UIC 406 compression technique for capacity analysis (UIC, 2004), Vromans (2004; 2005) developed two representative indices termed “SSHR (Sum of Shortest Headway Reciprocals)” and “SAHR (Sum of Arrival Headway Reciprocals)”, that take the headway interval of two consecutive nodes (stations, yards or junctions) in a network into account. Landex (2008) combined the two indices developed by Vromans and created a single compact index; however, the computational process relies on a pre-determined train schedule and is thus not applicable for freight-dominated corridors in North America.

Traffic heterogeneity on North American mainlines is quantified by the variation in train priority and speed (Dingler et al., 2009). Krueger (1999) suggested that the impact of speed and priority variation between trains can be captured by the average speed and the expected number of meets and passes. Additionally, his parametric model and a regression model presented by Gorman (2009) can be used to model the performance of traffic with multiple train types. Harrod (2009) used a train dispatching optimization model to capture the effect of passenger operation on a freight corridor and also observed the negative impact of frequency and speed of passenger trains on freight traffic delay. Lai et al. (2012) proposed a Base Train Equivalent unit to quantify the relative effect of traffic heterogeneity on line capacity. Sogin (2012) and Sogin et al. (2013b) investigated the performance of heterogeneous traffic on several incremental capacity expansion strategies for a single-track line with a high density of passing sidings and equal siding spacing. Atanassov et al. (2014) applied a similar approach to quantify the impact of traffic heterogeneity on the performance of several capacity expansion strategies for single-track lines with unequal siding spacing.

The North American studies mentioned above only considered the impact of priority and speed variation as the impact of traffic heterogeneity, but not changes in train operating style and schedule flexibility (Figure 1.2). This dissertation research addresses the combined impact of schedule flexibility, operating style, priority, and speed on train performance and line capacity. A more comprehensive definition of operating style and schedule flexibility will be introduced in the next subsection.

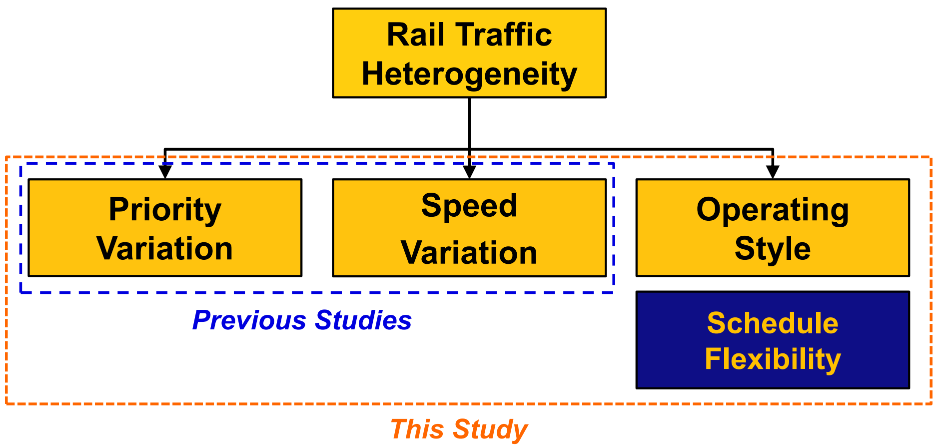


Figure 1.2. Factors considered in the previous and this study

1.2.3. Operating Styles and Schedule Flexibility

In this dissertation the schedule flexibility of a train is defined by the variation in its departure time and trip time (Figure 1.3). A train's departure time flexibility is defined as the potential range of its departure time from an initial terminal, or the beginning of a particular route segment under study. Once a train departs, there will also be variability in its travel time to its final terminal, or the opposite end of the route segment. Trip time flexibility can also be described by the range in the time-space path of a train. Departure and trip time flexibility have a direct relationship to schedule flexibility. Since higher schedule flexibility leads to higher uncertainty in train arrival time, it is inversely related to Level of Service (LOS) (Figure 1.4).

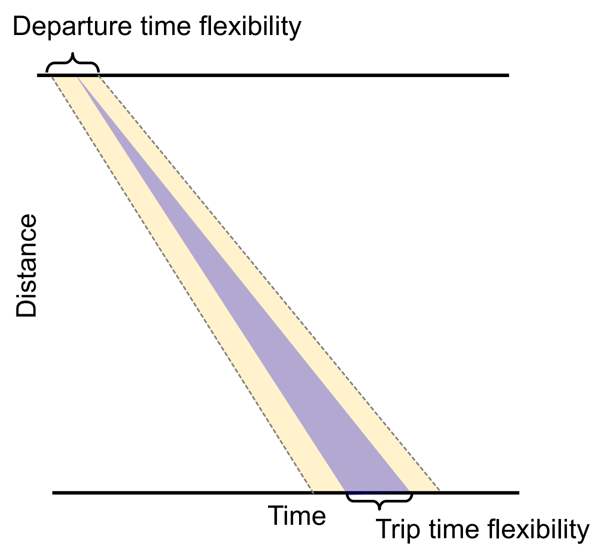


Figure 1.3. Departure and trip time flexibility

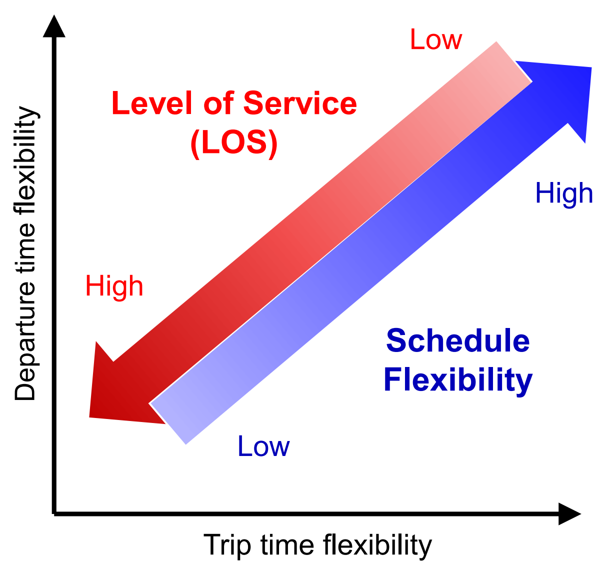
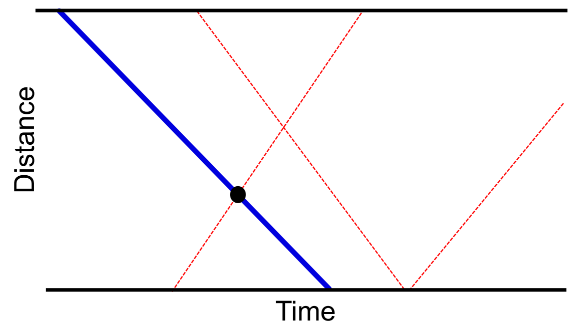
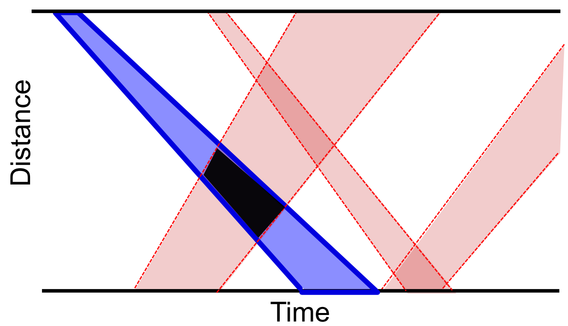


Figure 1.4. Relationship between departure time flexibility, trip time flexibility, schedule flexibility, and Level of Service (LOS)

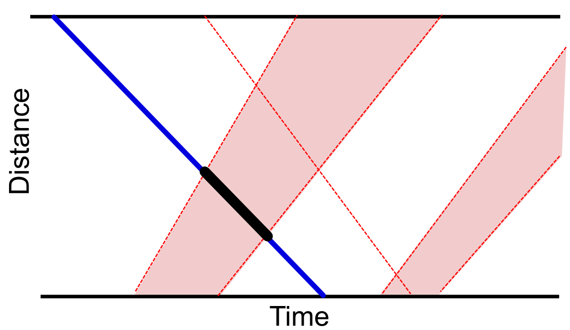
Operating style refers to the variation in schedule flexibility observed across the individual trains operating on a mainline during a given period. Rail systems adopt different operating styles according to their customer requirements and business needs (Figure 1.5).



(a)



(b)



(c)

Figure 1.5. Examples of different railway operating styles (a) structured operation (b) flexible operation (c) mixed operation

For North American freight railroads, the business objective of maximizing the length of trains in carload freight service requires terminal operators and dispatchers to dynamically adjust predefined train plans. Additional departure time flexibility is required in the event of insufficient crew or locomotive availability, or train makeup requirements. After trains depart terminals, running time flexibility is necessary to accommodate random disruptions such as unanticipated meets or passes with other late trains, mechanical failures, signal failures, temporary slow orders, or track inspection delays. As a consequence, predefined train operating plans in North America are relatively imprecise compared to railway operations with fixed timetables.

It was not always this way; for more than a century, North American freight trains generally conformed to scheduled operation. Meets and passes generally occurred at predetermined times and locations according to a detailed timetable. However, beginning in the 1950s and increasingly in the 1960s and 1970s, North American railroads gradually adopted a new operating style in which trains were "held for tonnage", meaning they only departed a terminal when some maximum number of cars had accumulated. This practice increased productivity but at the cost of reliability. The development of unit trains in the 1970s also contributed to less predictable train schedules. At the same time, scheduled passenger trains were gradually discontinued on all but a few mainline routes. Meanwhile, infrastructure was over-built for the amount of traffic so there was excess capacity. Without scheduled passenger trains and with excess capacity, there was less need to maintain precise disciplined timetable operations. Meet and pass times and locations were dynamic, arranged by dispatchers monitoring the progress of trains over a line. This operating style largely persists to this day and was named "improvised operation" by Martland (2010) and is termed "flexible operation" in this study.

The opposite operating style, where the operators carefully adhere to planned train paths, meet locations, dwell times and routes from origin to destination is termed "structured operation" (Martland, 2010). In contrast to freight operations, North American passenger and transit systems try and follow this type of operating style, although most operations still must cope with a variety of unplanned events. Under structured operations in Europe, dispatchers often have little flexibility, and their responses to schedule disruptions are usually prescribed by some emergency handling procedures or a pre-set rescheduled timetable (Norio et al., 2005; Luethi et al., 2007). Both flexible and structured operating styles occur together on North American shared corridors. The operating style on these corridors is referred to as "mixed operation" in this study.

Operating style affects train delay and line capacity. Each of the three time-distance diagrams in Figure 1.5 contain four train paths under a different operating style. The schedule flexibilities of each train follow the characteristics of the corresponding operating style. The train paths are indicated by the blue line or band, and the conflicts by the black dot (Figure 1.5a), area (Figure 1.5b), and line (Figure 1.5c). They represent the different range of traffic conflicts in the time-distance space that trains occupy. A particular single-track line will have different line capacities depending on the variation in the range of traffic conflicts created by each operating style (Shih et al., 2016a; 2016b).

In many countries outside North America, both passenger and freight trains use structured operation if a predetermined train plan is not disrupted. By contrast, all three operating styles can be found in different places in the North American rail network. In order to understand the fundamental interaction between operating style and train delay on a typical North American single-track mainline, Dick and Mussanov (2016) examined the capacity impact of different operating styles by quantifying operational flexibility. They measured the effect of varying train departure randomness on train delay and LOS for a given traffic volume. The authors examined homogeneous traffic, but did not quantify the combined impact of train priority, speed variation and operating style.

A comparison between the operating style of passenger-dominant and freight-dominant systems and its relationship to capacity evaluation was discussed by Pouryousef et al. (2015). Capacity studies related to freight rail systems focus on simulating the random factors involved in their stochastic operating environment (Pouryousef et al. 2013). Passenger rail capacity studies emphasize the efficiency and robustness of the predefined schedules through optimization, and the strength of the emergency schedule to mitigate disruption (Ekman, 2004; Norio et al. 2005; Pouryousef et al. 2013). Neither of these capacity study types can adequately address the current capacity problems on shared corridors with a mixed operating style. Optimization tools for structured operations cannot handle unscheduled trains. Simulation-based approaches for mixed and flexible operations can generate basic statistics depicting traffic performance, but a standardized approach to evaluate line capacity under specific passenger level-of-service and timetable requirements at intermediate stops is still required.

The operating style on a corridor can also be connected to the length of infrastructure or operations planning period. The longer the planning period, the greater the uncertainty in the specifics of future train plans. Methods, processes, and tools applicable for scenarios with a larger degree of randomness, or more flexible operating styles, can be used in these cases. Similarly, tools developed for structured operating styles are more appropriate for short-term planning. Understanding the effect of different operating styles on train delay and mainline capacity is another important characteristic that modern capacity analysis tools should consider. This study seeks to develop new tools and approaches for capacity evaluation that are better-suited to different operating styles found on North American railroads.

1.2.4. Single track with Insufficient Capacity

In general, North American freight railroads attempt to construct and maintain as little excess infrastructure as possible. This implies that network capacity available for new traffic is always limited. The steady growth in traffic in the 1990s and early 2000s (Figure 1.1) led railroads to invest in substantial infrastructure expansion and improvement projects. Many of these involved adding second-main tracks to key segments of high-density rail corridors on their core network. Other portions of the network with historically lower traffic density remained single track with widely spaced passing sidings; however, some parts of this lower density network have experienced recent increases in traffic due to growth in transportation of ethanol and crude oil. In response, railroads have shifted part of their capital investment plans to construct new passing sidings or upgrade signal systems on these lines. In order to accommodate this expanded traffic, infrastructure or operating solutions must be applied to solve the resultant congestion. Most North American railway capacity research has focused on higher-density lines with much less attention given to developing methodologies to effectively plan expansion for these types of low density lines. In this dissertation I describe new tools I have developed for evaluating both passing siding and double-track projects on single-track lines.

1.2.5. Existing Capacity Evaluation Tools

According to Lai (2008), existing railway capacity analysis tools can be categorized into four types: simulation (Petersen, 1982; Leilich, 1998; Salido et al., 2012; Stenstrom et al., 2013; Sipilä, 2015), optimization (Ahuja et al., 1993; Marin and Salmeron, 1996; Lai 2008; Lai and Barkan 2011), analytical approaches (Bronzini and Clarke, 1985; Chen and Harker, 1990; Parkinson and Fisher, 1996; Burdett and Kozan, 2006; Lindner, 2011; Bonsra and Harbolovic, 2012; Salido et al., 2012; Jensen et al., 2017), and parametric models (Prokopy and Rubin, 1975; Krueger, 1999; Mitra and Tolliver, 2010; Murali et al., 2010; Lai et al., 2012).

In general, current simulation tools capture the interaction between trains and infrastructure (Bronzini and Clarke, 1985) but respond slowly to changes in infrastructure or traffic inputs. Additionally, simulation outputs (i.e. train dispatching results) represent feasible, but not necessarily optimal, solutions. Additional analysis with an analytical or optimization tool is required to obtain optimal solutions. Optimization models can generate the optimal solution to a problem in a reasonable time, but sometimes over-simplify details or ignore stochastic factors due to computational constraints.

Analytical approaches, such as UIC 406 (UIC, 2004; UIC, 2013) or TCRP Report 13 (Parkinson and Fisher, 1996), respond quickly to changes in inputs to obtain an optimal solution but have little computational power. They do not typically consider the details of operational randomness precisely. Analytical approaches are often limited in the number of inputs they can consider, resulting in sub-optimal solutions.

Parametric models are often statistical models based on regression of simulation results or field data (Lai, 2008). They are responsive and consume little computational power, making them well-suited for network analysis. Although parametric models may have the ability to consider a certain degree of randomness, they do not directly generate optimal solutions. Like a simulation model, combined use with optimization or analytical model is needed to obtain optimal solutions.

1.3. Objective and Scope of the Study

An overall theme of this dissertation is to demonstrate how operating style directly influences the types of capacity analysis tools that are most applicable to a given rail corridor under study. The objective of this study is to develop new railway capacity evaluation tools and infrastructure planning techniques to address infrastructure or operations planning challenges under different operating styles. Two main research questions will be answered during the development of
these tools:

- What is the relationship between the operating style, variability of train priority and speed, and the capacity of a single-track line?
- Can the properties of this relationship be used to gain insight on where to conduct capacity expansion projects or operational changes?

The type of planning tool should be matched to capacity and infrastructure planning scenarios based on properties of the infrastructure, traffic and operating styles. Several previous studies quantified fundamental relationships between infrastructure and traffic but did not systematically study operating style (Peterson and Taylor, 1987; Pawar, 2011; Lai et al., 2012). Consequently, this study aims to develop tools based on the demands of different operating styles. The types of tools appropriate for each operating style must have the ability to handle the level of traffic heterogeneity and schedule flexibility associated with that operating style. The developed tools can help practitioners expedite the planning process and yield new knowledge of railway capacity relationships that will allow railroads to maximize their operating efficiency.

1.4. Structure of Dissertation

This dissertation is presented in five chapters (Figure 1.6). In Chapter 1, I discuss current railway capacity topics, review existing tools and their drawbacks, and summarize my overall research plan. A capacity evaluation technique for mixed or flexible operation is developed in Chapter 2. The technique involves development of a regression model and a transformation process for calculating the maximum train throughput per day given the different LOS specific to individual types of trains. The technique can be used to evaluate the additional capacity consumption arising from rail traffic heterogeneity. In Chapter 3, I propose a capacity evaluation process similar to the one developed in Chapter 2 to compare four different infrastructure expansion strategies for single-track lines with sparse sidings under mixed or flexible operation. This provides a general guideline for evaluating these types of capacity expansion projects.

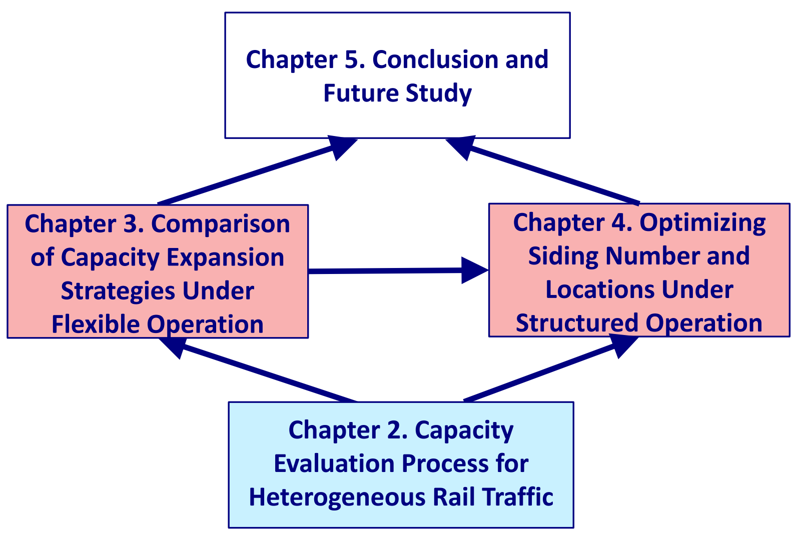


Figure 1.6. Structure of dissertation and the relationship between Chapters

The work in Chapter 3 suggests that optimization approaches may be useful in efficiently selecting locations for mainline capacity expansion projects. In Chapter 4 I develop an optimal siding location model that can identify the optimal number and location of passing siding projects and evaluate the performance of rail traffic under structured operations. The model can be used to determine the optimal infrastructure expansion plan for corridors dominated by passenger or premium-intermodal traffic that operates in a structured manner. In Chapter 5, I summarize the conclusions stemming from the research and outline directions for future study.

1.5. Summary of Dissertation Contributions

This dissertation expands the current understanding of capacity analysis techniques and introduces new approaches for rail capacity evaluation and planning (Figure 1.7). Most past studies or tools can be categorized by the degree of schedule flexibility they can consider and the practicality of their application. There are three levels of practicality: theoretical concepts, applied methods, and practical tools. There are also three levels of operating styles as mentioned previously.

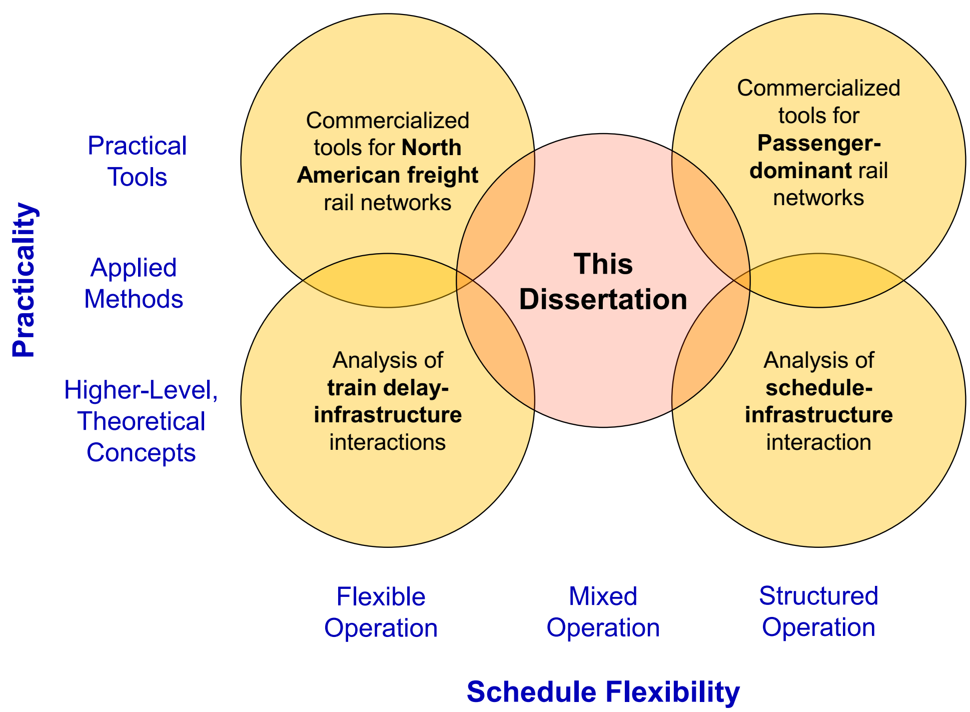


Figure 1.7. Relationship between techniques in this dissertation, the past studies, and tools

Within this space, four major groups of tools and techniques can be identified: analysis of schedule-infrastructure interaction (common in European capacity research literature) or train-delay-infrastructure interactions (common in North American capacity research literature), and commercialized tools for the European passenger-dominated corridors or commercialized tools for the North American freight-dominated rail network. There exists a direct connection between the theoretical concepts and practical tools within the realms of structured or flexible operation. However, there only exists a weak connection between structured and flexible operation formed by the few quantitative studies conducted or tools developed for mixed operations. This study develops applied methods for mixed operations to strengthen the connection between flexible and structured operations, while also increasing the tools available and understanding of all operation types. The tools developed in this dissertation incorporate fundamental interactions between rail traffic, infrastructure layout, and operating style on mainline capacity.

1.5.1. Contributions of Chapter 1

“Operating style” and “schedule flexibility” proposed in Chapter 1 offer a new perspective for quantifying and addressing rail traffic heterogeneity. They emphasize the importance of considering operational randomness in planned train departure times while making capacity-related decisions. Most past studies have focused on either passenger-dominant corridors (structured operation) or US freight rail corridors (flexible operation). The results of these studies cannot be applied on corridors with mixed operation. Mixed operation, is a common and growing type of operating style in the US freight rail network, and requires new methods for capacity evaluation and planning such as those presented in this dissertation.

1.5.2. Contributions of Chapter 2

The research in Chapter 2 evaluates and enables visualization of the capacity interactions between more than two different train types under flexible operations. The study improves the general understanding of the impact of traffic heterogeneity on railway operations. The analysis of additional passenger trains suggests that both the volume and the mixture of existing traffic are related to the ability of a line to handle additional traffic. The sensitivity analysis shows that relaxing LOS of a particular train type does not necessarily yield additional capacity. The capacity contour plots suggest that reducing train speed heterogeneity is an effective strategy to gain additional capacity when the required minimum run time of each train type is not violated. The capacity evaluation approach developed in this chapter provides a standardized method for practitioners to calculate the trade-off between line capacity, traffic heterogeneity and train-specific level of service requirements under mixed operations.

1.5.3. Contributions of Chapter 3

Chapter 3 proposes an evaluation process for assessment of capacity expansion alternatives of a single-track line with sparse sidings. The case study evaluates the efficiency and reliability of alternatives. The results indicate the relative effectiveness of the different capacity expansion alternatives and that alternative 1a "center out" provides the greatest efficiency with the lowest variability in delay for the evaluated single-track line. While this is not intended to be a general finding, the capacity evaluation process itself is a contribution. This approach can be used by railroads to develop high-level siding expansion program plans on single-track lines experiencing congestion under mixed operations. The trade-off between total length of second track, train delay, level of service, and line capacity based on the selected alternative are also shown. The results can help practitioners and researchers better understand fundamental relationships between changes in infrastructure, traffic volume, heterogeneity, and traffic delay.

1.5.4. Contributions of Chapter 4

The model developed in Chapter 4 provides the optimal number and locations of additional passing siding projects for a single-track line with sparse sidings under structured operation. It extends the Higgins et al., (1997) model by introducing practical engineering cost constraints. Although constraints on computation time restrict the model from generating a more robust infrastructure plan based on multiple train schedules, it can still be used as a prototype for future researchers or railroads to develop their own models.

1.5.5. Contributions of Chapter 5

The possible research directions introduced in the last chapter highlight the concept of traffic conflict analysis that can be used to evaluate the combined impact of variability in train priority, speed, and operating style. Two applications of traffic conflict analysis are provided as examples to demonstrate its use in developing generalized planning tools for all operating styles and ranges of schedule flexibility. One application prioritizes the infrastructure capacity projects on a single-track mainline based on the distribution of traffic conflicts. This application can provide railroad capacity planners with an alternative to detailed simulation or the method proposed in Chapter 3. The other application develops indices for the number and types of traffic conflicts encountered by a train to quantify the combined impact of variability in train priority, speed, and operating style. It is proposed that these indices can be used to build regression models that predict individual train delay. Both applications demonstrate the potential use of train traffic conflict analysis. They can also be used by future researchers as the basis for their own railway capacity evaluation methods.

CHAPTER 2: EVALUATING THE IMPACT OF TRAFFIC HETEROGENEITY AND LEVEL OF SERVICE ON CAPACITY

An earlier version of this research appears in:

Shih, M.C., C.T. Dick, and C.P.L. Barkan. 2015a. Impact of passenger train capacity and level of service on shared rail corridors with multiple types of freight trains. Transportation Research Record: Journal of the Transportation Research Board, 2475: 63-71.

In this chapter I develop a capacity evaluation process for mixed or flexible operation. The process can use either simulation or actual traffic data from a mainline segment of interest as inputs. The technique generates a relationship between line capacity (maximum possible throughput) and the variability of speed and priority according to a given level of service (LOS) requirement for each type of train on that line segment.

Reducing traffic heterogeneity or relaxing LOS can increase line capacity (Figure 2.1) (Krueger, 1999; Mattsson, 2007; Abril et al. 2008). The area of the triangle defined by the three axes is constant for a fixed infrastructure arrangement. In scenario A (red), the degree of traffic heterogeneity and the LOS is higher than scenario B (blue), resulting in lower relative capacity. Krueger (1999) attempted to capture this relationship using average traffic delay,

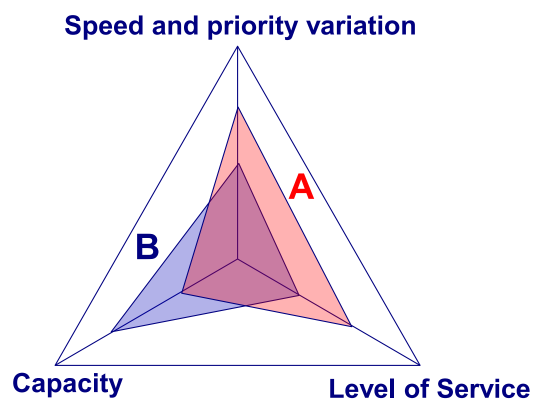


Figure 2.1. Example of the trade-off between line capacity, speed and priority variation, and LOS

heterogeneity-related factors, and the average LOS of traffic in his parametric capacity model. However, individual train types may have very different LOS requirements due to their differing operational needs (UIC, 2004; UIC, 2013). The process developed in this chapter improves upon previous methods by considering the impact of multiple train types, each with their own LOS requirements, on line capacity.

2.1. Overview of the Current Status

Rail line capacity can be measured in a number of ways (UIC, 2004; Landex et al., 2007; Abril et al., 2008; UIC, 2013), but two main approaches are used most frequently. The first is by track occupation rate or the proportion of time a segment is occupied over a defined time period (UIC, 2004; 2013). The compression method proposed by the International Union of Railways (referred to as UIC 406) uses this approach. Capacity can be calculated by the UIC 406 compression method based on a predetermined train schedule.

The second approach is to use the average train delay for a given traffic volume and to define capacity as the maximum allowable delay. Krueger (1999) used this concept to obtain maximum throughput of traffic per unit of time. He constructed a delay-volume curve with an exponential increase in average train delay as train volume increased on a given line segment (Figure 2.2). By specifying a maximum allowable average train delay, the largest amount of throughput that does not violate this maximum allowable delay defines the capacity of the rail line.

Since the flexible operating environment in North America does not fit into the strict schedule requirements of the first method, I use the concept of average train delay and maximum allowable delay as metrics for line capacity and traffic performance.

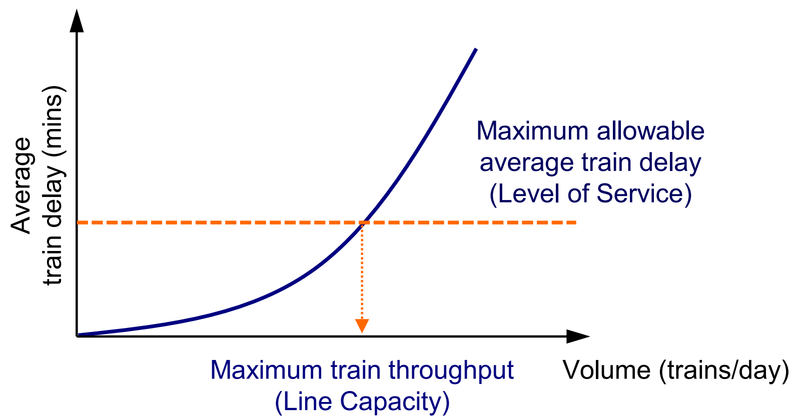


Figure 2.2. Defining capacity by maximum allowable train delay

Currently, both parametric and simulation models are used to evaluate the capacity of mixed or flexible operations on mainlines using the concept of maximum allowable train delay. Two parametric models frequently used by railroads and researchers in North American are the Federal Railroad Administration (FRA) parametric model (Prokopy and Rubin, 1975) and the Canadian National (CN) parametric model (Krueger, 1999). These models do not fully account for traffic heterogeneity, or the different operating styles and service quality requirements of different train types.

Simulation models can account for details of train operation randomness and traffic heterogeneity, but they require considerable time and effort to develop. Dingler et al. (2013) and Sogin et al. (2013a) investigated the impact of traffic heterogeneity on line capacity using simulation. They considered the interactions between two train types to understand the basic relationship between train delay and heterogeneity. However, the traffic mixture on most railway lines, and shared corridors in particular, usually contains more than two train types. Also, previous simulation studies only considered average train delay in their determination of capacity and thus do not directly apply to corridors where different train types have differing LOS requirements. The capacity evaluation technique developed in this chapter is capable of considering the trade-off between line capacity, traffic heterogeneity and LOS for different train types under mixed or flexible operation.

2.2. Methodology

The capacity evaluation process developed in this chapter requires railway traffic scenarios and their corresponding train delay data as inputs. A railway traffic scenario is a specific combination of traffic volume and mixture of train types comprising that traffic volume on a given mainline segment of interest. In the context of North American railway capacity research, and in this dissertation, train delay for a particular train is defined as the difference between its minimum free running time and its actual running time across a segment of interest for which line capacity is being determined. The minimum free running time is the time required for a train to traverse the segment of interest with no stops for meets or conflicts with other trains, while obeying all maximum authorized speeds and considering the acceleration and braking capabilities of the train. The actual running time is the time required for a train to traverse the same segment using simulated rail operations, or as observed in historical train operating data for that segment.

For this study, train delay data were obtained using Rail Traffic Controller (RTC) (Wilson, 2012) to simulate operation on a hypothetical rail line. Actual train delay data from rail lines with different traffic scenarios or outputs of other simulation platforms can also be used by the process to evaluate rail line capacity.

To develop the required train delay data for regression, the potential traffic scenarios of the target line should be simulated. Simulating all the potential scenarios could be time consuming, thus an experimental design was created to select a portion of the potential scenarios as the representative traffic scenarios for RTC simulation analysis. The scenarios were then simulated with given train characteristics and infrastructure properties to obtain corresponding train delay information. A polynomial regression model considering parameter interactions was constructed based on the delay output. The delay model was then transformed into a model for line capacity according to the minimum LOS (maximum allowable average train delay) for each train type. The transformed model offers a general capacity evaluation process that can be applied to any rail line; however, the particular line capacity model developed based on the traffic and infrastructure characteristics of the mainline used in this study is specific to that line. The process is summarized in Figure 2.3, and each step in the process is described in the following sections.

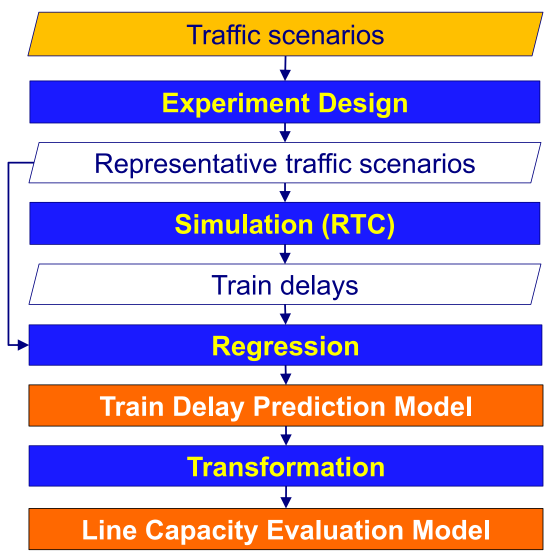


Figure 2.3. Flowchart of the capacity evaluation process

2.2.1. Experimental Design

The general capacity evaluation process developed in this chapter can be applied to lines with any number of train types. This study examines three types (passenger, intermodal and bulk unit freight trains with high, medium and low priority, respectively) because the interactions between train types are simpler to visualize.

Three indices are used to quantify the traffic volume and mixture: total traffic volume (*Q*), number of passenger trains (*P*), and bulk unit train traffic as a percentage of total freight traffic (*U*). The last value (termed "percent unit trains" in this chapter) provides a measure of the level of freight train heterogeneity on the route and, given values for the other two factors (total volume and number of passenger trains), allows for calculation of the number of intermodal and bulk unit freight trains. The number of passenger trains and percent unit trains also enable the analysis of the incremental impact of additional passenger trains on different freight traffic mixtures. Another train type can be used if the incremental impact of that train type is of interest. Two constraints are applied when the experiment matrix is created: the number of passenger trains cannot exceed the total traffic and, for each train type, the number of trains in each direction must be balanced.

Although not considered in this dissertation, the approach can be expanded to evaluate the capacity of scenarios with more than three types of trains. The total traffic volume and percentage of each train type (frequency of each train type as a percentage of total traffic) can be used as indices to quantify the volume and mixture of traffic with many train types. The constraint that specifies a balanced train type volume in each direction can be removed if the directions of the trains are not balanced.

An experimental matrix of traffic simulation scenarios is needed to obtain a delay response surface for the line under study across a range of traffic volumes and mixtures as defined by the three indices described above. Partial factorial design was used to select a subset of simulations from a full factorial design to eliminate redundant trials (Box and Soren, 1987). This partial factorial subset has similar delay response to that of the original experiment but is more efficient to run because it uses fewer traffic simulation scenarios.

2.2.2. Regression and Transformation

In the next step of the capacity evaluation process, the traffic scenarios and delay data from RTC simulation were used to construct a multivariate regression model for train delay of each individual train type. If historical train delay data were being used instead of simulation, a similar regression model would be developed based on the historical train delays and corresponding traffic indices. To provide a measure of capacity, the regression model with volume as an input and delay as an output must be transformed into a model for volume based on allowable delay (LOS) for each train type.

The transformation step in the process can be done graphically (Figure 2.4) or mathematically. The upper set of axes in Figure 2.4 displays the relationship between the average delay of intermodal trains and the freight traffic mixture (percent unit trains) for profiles of constant total traffic volume (ranging from 20 to 28 trains per day). By setting the maximum allowable average delay for intermodal trains to the LOS for intermodal traffic (D^{max}, 25 min in Figure 2.4), and intersecting this delay value with the profiles, the maximum traffic volumes that can be handled without violating the intermodal train LOS can be obtained for corresponding traffic mixtures. These points can be transferred to the lower set of axes and used to construct an intermodal capacity profile for a given LOS. This transformation process must be repeated for each train type, and the minimum of all the capacity profiles obtained is the final capacity profile.

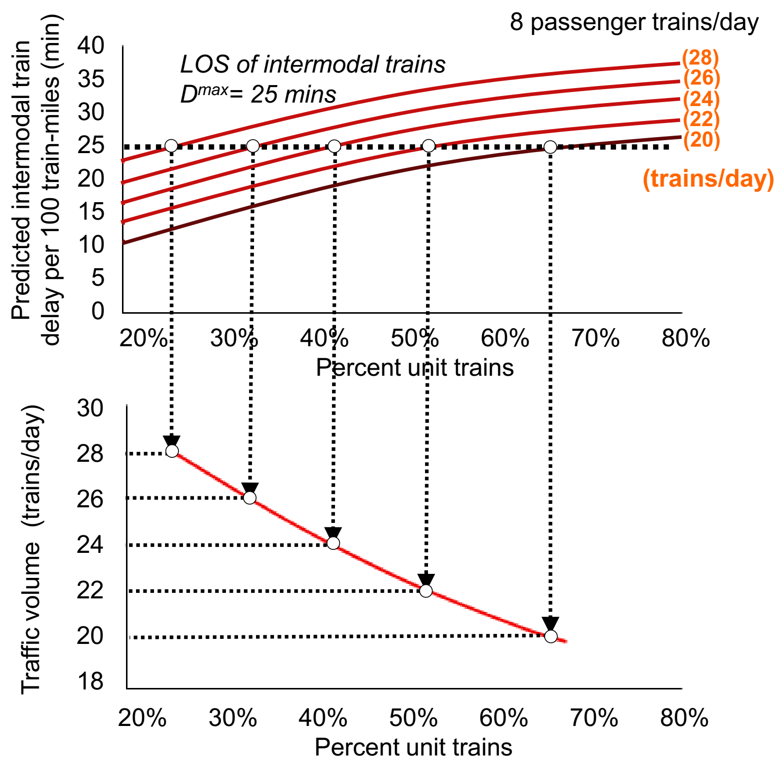


Figure 2.4. Geometrical concept of the transformation process

The values of this final profile represent the maximum throughput of trains for different freight traffic mixtures without violating the LOS of all train types.

Mathematically, the original polynomial regression model for delay of each train type can be represented as a quadratic function of total traffic volume (Equation 2.1). The quadratic function is adopted to approximate the exponential delay-volume relationship identified in previous studies (Krueger, 1999; Lai, 2008). The quadratic equation can be used to solve for traffic volume and transform the original function into Equation 2.2. The functions predicting D_t can be obtained through a linear regression model that contains the three indices (Q, P, U), and all the "train types" t, as categorical variables. The capacity profile of each train type can be obtained from Equation 2.2 by substituting the delay of train type t (D_t) with maximum allowable delay (D_t^{max}) according to the desired LOS for that train type. This transformation process is applied to each train type. Thus each combination of passenger train volume and freight traffic mixture will have three different allowable total traffic volumes (Q_t^{max}) based on the specific LOS for each train type.

$$D_t = f\,Q^2 + g_t(P,U)\;Q + h_t(P,U) \tag{2.1}$$

$$Q_t^{\max} = \frac{-g_t(P,U) + \sqrt{g_t\,(P,U)^2 - 4f\left[h_t(P,U) - D_t^{\max}\right]}}{2f} \tag{2.2}$$

Where:

D_t: Predicted average train delay per 100 train-miles of train type t

$D_t^{'max}$: Maximum allowable predicted average delay (LOS) of train type t

P: Number of passenger trains in total train traffic volume

U: Bulk unit freight trains as a percentage of total freight traffic (percent unit train)

Q: Total train traffic volume (includes the number of passenger trains)

$Q_t^{'max}$: Maximum throughput without violating LOS of train type t

(capacity profile of train t, including the number of passenger trains assigned)

f: Coefficient of the second order term of the delay-volume function of train type t

$g_t(P,U)$: Function represents the first order coefficient of the delay-volume function of train type t

$h_t(P,U)$,: Function represents the intercept of the delay-volume function of train type t

A maximization term added into Equation 2.2 can prevent “imaginary number” capacity when the LOS of a certain train type is not feasible relative to its predicted delay performance (Equation 2.3). Additionally, if $Q_t^{'max}$ is equal to or smaller than zero, it means LOS for that train type is infeasible.

$$Q_t^{\max} = \frac{-g_t(P,U) + \sqrt{\max\left\{g_t(P,U)^2 - 4f\left[h_t(P,U) - D_t^{\max}\right],\ g_t(P,U)^2\right\}}}{2f} \tag{2.3}$$

Equation 2.3 can only be used for three train types. For scenarios with more than three train types, the indices suggested below can be used to characterize the mixture of traffic:

> $m_1, m_2, \ldots, m_t, \ldots, m_T$: each represents the number of train type *t* as a percentage of total traffic volume *Q*, *T* is the total number of train types

If the incremental impact of a certain train type is the focus of the analysis, then the number of that train type can be used as an index as well, like the number of passenger trains (*P*) used in this study.

Based on the suggested indices, more general forms of the train delay prediction model (Equation 2.4) and transformed line capacity evaluation model (Equation 2.5) are proposed. Similar to the general indices for traffic with more than three train types, Equations 2.3, 2.4, and 2.5 are suggested as possible general forms; however, they were not validated in this research. Any appropriate indices can be used in both general functions to describe more complicated traffic mixtures, but the polynomial structure must be maintained to approximate the exponential delay-volume relationship and facilitate the transformation process described in this chapter.

$$D_t = f\,Q^2 + g_t(m_1, m_2, \ldots, m_T)\,Q + h_t(m_1, m_2, \ldots, m_T) \tag{2.4}$$

$$Q_t^{\max} = -g_t(m_1, m_2, \ldots, m_T) + \frac{\sqrt{\max\left\{g_t(m_1, m_2, \ldots, m_T)^2 - 4f\left[h_t(m_1, m_2, \ldots, m_T) - D_t^{\max}\right],\ g_t(m_1, m_2, \ldots, m_T)^2\right\}}}{2f} \tag{2.5}$$

$$Q^{fin} = \min(Q_1^{\max}, Q_2^{\max}, \ldots, Q_t^{\max}, \ldots, Q_T^{\max}) \tag{2.6}$$

For Equations 2.3 and 2.5, the final capacity profile is constructed from the lowest of the calculated individual train-type maximum traffic volume values to create a minimum profile (Q^{fin}) that governs the capacity of the line (Equation 2.6). This final capacity profile represents the maximum throughput of trains under different traffic mixtures without violating the LOS of all train types. Since different train types may govern capacity for different traffic mixtures, the final capacity profile may not be a smooth function.

In calculating the line capacity evaluation model, the transformation process (Equation 2.3) could potentially magnify the train delay prediction model error, resulting in large uncertainty in the predicted line capacity (Figure 2.5). A delay-volume relationship under a certain traffic mixture (quantified by *P* and *U*) will have a corresponding train delay confidence representing the potential range of the true mean value. The intercepts between the required LOS (maximum allowable train delay) and upper and lower bounds of this confidence interval, define a range of possible line capacity predictions given the error in the train delay model.

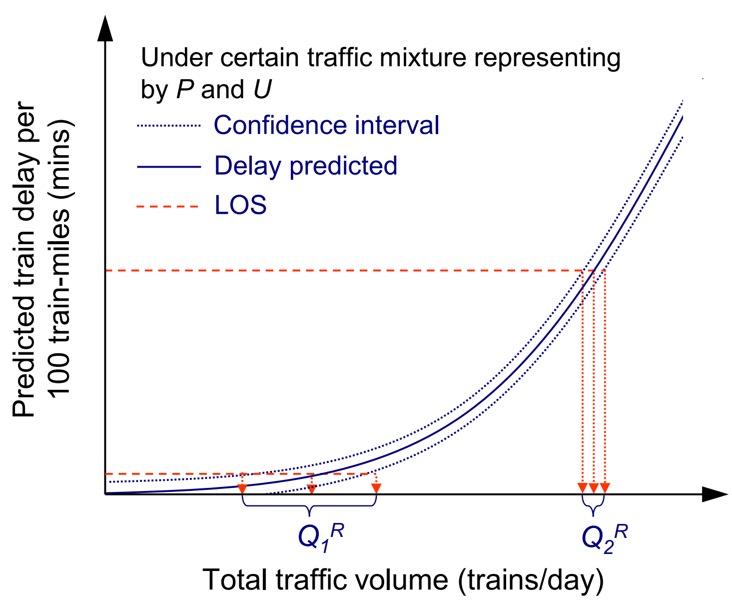


Figure 2.5. Magnification of delay prediction error in the transformation process

Two examples of the uncertainty in predicted line capacity due to delay prediction error are shown by Q_1^R and Q_2^R. Comparing the two examples, the potential range of predicted line capacity is greater (Q_1^R is greater than Q_2^R) when the LOS is set to a low value on the relatively flatter portion of delay-volume relationship. When the LOS of a train type is too strict, the robustness of the line capacity estimate is low.

To deal with the potential range of line capacity estimates, the minimum value within the range of these estimates could be defined as the line capacity; however, this could be an underestimate. If more robust capacity estimation is desired by a user, the assumed LOS can be relaxed. If it cannot be relaxed, more simulation or historical data could be collected to reduce the original delay prediction error and line capacity uncertainty.

The line capacity evaluation model is difficult to directly validate by simulation or historical data. Obtaining the LOS-heterogeneity-capacity relationship from simulation outputs or historical data is infeasible, so the validation process can only be done indirectly. Using the r-squared value of the train delay prediction model to assess its performance would be one approach. Another way to validate the line capacity evaluation model is to compare the average delay of each train type and given LOS based on the simulation results of validation scenarios. A validation scenario could be derived from a calculated line capacity value and its corresponding traffic mixture. For a scenario with a predicted line capacity of 22 trains per day, six passenger trains, and 50 percent unit trains under a given set of LOS requirements, a corresponding validation scenario with six passenger, eight intermodal, and eight unit trains must be simulated. The line capacity evaluation model is valid if the average train-type-specific delay obtained from the simulation is equal to its corresponding LOS requirements for at least one of the train types (i.e. at capacity at least one train type must be operating at its maximum allowable train delay).

2.3. Case Study

To illustrate the insights that can be gained through application of the capacity evaluation process developed in this chapter, it is demonstrated using a case study capacity analysis of a single-track, shared-use rail corridor. It begins with an evaluation of the incremental impact of passenger trains on the capacity of the line with different mixtures of existing freight train types. The analysis determines the relative impact of passenger trains on each freight train type. Since the LOS is somewhat subjective and may vary between railways and other practitioners, a sensitivity analysis was conducted on the maximum allowable delay of each train type to better understand its effect on capacity. The final part of the case study aims to capture the impact of speed heterogeneity on line capacity by comparing scenarios with a different set of maximum speeds assigned to each train type.

2.3.1. Simulation Parameters

Although any combination of three train types can be used, traffic composed of two types of freight trains and one type of passenger train was selected to represent the general traffic mixture on shared-use corridors for the case study. To provide the greatest contrast between train types, the freight traffic is composed of intermodal and bulk unit trains. The intermodal train type is used to represent freight trains with higher speed, priority and LOS. The bulk unit train type represents freight trains with lower speed, priority and LOS. The attributes of each type of freight train were set according to the characteristics of train types in the Association of

American Railroads *National Rail Freight Infrastructure Capacity and Investment Study* (2007) (Table 2.1).

Table 2.1. Simulation parameters of train characteristics

Parameter	Passenger trains	Intermodal trains	Bulk unit trains
Locomotive	2 GE P42	3 EMD SD 70	3 EMD SD 70
Number of cars	7 articulated Talgo cars	93 platforms	115 loaded hopper cars
Length (ft)	500	5,659	6,325
Weight (ton)	800	5,900	16,445
Horsepower per total ton	15.4	3.64	0.78
Maximum speed (mph)	75	55	35
Schedule stops	30-mile station spacing	None	None

Passenger trains are modeled after those used in short-haul, regional intercity service typical of those being used to increase passenger service speed and frequency (Table 2.1). The particular passenger train consist matches those used for the Amtrak Cascades service in the Pacific Northwest. For purposes of this study, the trains are scheduled to make station stops for 3 minutes every 30 miles.

The route infrastructure is a 242-mile single-track line with sidings uniformly spaced 10 miles apart (Table 2.2). These characteristics emulate a relatively busy, single-track line. Use of a general route helps avoid variance due to uneven siding spacing, specific curvature, and grade profiles, so as to isolate the more fundamental relationships between delay and traffic mixture.

Table 2.2. Simulation parameters of mainline properties

Parameter	Value
Total length (mile)	242
Siding spacing (mile)	10
Average signal spacing (mile)	2
Turnout speed (mph)	45
Traffic control system	2-block, 3-aspect CTC
Grade and curvature (%)	Both 0%

The index levels used in the experimental design process are listed in Table 2.3. A partial factorial design was used to select 32 traffic scenarios out of the full 243 traffic scenarios. Each scenario was simulated with an RTC run of six replicates, each lasting five days, to develop train performance data for a total of 30 days. The repetitions of each scenario generated enough traffic data with realistic variation to support statistical analysis.

Table 2.3. Index levels used in the experiment

Index	Low	Medium	High
Total traffic (trains/day)	6	22	38
Number of passenger trains (trains/day)	0	18	36
Percent unit trains (%)	6	50	94

In the RTC simulations, the departure pattern of trains is randomized to represent possible variation in train schedules. A 30-day simulation for each scenario was repeated six times in RTC with different randomization values to generate 180 days of train delay data to support statistical analysis. The delays of all trains of a given type are averaged over all 180 simulated days of train operations to calculate the average train delay response for that train type under the simulated traffic scenario in the experiment design. The average train delay response is further normalized by the length of the route to produce an output value of train delay in minutes per 100 train-miles.

2.3.2. Constructing Train Delay Prediction and Line Capacity Evaluation Models

The regression approach used the calculated train delay response from the 24 traffic scenarios (Table 2.4) in the experiment matrix to construct a prediction model for average train delay. There are repeated scenarios used to balance the experiment matrix. These scenarios were simulated with a different set of random seeds to obtain variable delay responses. Instead of the average delay for a simulation scenario, the delays for each individual train on each day of the simulation scenario were used to fit the regression model. There are a total of 30,526 train delay points: 10,425 passenger trains, 9,870 intermodal trains, and 10,231 unit trains. The large amount of data in the regression reduces the error of the train delay prediction model. The estimated error of the line capacity evaluation model based on the confidence interval of the train delay prediction model output ranges from approximately 0.3 to 0.8 trains/day. This error estimate shows the stability of the line capacity evaluation model constructed in this case study.

Table 2.4. Experiment scenarios and the corresponding average delay of each train type

	Index			Average train delay per 100 train-miles (mins)		
Scenario	Number of passenger trains *(P)*	Total traffic *(Q)*	Percent unit trains *(U)*	Passenger trains	Intermodal trains	Bulk unit trains
1	2	6	50.0%	1.0	3.8	6.0
2	2	6	50.0%	0.9	3.0	7.2
3	2	20	88.9%	1.7	12.9	21.5
4	2	22	90.0%	2.2	12.8	31.7
5	2	22	90.0%	1.8	13.1	24.1
6	2	24	90.9%	1.9	18.5	27.1
7	2	38	94.4%	4.7	26.9	74.8
8	2	38	55.6%	2.8	36.1	58.6
9	2	22	10.0%	1.0	12.1	34.1
10	2	38	50.0%	5.0	66.0	42.1
11	2	36	23.5%	4.0	28.4	64.8
12	2	38	5.6%	6.1	45.2	55.2
13	2	38	5.6%	2.6	52.1	60.4
14	10	38	92.9%	6.2	54.7	75.3
15	16	20	50.0%	8.0	42.4	94.3
16	16	38	9.1%	3.2	15.7	93.2
17	18	22	50.0%	3.6	14.4	45.1
18	18	38	90.0%	4.1	38.1	67.4
19	18	38	10.0%	7.3	31.9	88.1
20	18	38	10.0%	6.8	34.1	79.7
21	20	24	50.0%	7.8	63.4	101.4
22	20	38	88.9%	7.8	65.6	107.3
23	34	38	50.0%	7.9	40.0	102.1
24	34	38	50.0%	6.7	44.6	116.2

To capture the response surface of average train delay, the first order, second order, and interaction terms of each index were used to construct the prediction model. Box and Wilson (1951) suggested using a second-degree polynomial model including the interaction terms of each factor as an approximation of the response variable surface. Since this model is easy to estimate and apply, it was used to construct the regression model in this study. A stepwise regression approach using combined search (SAS Institute, 2017) based on BIC ratio (Burnham and Anderson, 2003) was conducted to select the terms that form the train delay prediction model. There are several frequently used likelihood ratios for model selection, including p-value, BIC, and AICC (Burnham and Anderson, 2004). The BIC ratio was chosen since it penalizes the number of selected terms more than the other two methods, which is consistent with the principle of Occam's Razor and prevents overfitting.

The model developed through regression analysis (Table 2.5) has an R-squared value of 0.87, indicating the regression model is precise enough to capture the delay response of the traffic, thus no further validation process was conducted. All of the terms listed in Table 2.5 were selected by the stepwise regression approach. The results of F-test and t-tests show the statistical significance of the train delay prediction model and each selected term with a confidence level of 0.95.

Elaborating on the form of the model in Table 2.5:

- The P, Q, and U terms are the indices of number of passenger trains, traffic volume, and percent unit trains, respectively. The $\bar{P}$, $\bar{Q}$, and $\bar{U}$ terms are the means of P, Q, and U term values, respectively.
- Before regression, the continuous variables were centered using the average values to reduce collinearity (SAS Institute, 2017), hence the presence of the average $\bar{P}$, $\bar{Q}$, and

Table 2.5. Statistics of train delay prediction model and the selected terms

Term	Coefficient	t Ratio	p-value of t-test
Intercept	-137.8496	-4.25	<0.0001
$(P-\bar{P})$	0.3460	5.96	<0.0001
$(Q-\bar{Q})$	2.7110	2.22	0.0262
$(U-\bar{U})$	25.8028	12.50	<0.0001
Type:P	-36.1024	-20.78	<0.0001
Type:I	-15.0820	-22.39	<0.0001
Type:U	51.1844	73.58	<0.0001
$(P-\bar{P})^2$	-0.0228	-8.38	<0.0001
$(Q-\bar{Q})^2$	0.0147	3.32	<0.0001
$(U-\bar{U})^2$	-26.4025	-17.15	<0.0001
$(P-\bar{P})\times(U-\bar{U})$	0.6778	5.09	<0.0001
$(Q-\bar{Q})\times(U-\bar{U})$	1.0336	10.68	<0.0001
$(P-\bar{P})\times(Q-\bar{Q})$	0.1016	8.12	<0.0001
$(P-\bar{P})\times$*Type:P*	-0.7874	-18.16	<0.0001
$(P-\bar{P})\times$*Type:I*	0.1150	2.29	0.0234
$(P-\bar{P})\times$*Type:U*	0.6724	13.19	<0.0001
$(Q-\bar{Q})\times$*Type:P*	-1.52521	-6.44	<0.0001
$(Q-\bar{Q})\times$*Type:I*	-0.7077	-3.72	<0.0001
$(Q-\bar{Q})\times$*Type:U*	2.2329	4.24	<0.0001
$(U-\bar{U})\times$*Type:P*	7.9698	5.43	<0.0001
$(U-\bar{U})\times$*Type:I*	16.5163	12.51	<0.0001
$(U-\bar{U})\times$*Type:U*	-24.4861	-22.18	<0.0001
R-squared value:	0.8701	p-value of F-test:	<0.0001

Where: $\bar{P}$ = 11.7968, $\bar{Q}$=29.2226, $\bar{U}$= 0.4097 (40.97%)

$\bar{U}$ terms in the model.

- The train type t is split into three binary terms, *Type:I*, *Type:P*, and *Type:U*, to represent the intermodal, passenger and bulk unit train types, respectively. The binary terms allow the single model to predict the average train delay for a specific train type.
- The $(P-\bar{P})^2$, $(Q-\bar{Q})^2$, and $(U-\bar{U})^2$ terms are the second-order polynomial terms.
- The interaction terms are indicated by a multiplication sign "×" between the different first order terms, e.g. $(P-\bar{P})\times(U-\bar{U})$ or $(U-\bar{U})\times$*Type:U*.

The train delay prediction model uses a single expression for all train types; however, to facilitate the interpretation of its coefficients, the model can be split into three functions, one for each train type (Table 2.6). These train-specific models were obtained through the following four algebraic simplification steps:

Table 2.6. Coefficients of the train delay prediction function of each train type

	Train types		
Term	Passenger trains	Intermodals	Bulk units
Intercept	-173.9520	-152.9316	-86.6652
P^*	-0.4414	0.4610	1.0184
Q^*	1.1858	2.8260	1.0184
U^*	33.7726	42.3191	1.3167
P^{*2}	-0.0228	-0.0228	-0.0228
Q^{*2}	0.0147	0.0147	0.0147
U^{*2}	-26.4025	-26.4025	-26.4025
$P^*\times U^*$	0.6778	0.6778	0.6778
$Q^*\times U^*$	1.0336	1.0336	1.0336
$P^*\times Q^*$	0.1016	0.1016	0.1016

- Add the appropriate "*Type:I*", "*Type:P*", or "*Type:U*" term to the overall intercept to obtain the intercept for each train type.
- Add the appropriate $(P-\bar{P})\times Type{:}I$, $(P-\bar{P})\times Type{:}P$, or $(P-\bar{P})\times Type{:}U$ term to the coefficient of the $P-\bar{P}$ term to obtain the coefficient of the P^* term for each train type.
- Add the appropriate $(Q-\bar{Q})\times Type{:}I$, $(Q-\bar{Q})\times Type{:}P$, or $(Q-\bar{Q})\times Type{:}U$ term to the coefficient of the $(Q-\bar{Q})$ term to obtain the coefficient of the Q^* term for each train type.
- Add the appropriate $(U-\bar{U})\times Type{:}I$, $(U-\bar{U})\times Type{:}P$, or $(U-\bar{U})\times Type{:}U$ term to the coefficient of $(U-\bar{U})$ term to obtain the coefficient of the U^* term for each train type.

Additionally, to visually simplify the written model, the notation $(P-\bar{P})$, $(Q-\bar{Q})$, and $(U-\bar{U})$ is replaced by P^*, Q^*, and U^*, respectively, in Table 2.6.

The train delay prediction model coefficients can be interpreted as follows:

- Since the average delay of the passenger trains is the lowest, with intermodal second, and bulk unit trains the highest, the relative order of the intercepts is reasonable.
- The coefficients of all Q^* and Q^{*2} terms yield a convex delay-volume curve, similar to the shape of the exponential delay-volume curve in Krueger's study (1999).
- The coefficients of all U^* and U^{*2} terms indicate the delay-heterogeneity curve is concave when values of P^* and Q^* terms are fixed. This result is consistent with the shape of the delay-heterogeneity curve obtained by Dingler et al. (2009).
- The coefficients of all P^* and P^{*2} terms show a concave relationship between number of passenger trains and train delay where there is a critical number of passenger trains under a certain traffic volume, and freight traffic mixture. Below this number of passenger trains, the passenger train type does not comprise the majority of the traffic and additional passenger trains increase traffic heterogeneity. Above this number of passenger trains, the passenger train type comprises the majority of the traffic, and additional passenger trains decrease heterogeneity (increase traffic homogeneity).
- The positive coefficients of $P^* \times U^*$ and $Q^* \times U^*$ terms indicate that the impact of additional passenger trains or overall increase in traffic volume is greater if the current freight traffic contains more bulk unit trains.
- The positive coefficients of $P^* \times Q^*$ terms denote that the impact of P or Q term on delay is greater when the value of the other term is greater; additional traffic volume and passenger trains have a compounding effect.

The parameters in Table 2.6 are used to calculate the case-study specific f coefficient Equation 2.7 and the g_t and h_t functions Equations 2.8 and 2.9 within the general train delay prediction model (Equation 2.1) and line capacity evaluation model (Equation 2.3). Equation 2.10 is the domain of the model inputs. The models constructed in the case study are only valid when its input is within this domain. The values in Equation 2.10 were calculated based on the mean value (Table 2.5) and range (Table 2.3) of each index in the traffic scenarios. The train delay prediction models can be obtained by substituting Equations 2.7 to 2.9 into Equation 2.1. The line capacity evaluation model can similarly be obtained by substituting Equations 2.7 to 2.9 into Equation 2.3 then using Equation 2.6 to determine the final capacity profile. Additionally, if the line capacity calculated is not within the range of simulated traffic volumes, it is recommended that the user train a new regression model with additional scenarios covering a traffic level equal to the calculated line capacity.

$$f = 0.0147 \tag{2.7}$$

$$g_t = -0.1016P^* - 1.0336U^* + \begin{cases} 1.1858 & t \in Type:P \\ 2.8260 \quad if & t \in Type:I \\ 1.0184 & t \in Type:U \end{cases} \tag{2.8}$$

$$h_t = -0.0228P^{*2} + 0.6778P^*U^* - 26.4025U^{*2} + \begin{cases} -173.9520 - 0.4414P^* + 33.7726U^* & t \in Type:P \\ -152.9316 + 0.4610P^* + 42.3191U^* \quad if & t \in Type:I \\ -86.6652 - 1.0184P^* + 1.3167U^* & t \in Type:U \end{cases} \tag{2.9}$$

$$2 \le P = P^* + \overline{P} \le 36,\ 6\% \le U = U^* + \overline{U} \le 94\% \tag{2.10}$$

The forms of the interaction in the g_t and h_t functions suggest the impact of additional passenger trains on delay is distributed disproportionally between train types. This impact is investigated and discussed in the next subsection using the line capacity evaluation model.

2.3.3. Incremental Impact of Passenger Trains on Line Capacity

The case study simulation results demonstrate that when passenger trains are added to lines with different existing freight traffic mixtures, the impact of the passenger trains is distributed disproportionally. For example, intermodal trains experience little additional delay when passenger trains are added to a line where the intermodal trains comprise the majority of freight traffic (Figure 2.6); however, on lines where bulk trains dominate, the added passenger trains have a greater impact on intermodal train delay.

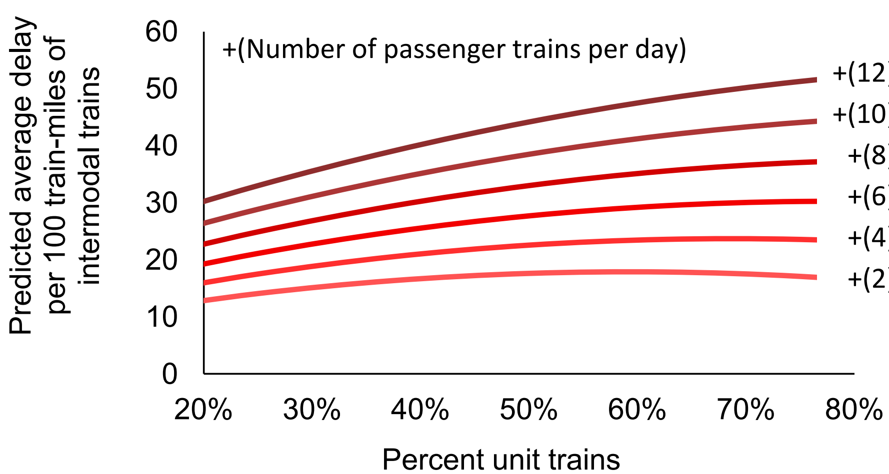


Figure 2.6. Example of disproportional impacts of passenger trains on intermodal train delay under different freight traffic mixtures (20 freight trains scenario)

To further illustrate the disproportional impact of additional passenger trains on different types of freight trains, a case was considered in which the maximum allowable delays are fixed at: 8, 20, and 60 minutes for passenger, intermodal and bulk unit trains, respectively. The line capacities defined by the LOS of each train type over a range of freight traffic mixtures changes when there are 0, 2, 6 and 8 passenger trains operating on the line (Figure 2.7). In a comparison of the graphs, the passenger capacity profile becomes more critical (moves downward) as the number of passenger trains increases, followed by the intermodal and unit train profiles. This finding implies that, in this case, added passenger trains impact the performance of other passenger trains the most, followed by intermodal and then bulk trains.

Moreover, the shape of the final capacity profile changes as the number of passenger trains changes (Figure 2.7). For the scenario with zero passenger trains per day, capacity increases with percent unit trains. When the number of passenger trains increases to more than two per day, the portion of the final capacity profile corresponding to a higher percent unit trains starts to decline when the percent unit trains increases. This finding implies that the freight traffic mixture (percent unit trains) with the lowest capacity changes as the number of passenger trains is increased. Thus, it is not just the volume of existing freight trains that is important when the ability of a line to support additional passenger traffic is assessed, but the exact mixture of freight trains operating on the line. This finding may help planners better predict potential congestion when new passenger service is proposed on different types of freight corridors.

The analysis demonstrates how additional passenger trains disproportionally reduce freight train capacity, depending on the initial freight traffic mixture. This incremental impact of passenger trains can be evaluated using an index called the Equivalent Freight Capacity Loss (EFCL). EFCL is calculated by dividing the total loss of freight capacity by the number of additional passenger trains added.

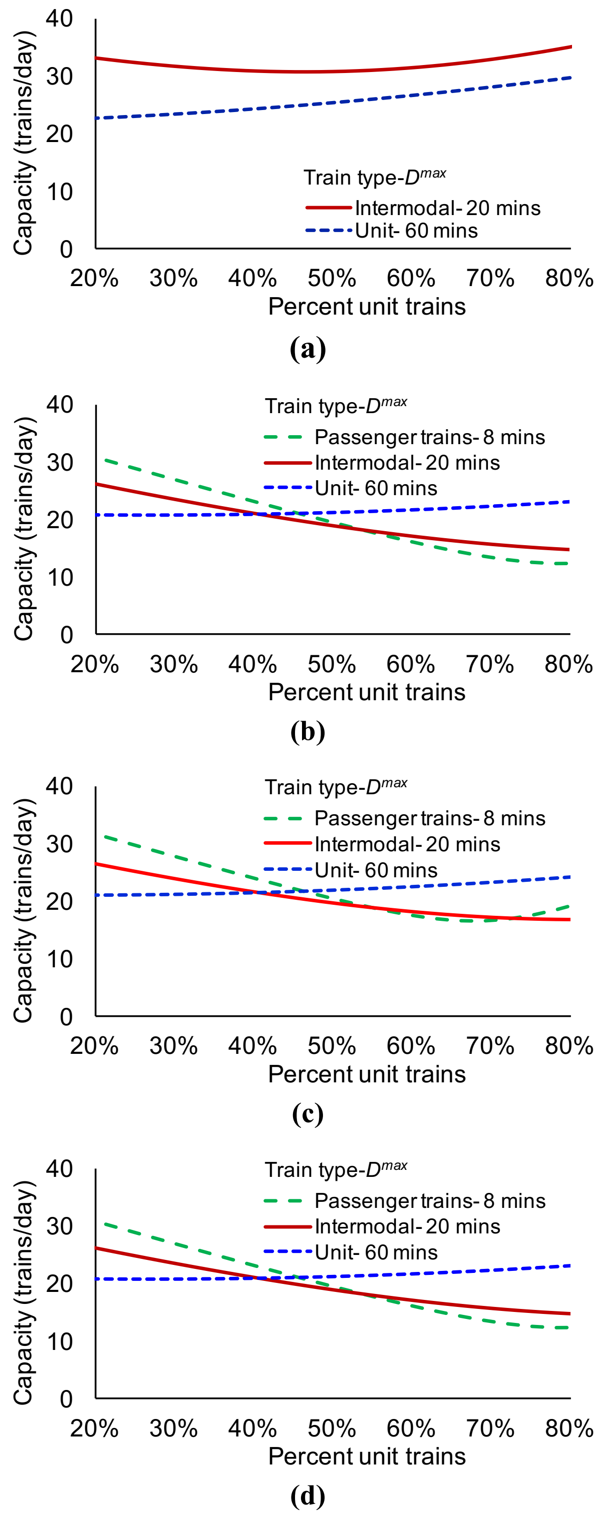


Figure 2.7. Final capacity profile under scenarios with (a) freight traffic only and, (b) 2 additional, (c) 6 additional, and (d) 8 additional passenger trains

For the combinations of freight and passenger traffic considered in this study, the region where a single passenger train has the greatest impact on capacity is between four and eight passenger trains per day with the initial freight traffic mixture greater than 80 percent unit trains (Figure 2.8a). This region corresponds to the most heterogeneous conditions on the line. The critical location is not at the point of highest percent unit trains and number of passenger trains because the ratio of passenger trains to total traffic increases when the number of passenger trains increases. For example, the capacity of a scenario with six passenger trains and 80 percent unit trains is approximately 17 trains per day and the capacity of a scenario with ten passenger trains and the same percentage of unit trains is approximately 12 trains per day (Figure 2.8b). The proportion of passenger trains in the first scenario is approximately 35 percent, and 83 percent in the second. Since most of the trains in the second scenario are passenger trains, the average traffic speed is higher and the interference from train type heterogeneity is lower compared to the first scenario.

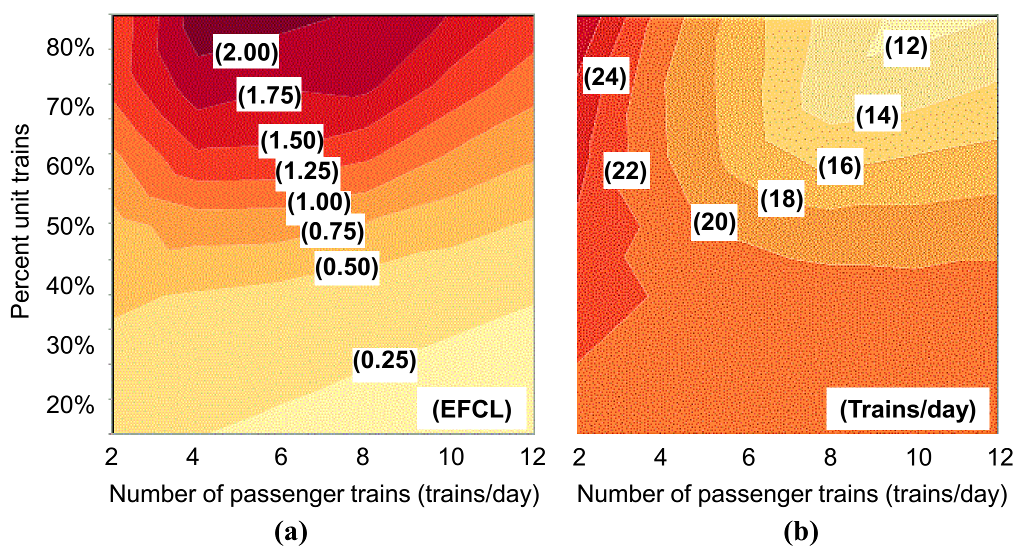


Figure 2.8. Variation of (a) Equivalent Freight Capacity Loss (EFCL) per passenger train (b) capacity contours under different traffic mixture

2.3.4. LOS Sensitivity Analysis

The LOS for each train type may change according to shipper demands, individual railway business objectives and the condition of the rail network. For example, lines connecting through a congested terminal may require a stricter LOS for certain trains to maintain the on-time performance of traffic. Since changing the LOS (maximum allowable average train delay) of a particular type of train alters its capacity surface, a change to one train type may cause changes in the final capacity surface.

To investigate the sensitivity of capacity to LOS, the change in line capacity caused by 10 percent improvements and relaxations in LOS of each train type were calculated (Figure 2.9). For positive and negative changes to the LOS of each train type, the contours plotted in the figure represent the resulting absolute change in the maximum number of trains per day over a range of traffic mixtures. For example, if the intermodal LOS must be improved by 10 percent (i.e. intermodal train delay decreased by 10 percent) under a traffic mixture includes four passenger trains and 50 percent unit trains, a capacity reduction of approximately two trains per day is required to achieve the new LOS.

Careful study of Figure 2.9 reveals a pattern in which relaxing LOS (increasing maximum allowable delay) of a certain train type only increases capacity under certain combinations of passenger and freight traffic mixtures. This result indicates that changes of LOS of a certain train type do not necessarily change line capacity. Changes in line capacity under certain traffic mixtures are usually related to the LOS of one or two train types even though there are three train types present on the case study corridor. If the delays of a certain train type under different traffic mixtures are generally much smaller than an improved or relaxed LOS, the

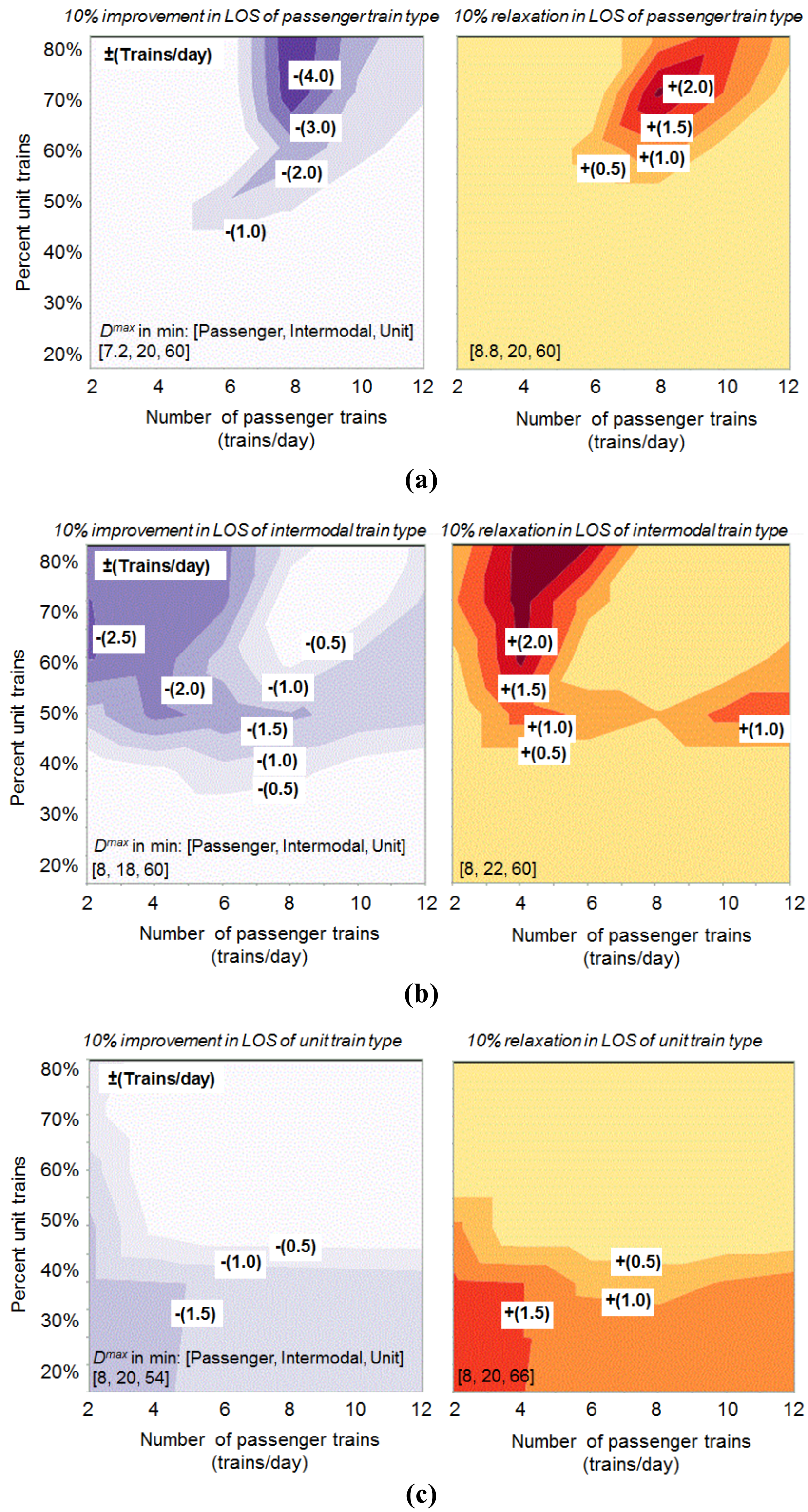


Figure 2.9. Changes in line capacity (trains per day) resulting from changes in maximum allowable delay for (a) passenger (b) intermodal and (c) bulk unit trains

capacity contours for that train type may not determine the line capacity under any traffic mixture. Conversely, zones of zero capacity exist when the LOS of a certain train type is very strict and the maximum allowable delay is set to be lower than what can be achieved given the traffic mixture and volume on the line.

2.3.5. Speed Homogeneity

Dingler et al. (2009) found that reducing the heterogeneity of train speed or priority increased line capacity. However, homogenizing train priority may not be appropriate because it can reduce service reliability of certain time-sensitive trains, or unduly increases the operating cost of less time-sensitive trains. Changing train speed has a relatively low impact on service reliability of time-sensitive trains as long as the minimum run time is satisfied. This subsection analyzes a scenario where train speeds are made more homogeneous to evaluate the effect on capacity.

To reduce heterogeneity, the maximum speed of passenger, intermodal and bulk unit trains are adjusted to 60, 55 and 50 mph, respectively, from the original 75, 55 and 35 mph. The benefit of reducing speed heterogeneity varies based on the initial freight traffic mixture (Figure 2.10).

When the percentage of bulk unit trains increases, the benefit from adjusting speeds becomes much more pronounced. This variation in capacity improvement under different freight mixtures suggests the relative impact of passenger trains on intermodal and bulk unit trains changes with speed. The impact of passenger trains on intermodal and bulk unit trains both decrease, but the decrease for intermodal trains is less than that for bulk unit trains.

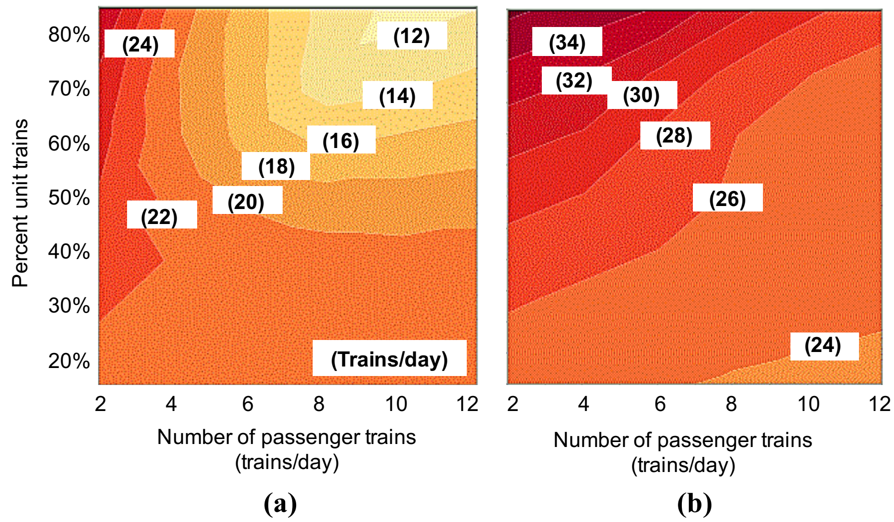


Figure 2.10. Comparison of capacity contours between (a) original case, and (b) case with more homogenous speed

The implication of this finding is that altering train speed may increase the minimum running time of some trains but can help accommodate more passenger and freight traffic while maintaining the LOS. Operational strategies related to altering train speed may be an option to temporarily increase capacity in order to recover from disruptions when the minimum running time of each train has already been exceeded.

2.4. Discussion and Conclusion

A capacity evaluation process is proposed in this chapter to evaluate line capacity under different traffic mixtures of trains with unique LOS requirements. An application of the process using traffic with three types of trains was demonstrated. Using the values of number of passenger trains per day, percent unit trains and LOS of each train type as input, the process develops a capacity profile for each individual train type. The final capacity profile is defined by the minimum value of all profiles. This process can be extended to lines with any combination of three or more train types if appropriate indices are used for the construction of a regression model with a polynomial structure.

The case study demonstrates the incremental impact of adding passenger trains to lines with mixtures of different types of freight trains. The capacity evaluation process can depict the incremental impact of one train type on the other train types and the overall capacity of the line. In general, the addition of a priority passenger train has a disproportionate impact on train types. For example, on a freight rail line dominated by bulk unit trains, intermodal trains are the most negatively affected by the addition of passenger trains since the intermodal trains must relinquish a preferred schedule spot for use by a priority passenger train. Despite being in the majority, bulk trains sustain relatively little impact, even though they exhibit a greater speed differential compared with passenger trains. Instead of only looking at average delay across all train types, practitioners can use this process to identify the impact on other types of trains as a result of adding trains. The changes in the final capacity profiles also show that both the volume and mixture of existing trains are important when assessing the ability of a line to support additional traffic.

The sensitivity analysis of capacity to LOS illustrates how the capacity benefit of relaxing the allowable delay for certain train types varies according to the freight traffic mixture on the line. Increases in allowable delay (relaxing LOS) for a particular type of train does not always increase line capacity and decreasing delay (improving LOS) of a train type does not always reduce line capacity. Finding the critical train types of a mainline under certain traffic mixtures is necessary before using LOS relaxation as the strategy to gain additional capacity. Increases in allowable delay for a particular type of freight train only tend to increase capacity when there are few passenger trains and that particular type of freight train represents a minority of freight traffic. The case study findings suggest that the LOS of the minority freight train type plays a key role in establishing the capacity of a line with three or more types of trains.

The speed homogeneity portion of the case study demonstrated that reducing speed heterogeneity can enhance capacity and reduce the incremental impact of additional passenger trains on line capacity. Since minimum running times must still be met, harmonizing operating speeds could be a method to add "temporary capacity" to recover from disruptions.

Besides its use in evaluating line capacity, the types of train delay prediction models developed as part of the presented capacity evaluation process have additional utility in comparing the delay performance of different infrastructure and operating scenarios for a given traffic volume and mixture. An application of a train delay prediction model for measuring the performance of different capacity expansion alternatives is presented in the next chapter.

CHAPTER 3: COMPARING CAPACITY EXPANSION STRATEGIES FOR SINGLE-TRACK LINES UNDER MIXED OR FLEXIBLE OPERATION

An earlier version of this research appears in:

Shih, M.C., C.T. Dick, S. Sogin, and C.P.L. Barkan. 2014a. Comparison of capacity expansion strategies for single-track railway lines with sparse sidings. Transportation Research Record: Journal of the Transportation Research Board, 2448: 53-61.

As mentioned in the introduction, North American railways have a strong business incentive to properly match line capacity to traffic demand. A poorly located capacity expansion project is an unwise investment that will do little to meet future traffic demand. Since single-track lines provide less capacity than multiple-track lines, they are more likely to become bottlenecks on a rail network with growing traffic. In this chapter I investigate several approaches for increasing the capacity of single-track lines with sparse siding spacing under mixed or flexible operations in order to develop greater understanding of the efficacy of different strategies.

3.1. Overview of the Current Status

In general, rail capacity can be improved through changes in operational strategy or improvements to the infrastructure (Dingler, 2009; Dingler et al., 2013; Lai and Shih, 2013). Changes in operational strategies tend to have lower capital cost and can be implemented more quickly than infrastructure investment but may not be adequate to accommodate sustained growth in traffic. Given the projected increase in long-term demand for rail capacity, both infrastructure and operational strategies are needed.

North American railroads have been making infrastructure investments to increase capacity for over 15 years. On high volume segments of their core network, this has often been in the form of adding a second main track. However, growth of freight traffic and recent changes in commodity flows have changed traffic patterns, resulting in growth on lines with historically lower traffic density. Such lines may have fewer passing sidings of sufficient length to handle modern unit trains. To meet demand on these lines with widely spaced passing sidings, railroads have shifted capital to projects that increase their capacity. For example, in 2011 and 2012, Canadian Pacific invested $97 million to renew and improve its network in the Bakken region of North Dakota to provide better service to the energy industry (Wanek-Libman, 2013). BNSF Railway initiated several siding projects related to energy industry development in 2012 (BNSF, 2012) along with more recent infrastructure projects to improve their network capacity (BNSF, 2016). Canadian National spent $68 million in 2013 to upgrade two of its branches in Wisconsin to accommodate growth in transportation demand for hydraulic fracturing sand (Wanek-Libman, 2013). Besides these examples, additional projects are in the planning and engineering stages. Because of the large capital investment required, understanding the relationship between infrastructure improvement and capacity increase on single-track lines with sparse sidings will help the railroads plan a more effective and efficient capacity expansion strategy.

A number of previous studies have investigated the effectiveness of infrastructure improvements for increasing line capacity. European studies tend to focus on passenger rail corridors (Fransoo and Bertranda, 2000; Lindfeldt, 2007; 2009; 2012a) while this chapter focuses on freight or shared-use rail corridors. Petersen and Taylor (1987) used simulation analysis to determine longer siding locations to improve the efficiency of passenger train operation on freight lines. An analytical model was proposed by Pawar (2011) to determine the appropriate length of long sidings for train meets. Both studies focused on a specific type of capacity expansion alternative. Lindfeldt (2012b) used an analytical approach to find feasible strategies to increase capacity through incremental infrastructure projects, but he only analyzed a particular real-world line with specific existing characteristics. Sogin (2013) and Sogin et al. (2013b) used simulation methods to study the relationship between the general length of second main track and train delay. Their studies were more general but did not cover the transition from single-track lines with sparse sidings to single-track lines with dense siding-spacing. Single-track lines with sparse sidings are common in North America and, as described above, have been the subject of recent and planned infrastructure investment. A study investigating and comparing capacity expansion strategies for single-track lines with sparse sidings will enable railway practitioners to make better-informed capital investment decisions.

3.2. Approach to Evaluating Performance of Alternative Expansion Strategies

In addition to providing insight specific to expanding the capacity of single-track lines with sparse sidings, this study formalizes a more general evaluation process to assess and compare the performance of potential railway line capacity expansion alternatives (Figure 3.1). As introduced in the following sections, the major steps in this general performance evaluation process are identification of infrastructure and traffic scenarios, experimental design, simulation, regression and performance analysis.

3.2.1. Infrastructure and Traffic Scenarios

To begin the evaluation process, a railway practitioner must identify a number of proposed capacity expansion project alternatives or alternative strategies for consideration. In this study, a capacity expansion *project alternative* is used to describe one option for construction of additional mainline or siding track at a single location on a mainline corridor. An *alternative strategy* for capacity expansion refers to a unique series of capacity expansion projects to be construction. Although the process can consider any number of project alternatives or alternative strategies, considering a large number of project alternatives or evaluating many individual steps in several alternative strategies requires a large number of simulation experiments. Screening alternatives using general capacity insights from this dissertation and other published research can reduce the required simulation effort.

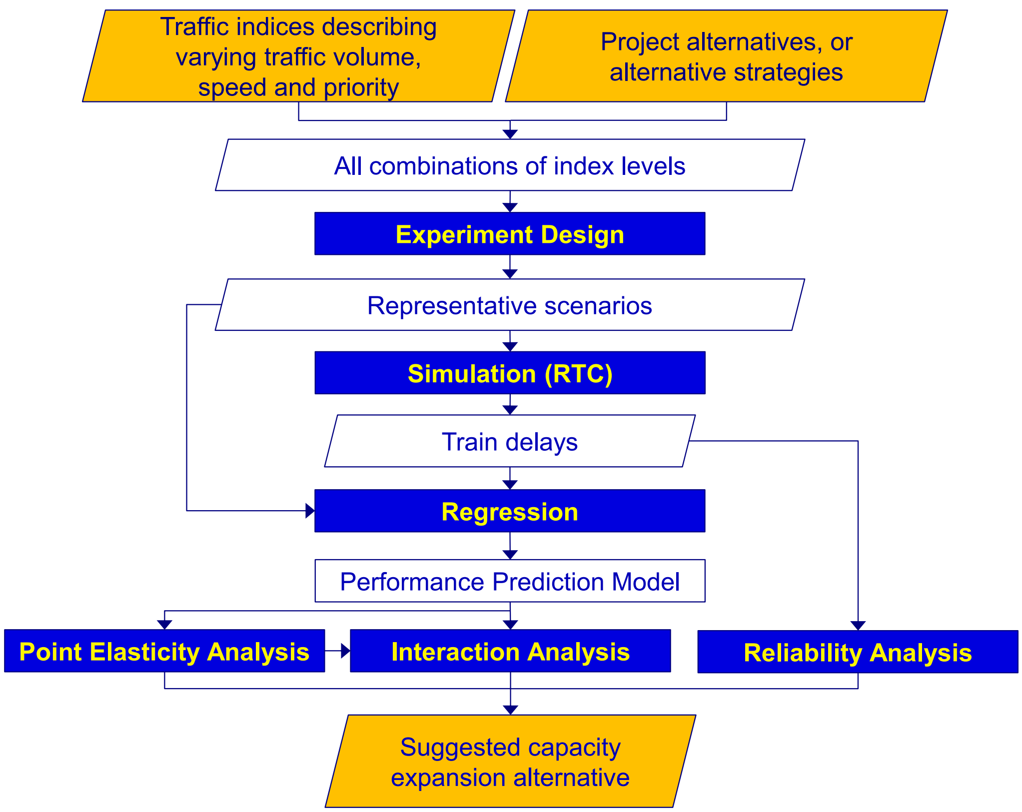


Figure 3.1. Flowchart of the strategy comparison process

In selecting railway traffic scenarios, practitioners are often interested in the performance of alternatives under future traffic growth beyond current conditions. Practitioners may have specific traffic forecasts and growth factors or, in the case of shared-corridors, they may be more interested in future traffic mixtures that are quite different from current operations. The latter case requires a more involved approach to construct traffic scenarios that will be representative of possible shared-corridor operating conditions.

Past studies have used various approaches to construct representative shared-corridor rail traffic. Among them, Krueger (1999), Gorman (2009), and Sogin et al. (2013b) identified several important traffic factors that affect the capacity of a single-track line. On the basis of these studies, traffic scenarios were selected to include a range of traffic volume (TV), maximum speed of freight trains (MFS), maximum speed of passenger trains (MPS), and traffic mixture. Traffic volume is defined as the total number of trains traversing the study route per day. The maximum speed values for freight and passenger trains are the highest authorized track speed for each group of trains under free-flow conditions. The actual traveling speed may often be constrained below these values because of the acceleration and braking required for different stopping patterns and to negotiate turnouts, the number and power of the locomotives assigned to the trains, and interference between train types. Traffic mixture is expressed as the percentage of the total number of trains that are freight trains (percent FT) (Sogin et al., 2013b). Varying the percent FT changes the level of interference caused by differences between train types; this process allows a capacity expansion study to consider both scenarios that are dominated by freight traffic and those dominated by passenger traffic.

3.2.2. Experiment Design and Simulation

A number of simulation scenarios are needed to develop a delay response surface across a range of traffic and infrastructure conditions. Potential traffic scenarios with different combinations of volume, speed, and train types, together with the proposed capacity expansion alternatives, are used to develop a matrix of simulation experiments using partial factorial design techniques. Partial factorial design selects a representative subset of simulations from a full factorial design to reduce the number of simulation runs required (Montgomery, 1984; Box and Soren, 1987). As described in Chapter 2, the delay response of this subset is similar to the original experiment but fewer scenarios need to be simulated.

RTC software was used to simulate the partial factorial design scenarios and determine the average train delay response. For a single scenario in an experiment design, the train delay output is the average train delay across all trains for multiple days of operation that are replicated several times with different flexible train departures. In practice, any simulation model, analytical or parametric approach can be used to determine the average train delay for each scenario in the experiment design.

3.2.3. Performance Prediction Regression Model

After conducting the simulation experiments with RTC software, regression on the simulation results is used to construct a performance (train delay) prediction model with the form of Equation 3.1.

$$D = f(Q,\ U,\ P,\ F, A_n, S) \tag{3.1}$$

Where:

D: Predicted average train delay per 100 train-miles of the index train type. The index train type in this study is the freight train since it is more sensitive to changes in infrastructure properties or traffic characteristics, and is usually the concern of the infrastructure owner on the shared-corridor. Users can define their index train type or use the average delay of all trains depending on their objectives.

Q: Total traffic volume (TV, trains/day).

U: Number of freight trains as a percentage of total traffic volume (percent FT, %)

P: Maximum allowable speed of passenger train type (MPS, mph).

F: Maximum allowable speed of freight train type (MFS, mph).

A_n: A set of binary terms, each representing a different alternative n.

S: The total length of second track (tracks other than the original main track, including passing sidings) as a percentage of the total route length under study (percent ST, %).

When there are two or more train types, the total traffic volume and percentage of each train type can be used as indices. Equation 3.2 below shows the more generalized form of the train delay performance prediction model for scenarios with more than three train types.

$$D = f(Q,\ U_T,\ V_T, A_n, S) \tag{3.2}$$

Where:

U_T: Set of numbers of train types as percentages of total traffic volume (%).

V_T: Set of maximum allowable speeds for train types (mph).

3.2.4. Performance Analysis

The simulation outputs and performance prediction model support an evaluation of each capacity expansion alternative through three different analyses:

- The *point elasticity analysis* examines the sensitivity of the train delay response for each infrastructure alternative to changes in traffic characteristics and for alternative strategies. This calculation of the point elasticities uses the performance prediction model developed through regression. More sensitive traffic characteristics are selected for interaction analysis. Additionally, when the delays of any alternatives are close, the results of this analysis also provide information on the robustness of the infrastructure alternative to the assumed future traffic conditions. Alternatives whose train delay response is less sensitive to changes in the traffic variables may yield a more consistent return on investment if future traffic conditions are uncertain.
- The *interaction analysis* uses the performance prediction model to compare the average train delay between alternatives for specific combinations of the traffic variables selected by point elasticity analysis. Alternatives with the lowest average train delay response for equivalent infrastructure investment will yield the greatest return.
- The *reliability analysis* uses the simulation output to directly construct a distribution of train delay for each scenario under a specific combination of traffic variables. In examining this distribution, the reliability analysis considers the best and worst-performing trains, not just the average. Alternatives that yield narrow train delay distributions will have more consistent and reliable performance.

The point elasticity, interaction and reliability analyses are described in more detail and demonstrated through the case study introduced in the following section. In recommending a preferred capacity expansion alternative or strategy, a practitioner may consider the results of any, or all three of the different analyses, depending on what is most important to their overall objectives.

3.3. Case Study

The case of capacity expansion strategies for single-track lines with sparse sidings is used to demonstrate the general evaluation process for assessing and comparing the performance of potential railway line capacity expansion alternatives. In this study, four potential capacity expansion alternatives for single-track lines with sparse sidings were identified on the basis of previous academic studies and industry suggestions. The capacity expansion alternatives evaluated are part of the larger transition process from a single-track line with sparse sidings to a full two-main-track line (Figure 3.2). The dashed lines (black) indicate the transition process from single-track line with dense sidings to a two-main-track line, previously studied by Sogin et al. (2013a), Atanassov et al. (2014), Atanassov (2015), and Atanassov and Dick (2015). The solid arrows (blue) are the focus of this study and the bold labels (blue) beside the arrows in Figure 3.2 indicate the alternatives evaluated.

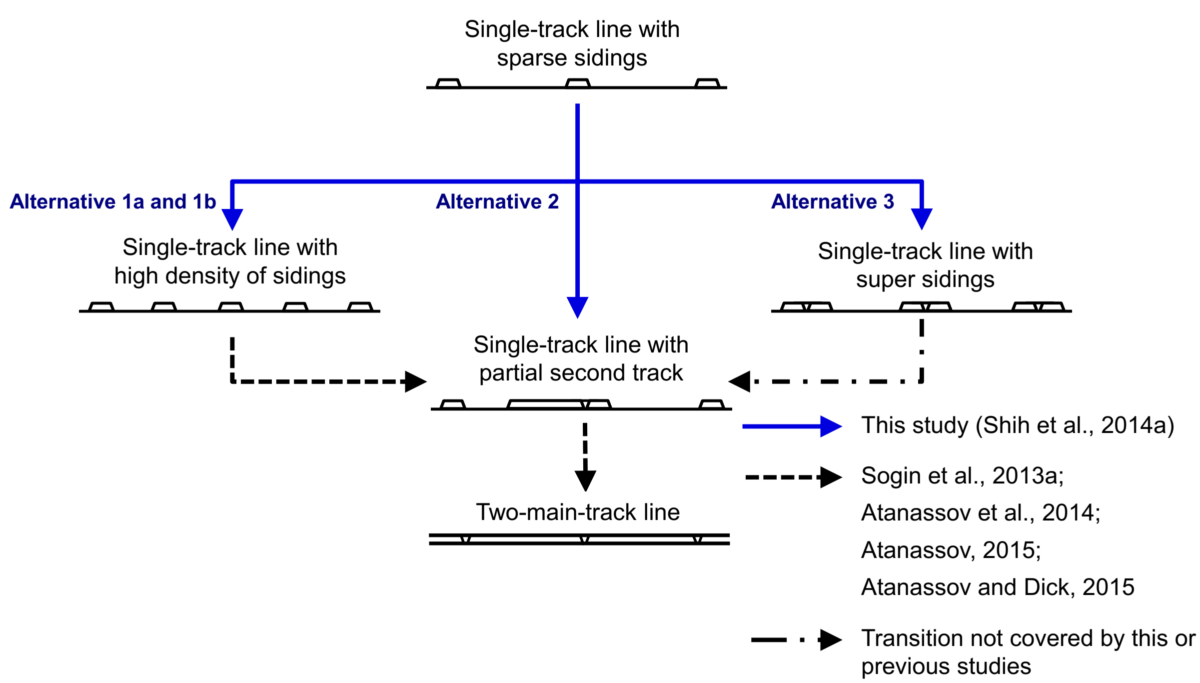


Figure 3.2. Flowchart of the capacity evaluation process

3.3.1. Infrastructure Parameters

All of the proposed capacity expansion alternatives start with the same baseline route infrastructure representative of a single-track line with sparse passing sidings (Table 3.1). Each proposed capacity expansion alternative differs in the type and location of added track infrastructure (Figure 3.3). Alternatives 1a and b both involve construction of new sidings between current passing siding locations to create denser siding spacing on the single-track line.

Table 3.1. Route parameters for simulation model

Parameter	Value
Total length of the line	240 miles
Initial siding spacing	20 miles
Initial percent two-main-track	10%
Average signal spacing	2 miles
Diverging turnout speed	45 mph
Traffic control system	2-block, 3-aspect CTC

In Alternative 1a, construction of new sidings begins at the middle of the corridor and moves outward toward the ends and is called "center out" (Figure 3.3a). In Alternative 1b, new sidings are evenly distributed along the corridor at each stage of construction and is called "spread evenly" (Figure 3.3b).

In Alternative 2, existing sidings are connected by an additional track to form an increasingly longer section of second main track and is called the "second-track" alternative (Figure 3.3c). This approach was selected on the basis of past research by Lindfeldt (2012b) who found that continuous double-track sections were the most effective approach to increase capacity.

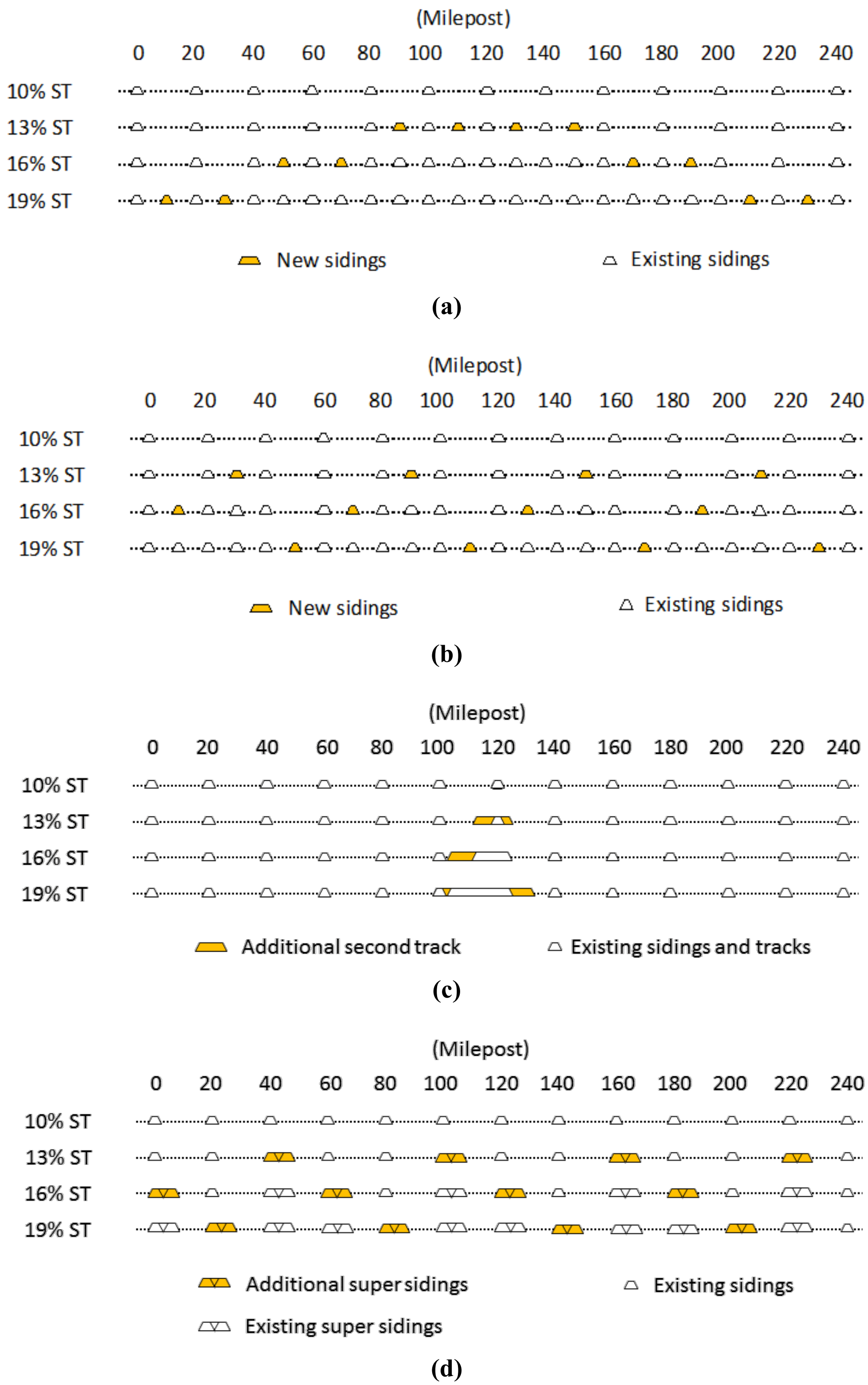


Figure 3.3. Capacity expansion strategies for single-track lines with sparse sidings, Alternative (a) 1a, center out (b) 1b, spread evenly, (c) 2, second-track (d) 3, super siding

In Alternative 3, "super sidings" are created by doubling the length of existing sidings and installing a universal crossover at the new midpoint. This strategy is used by some major North American railroads (CN, 2005).

For each alternative, three scenarios with different lengths of second track were constructed to represent the incremental process of capacity expansion. The unit used to quantify the length of second track is the total length as a percentage of total corridor length. Each different percentage of second track (percent ST) is intended to approximate a consistent level of capital investment across the different alternatives. In the experiment, each alternative received ST equal to 13 percent, 16 percent, and 19 percent, emulating the incremental infrastructure investment process above the base level of 10 percent ST.

3.3.2. Traffic Parameters

To focus on the interaction between passenger and freight trains on the capacity expansion alternatives, the case study traffic contains two train types: passenger and freight trains. This combination of traffic also helps maintain consistency between the experimental setting of this study and previous related studies (Sogin et al., 2013a; Sogin et al., 2013b) to facilitate comparison of the results.

The two sets of train parameters (Table 3.2) are similar to those used in Chapter 2, with the exception of the maximum allowable speed for each train type as it is a variable in the experiment design. Additionally, the ideal total running time of each train type to traverse the entire mainline, and the time to travel between adjacent sidings, are a function of the maximum train speed. For purposes of generality, there are no civil or other operating speed restrictions on the route considered in the case study.

Table 3.2. Train parameters for simulation model

Parameter	Freight trains	Passenger trains
Locomotives	3 EMD SD70	2 GE P42
Number of Cars	115 hopper cars	7 articulated Talgo cars
Length (ft)	6,325	500
Weight (tons)	16,445	800
Horsepower per total ton (hp)	0.78	15.4
Scheduled stops	none	30 miles station spacing
Ideal total running time (hr)	6.4- 9.6	3.4- 4.1
Ideal running time between adjacent sidings (hr)	0.4- 0.8	0.2- 0.3

In addition to the train characteristics in Table 3.2, the scheduled departure pattern may also affect line capacity. In this study, rail traffic follows flexible operation but the same method can also be applied to traffic under mixed operation. Under the completely flexible operations considered in this study, train departure time is determined using a random, uniform distribution over a 24-hour period. By considering different random departure times for each simulated day of train operations, RTC can determine average train performance over a range of possible schedule scenarios.

3.3.3. Experiment Design and RTC Simulation

In addition to the four infrastructure alternatives (Table 3.3a), five different factors with three index levels are considered by the partial factorial design (Table 3.3b). The highest TV tested in this study is 24 trains per day due to the limited capacity of the initial single-track line with sparse sidings. Higher traffic volumes lead to failed RTC simulation runs and a lack of valid simulation results for inclusion in the response surface. The value of percent FT ranges from 25 to 75 percent to capture the effect of heterogeneous traffic.

The percent ST starts at 13 percent (the base scenario with four additional sidings) instead of 10 percent (the base scenario) because all of the alternatives start with the same single-

Table 3.3. Index levels used in the experiment (a) categorical factor (b) numeric factors

(a)

Categorical factor	Category			
Alternative strategy	1a	1b	2	3

(b)

Numeric factor	Unit	Value Low	Medium	High
Percent second track (percent ST)	%	13	16	19
Traffic volume (TV)	trains/day (TPD)	8	16	24
Percent freight train (percent FT)	%	25	50	75
Maximum passenger speed (MPS)	mph	79	95	110
Maximum freight speed (MFS)	mph	30	40	50

track configuration with 10 percent ST. The high level of 19 percent ST reflects the scenario with the maximum number of sidings or the equivalent length of second track (second-track and super siding alternatives).

The partial factorial experiment matrix contains 162 scenarios (compared with 972 in the full factorial design). Each scenario was simulated using RTC for five days of train operations and replicated six times to yield train performance data for a total of 30 days of operation. The repetition of each scenario generates traffic data from different randomized schedules and also helps ensure that there is at least one feasible output from RTC for each scenario. The train delay response for each scenario in the partial factorial design (Table 3.4) is calculated as the average train delay across all trains over all 30 simulated days of operations.

3.3.4. Regression Approach and the Performance Prediction Model

The regression approach uses the calculated train delay response from each random seed for all 162 traffic scenarios in the experiment matrix to construct a model to predict the average train delay for each alternative under different traffic conditions. There are 25,610 data points,

Table 3.4. Representative scenarios in the experiment matrix

Scenario	TV (trains/day)	Percent FT (%)	MPS (mph)	MFS (mph)	Alternative	Percent ST (%)	Average freight train delay per 100 train-miles (min)
1	8	25	79	30	1a	13	37.3
2	8	25	79	30	1a	19	20.9
3	8	25	79	30	1b	13	40.0
4	8	25	79	30	2	16	46.3
5	8	25	79	30	2	19	45.6
6	8	25	79	30	3	13	41.5
7	8	25	79	30	3	19	27.7
8	8	25	79	50	1a	13	36.9
9	8	25	79	50	1a	19	11.9
10	8	25	79	50	1b	13	16.9
11	8	25	79	50	1b	19	11.9
12	8	25	79	50	2	13	22.7
13	8	25	79	50	2	19	19.9
14	8	25	79	50	3	13	18.0
15	8	25	79	50	3	19	15.3
16	8	25	95	40	1a	13	20.0
17	8	25	95	40	1a	19	13.2
18	8	25	95	40	1b	13	21.3
19	8	25	95	40	1b	19	13.2
20	8	25	95	40	2	13	22.3
21	8	25	95	40	3	13	24.3
22	8	25	95	50	1a	13	15.3
23	8	25	95	50	1a	19	13.4
24	8	25	95	50	1b	13	14.4
25	8	25	95	50	2	19	18.6
26	8	25	95	50	3	19	15.6
27	8	25	110	30	1a	13	36.9
28	8	25	110	30	1a	19	25.5
29	8	25	110	30	1b	13	38.1
30	8	25	110	30	1b	19	25.5
31	8	25	110	30	2	13	53.2
32	8	25	110	30	2	19	43.7
33	8	25	110	30	3	13	44.0
34	8	25	110	30	3	19	28.8
35	8	25	110	40	2	13	19.6
36	8	25	110	40	3	19	15.9
37	8	25	110	50	1a	19	11.3
38	8	25	110	50	1b	13	14.6
39	8	25	110	50	1b	19	11.3
40	8	25	110	50	2	13	18.1
41	8	25	110	50	2	19	13.8

Table 3.4 (cont.)

Scenario	TV (trains/day)	Percent FT (%)	MPS (mph)	MFS (mph)	Alternative	Percent ST (%)	Average freight train delay per 100 train-miles (min)
42	8	25	110	50	3	16	16.1
43	8	50	79	30	3	19	27.3
44	8	75	79	30	1a	13	27.6
45	8	75	79	30	1a	19	15.0
46	8	75	79	30	1b	19	15.0
47	8	75	79	30	2	13	24.0
48	8	75	79	30	2	19	24.1
49	8	75	79	30	3	13	23.2
50	8	75	79	40	1a	19	9.2
51	8	75	79	40	1b	13	13.0
52	8	75	79	40	2	13	13.0
53	8	75	79	40	3	13	13.2
54	8	75	79	50	1a	13	12.8
55	8	75	79	50	1a	19	7.4
56	8	75	79	50	1b	13	9.5
57	8	75	79	50	1b	19	7.4
58	8	75	79	50	2	16	10.6
59	8	75	79	50	3	13	10.6
60	8	75	79	50	3	19	10.0
61	8	75	95	30	1b	13	28.2
62	8	75	95	30	1b	19	15.7
63	8	75	95	30	2	13	25.2
64	8	75	95	50	1a	13	10.0
65	8	75	110	30	1a	13	27.2
66	8	75	110	30	1a	19	16.1
67	8	75	110	30	1b	13	19.9
68	8	75	110	30	2	19	25.3
69	8	75	110	30	3	13	24.1
70	8	75	110	30	3	19	18.3
71	8	75	110	40	1b	19	10.5
72	8	75	110	40	2	19	12.6
73	8	75	110	40	3	13	13.4
74	8	75	110	50	1a	13	9.7
75	8	75	110	50	1a	19	7.2
76	8	75	110	50	1b	19	7.2
77	8	75	110	50	2	13	9.6
78	8	75	110	50	2	19	8.8
79	8	75	110	50	3	13	9.8
80	8	75	110	50	3	19	8.9
81	16	25	79	30	1a	19	43.5
82	16	25	79	30	1b	19	43.5

Table 3.4 (cont.)

Scenario	TV (trains/day)	Percent FT (%)	MPS (mph)	MFS (mph)	Alternative	Percent ST (%)	Average freight train delay per 100 train-miles (min)
83	16	25	79	50	1b	13	87.4
84	16	50	95	40	1a	16	34.0
85	16	50	95	40	1b	16	30.7
86	16	50	95	40	2	16	43.6
87	16	50	95	40	3	16	38.6
88	16	50	110	50	2	16	32.5
89	16	75	79	30	1b	13	62.8
90	16	75	110	30	1b	19	35.5
91	16	75	110	50	1b	13	24.3
92	24	25	79	30	1a	13	180.1
93	24	25	79	30	1a	19	88.1
94	24	25	79	30	1b	13	183.0
95	24	25	79	30	2	13	170.6
96	24	25	79	30	2	19	235.0
97	24	25	79	30	3	19	122.0
98	24	25	79	40	1a	13	97.8
99	24	25	79	40	1b	13	91.7
100	24	25	79	40	1b	19	49.4
101	24	25	79	40	2	13	110.5
102	24	25	79	40	2	19	108.0
103	24	25	79	50	1a	13	69.3
104	24	25	79	50	1a	19	43.3
105	24	25	79	50	1b	19	42.2
106	24	25	79	50	2	16	87.5
107	24	25	79	50	3	13	70.3
108	24	25	95	30	1b	13	161.0
109	24	25	95	30	3	13	182.9
110	24	25	95	50	1b	19	37.7
111	24	25	95	50	3	13	68.6
112	24	25	110	30	1a	13	174.2
113	24	25	110	30	1a	19	87.7
114	24	25	110	30	1b	13	168.7
115	24	25	110	30	1b	19	87.7
116	24	25	110	30	2	13	167.8
117	24	25	110	30	2	19	161.8
118	24	25	110	30	3	16	209.9
119	24	25	110	40	1a	13	95.1
120	24	25	110	40	1b	13	84.2
121	24	25	110	40	3	19	77.7
122	24	25	110	50	1a	13	62.3
123	24	25	110	50	1a	19	40.1

Table 3.4 (cont.)

Scenario	TV (trains/day)	Percent FT (%)	MPS (mph)	MFS (mph)	Alternative	Percent ST (%)	Average freight train delay per 100 train-miles (min)
124	24	25	110	50	1b	13	69.1
125	24	25	110	50	1b	19	40.1
126	24	25	110	50	2	13	73.6
127	24	25	110	50	2	19	64.2
128	24	25	110	50	3	13	74.4
129	24	25	110	50	3	19	55.3
130	24	50	79	30	3	13	153.5
131	24	50	79	50	3	19	51.4
132	24	75	79	30	1a	19	66.9
133	24	75	79	30	1b	19	66.9
134	24	75	79	30	3	19	110.1
135	24	75	79	40	2	19	91.2
136	24	75	79	40	3	13	92.1
137	24	75	79	50	1a	13	48.1
138	24	75	79	50	1a	19	26.3
139	24	75	79	50	1b	13	48.3
140	24	75	79	50	1b	19	26.3
141	24	75	79	50	2	13	58.8
142	24	75	79	50	2	19	53.1
143	24	75	79	50	3	13	55.5
144	24	75	95	30	1a	13	137.7
145	24	75	95	30	1a	19	79.9
146	24	75	95	30	3	19	106.8
147	24	75	95	40	1a	19	35.3
148	24	75	95	40	1b	19	35.3
149	24	75	95	40	2	19	68.0
150	24	75	95	40	3	19	58.9
151	24	75	95	50	1b	19	28.4
152	24	75	95	50	2	13	64.0
153	24	75	110	30	1b	13	147.2
154	24	75	110	40	1a	13	73.0
155	24	75	110	40	3	19	63.7
156	24	75	110	50	1a	13	53.8
157	24	75	110	50	1a	19	31.3
158	24	75	110	50	1b	19	31.3
159	24	75	110	50	2	13	59.4
160	24	75	110	50	2	19	52.2
161	24	75	110	50	3	16	57.3
162	24	75	110	50	3	19	43.0

composed of 7,058 train delay points for Alternative 1a, 7,236 for Alternative 1b, 5,280 for Alternative 2, and 6,036 for Alternative 3.

The first order, second order, and interaction terms of each numeric index, together with the cross terms between each of the mentioned terms, and each alternative, were used to capture the average train delay response. These terms were used to approximate the delay response surface (Box and Wilson, 1951) for each alternative. Similar to the train delay prediction model in Chapter 2, the stepwise regression approach based on BIC ratio (Burnham and Anderson, 2003; SAS Institute, 2017) was used for model selection (Table 3.5). BIC ratio was used since it penalizes complex models more than the other frequently used ratios (Burnham and Anderson, 2004). The F-test result of the model is significant and the model has an R-squared value of 0.93. The model is meaningful and precise enough to capture the relationship between train delay, infrastructure alternative and traffic characteristics.

The following symbols and terms describe the form of the model in Table 3.5:

- The Q, U, P, F, and S terms are indices of traffic volume (TV), percent freight train (percent FT), maximum passenger train speed (MPS), maximum freight train speed (MFS), and percent second track (percent ST). Their mean values are represented by the $\bar{Q}$, $\bar{U}$, $\bar{P}$, $\bar{F}$, and $\bar{S}$ terms.
- Similar to Chapter 2, the continuous variables were centered based on the mean values before regression. This helps reduce collinearity (SAS Institute, 2017).
- The capacity alternative A_n is split into four binary terms: A_{1a}, A_{1b}, A_2, and A_3. They represent the identified capacity expansion alternatives in this study.

Table 3.5. Statistics of performance prediction model and the selected terms

Term	Coefficient	t Ratio	p-value of t-test
Intercept	35.8765	35.93	<0.0001
$(Q-\bar{Q})$	4.2254	70.49	<0.0001
$(U-\bar{U})$	-33.0494	-25.93	<0.0001
$(P-\bar{P})$	0.0132	0.63	0.5281
$(F-\bar{F})$	-2.0060	-47.38	<0.0001
$(S-\bar{S})$	-334.5839	-31.78	<0.0001
A_{1a}	*-8.2774*	-17.06	<0.0001
A_{1b}	-7.4193	-15.08	<0.0001
A_2	12.0635	21.7	<0.0001
A_3	3.6332	7.13	<0.0001
$(Q-\bar{Q})^2$	0.1264	7.99	<0.0001
$(F-\bar{F})^2$	0.1235	17.77	<0.0001
$(U-\bar{U})\times(P-\bar{P})$	0.4839	5.41	<0.0001
$(Q-\bar{Q})\times(F-\bar{F})$	-0.1965	-40.35	<0.0001
$(Q-\bar{Q})\times(S-\bar{S})$	-29.2949	-21.18	<0.0001
$(F-\bar{F})\times(S-\bar{S})$	19.0625	15.33	<0.0001
$(Q-\bar{Q})\times A_{1a}$	-0.8024	-12.33	<0.0001
$(Q-\bar{Q})\times A_{1b}$	-0.7499	-10.75	<0.0001
$(Q-\bar{Q})\times A_2$	1.0999	14.20	<0.0001
$(Q-\bar{Q})\times A_3$	0.4523	7.60	<0.0001
$(F-\bar{F})\times A_{1a}$	0.4375	9.50	<0.0001
$(F-\bar{F})\times A_{1b}$	0.5451	-9.85	<0.0001
$(F-\bar{F})\times A_2$	-0.7349	-3.97	<0.0001
$(F-\bar{F})\times A_3$	-0.2478	-21.18	<0.0001
$(S-\bar{S})\times A_{1a}$	-105.1697	-6.19	<0.0001
$(S-\bar{S})\times A_{1b}$	-119.1163	-6.97	<0.0001
$(S-\bar{S})\times A_2$	211.3640	10.93	<0.0001
$(S-\bar{S})\times A_3$	12.9220	0.69	0.4872
R-squared value:	0.9324	p-value of F-test:	<0.0001

Where: $\bar{Q}$=17.4388, $\bar{U}$= 0.5861 (58.61%), $\bar{P}$ = *94.3390*
$\bar{F}$= 41.9430, $\bar{S}$= 0.1635

- The $(Q-\bar{Q})$ to A_3 are the first order terms. The $(F-\bar{F})^2$ and $(Q-\bar{Q})^2$ terms are the only second-order polynomial terms.
- The alternatives are shown by four binary terms, A_{1a}, A_{1b}, A_2 and A_3, to represent the Alternative 1a (center out), 1b (spread evenly), 2 (second-track), and 3 (super siding)

strategies. The binary terms allow a single model to predict the average train delay for a specific alternative strategy.

- The interaction terms are indicated by a multiplication sign "×" between the different first order terms. The $(U-\bar{U})\times(P-\bar{P})$ to $(S-\bar{S})\times A_3$ terms are the interaction terms.

The results of t-test indicate that most of the terms selected are statistically significant with the exception of the first order term $(P-\bar{P})$ and interaction term $(S-\bar{S})\times A_3$. These two terms were kept in the model for the following reasons:

- The $(P-\bar{P})$ term is retained because the interaction term $(U-\bar{U})\times(P-\bar{P})$ was selected. The $(U-\bar{U})\times(P-\bar{P})$ term is important because it has a clear physical meaning: the impact of passenger train speed on freight train delay is greater when the traffic mixture contains more freight trains. To keep the interaction term, both of the first order terms in the interaction term must be kept in the model.
- The $(S-\bar{S})\times A_3$ term is retained due to the presence of other interaction terms comprised of the $(S-\bar{S})$ term and the other alternatives. The $(S-\bar{S})\times A_3$ term is included to balance the regression results (SAS Institute, 2017).

The performance prediction model is a single expression for all alternatives. In order to facilitate the interpretation of its coefficients, the model is split into four functions, one for each alternative (Table 3.6). These alternative-specific models were obtained through the following four algebraic simplification steps:

- Add the appropriate A_{1a}, A_{1b}, A_2 or A_3 term to the overall intercept to obtain the intercept for each alternative strategy. Add the appropriate $(F-\bar{F})\times A_{1a}$, $(F-\bar{F})\times A_{1b}$,

Table 3.6. Coefficients of the performance prediction function of each alternative

	Alternative			
Term	1a	1b	2	3
Intercept	27.5991	28.4571	47.9400	39.5097
Q^*	3.4230	3.4755	5.3253	4.6777
U^*	-33.0494	-33.0494	-33.0494	-33.0494
P^*	0.0132	0.0132	0.0132	0.0132
F^*	-1.5684	-1.4608	-2.7408	-2.2537
S^*	-439.7536	-453.7002	-123.2199	-321.6619
Q^{*2}	0.1264	0.1264	0.1264	0.1264
F^{*2}	0.1235	0.1235	0.1235	0.1235
$U^* \times P^*$	0.4839	0.4839	0.4839	0.4839
$Q^* \times F^*$	-0.1965	-0.1965	-0.1965	-0.1965
$Q^* \times S^*$	-29.2949	-29.2949	-29.2949	-29.2949
$F^* \times S^*$	19.0625	19.0625	19.0625	19.0625

$(F-\bar{F})\times A_2$, or $(F-\bar{F})\times A_3$ term to the coefficient of the $(F-\bar{F})$ term to obtain the coefficient of the F* term for each alternative strategy.

- Add the appropriate $(S-\bar{S})\times A_{1a}$, $(S-\bar{S})\times A_{1b}$, $(S-\bar{S})\times A_2$, or $(S-\bar{S})\times A_3$ term to the coefficient of $(S-\bar{S})$ term to obtain the coefficient of the S^* term for each alternative strategy.
- To visually simplify the written model, the notation $(Q-\bar{Q})$, $(U-\bar{U})$, $(P-\bar{P})$, $(F-\bar{F})$, and $(S-\bar{S})$ is replaced by Q^*, U^*, P^*, F^*, and S^*, respectively, in Table 3.6.

The performance prediction model coefficients can be interpreted as follows:

- The coefficients of all Q^* and Q^{*2} terms yield a convex delay-volume curve, similar to the shape of the exponential delay-volume curve in Krueger's study (1999).
- The coefficients of all U^* terms indicate a negative linear delay-heterogeneity curve when values of P^* and Q^* terms are fixed. This result is not consistent with the shape of the delay-heterogeneity curve obtained by Dingler et al., (2009); however,

in the train delay prediction model proposed by Sogin et al. (2013b), the delay-heterogeneity curve is also negative linear. The results of the previous studies and case studies of the previous chapter suggest the shape of the delay-heterogeneity curve could either be linear or concave.

- The coefficients of all F^* and F^{*2} terms show a concave relationship between the freight train speed and train delay. Increases in freight train speed produce diminishing improvements to train delay. An example relationship between freight train speed and delay is displayed in the analysis section.
- The coefficients of all P^* terms are positive. An increase in passenger train speed has a positive increase impact on freight train delay, matching the observation by Sogin et al., (2013a).
- The positive coefficients of all $U^*{\times}P^*$ terms indicate that the impact of passenger train speed on freight train delay is greater if the current traffic contains more freight trains.
- The negative coefficients of $Q^*{\times}F^*$ and $Q^*{\times}S^*$ terms indicate that the effect of increasing train speed or building new second track to mitigate train delay is greater if the traffic volume is higher.
- The positive coefficient of $F^*{\times}S^*$ term show that the effect of increasing percent ST is smaller if the freight speed is higher.
- Even though the cross terms between the second order, or interaction terms of numeric indices and alternatives were considered in the model selection step, they were not selected because they are not statistically significant.

Equation 3.3 is the constructed performance prediction model and Equation 3.4 is its domain. It was then used for two of the three evaluations: point elasticity and interaction analyses.

$$
\begin{aligned}
D_n = & -33.0494\ U^* + 0.0132\ P^* + 0.1264\ Q^{*2} + 0.1235\ F^{*2} + 0.4839\ U^* P^* \\
& -0.1965\ Q^* F^* - 29.2949\ Q^* S^* + 19.0625\ F^* S^* \\
& + \begin{cases} 27.5991 + 3.4230\ Q^* - 1.5684\ F^* - 439.7536\ S^* & n \in A_{1a} \\ 28.4571 + 3.4755\ Q^* - 1.4608\ F^* - 453.7002\ S^* & n \in A_{1b} \\ 47.9400 + 5.3253\ Q^* - 2.7408\ F^* - 123.2199\ S^* & n \in A_2 \\ 39.5097 + 4.6777\ Q^* - 2.2537\ F^* - 321.6619\ S^* & n \in A_3 \end{cases}
\end{aligned}
\tag{3.3}
$$

$$
\begin{aligned}
& 8 \le Q = Q^* + \overline{Q} \le 24,\ 25\% \le U = U^* + \overline{U} \le 75\%,\ 79 \le P = P^* + \overline{P} \le 110, \\
& 30 \le F = F^* + \overline{F} \le 50,\ 13\% \le S = S^* + \overline{S} \le 19\%,\ A_n \in \{A_{1a}, A_{1b}, A_2, A_3\}
\end{aligned}
\tag{3.4}
$$

3.3.5. Point Elasticity Analysis

Elasticity, or point elasticity in the mathematical field, is an index used to measure the effect of an independent variable on a dependent variable. Elasticity is calculated with Equation 3.5.

$$e = \frac{\Delta Y}{\Delta X} \cdot \frac{X_o}{Y_o} \tag{3.5}$$

Where:

e: point elasticity

ΔX, ΔY: changes of independent and dependent variables, and

X_o, Y_o: baseline condition

Elasticity is a dimensionless parameter, so this estimate is independent of the units of the two variables. Since the numerator and the denominator of elasticity are normalized, it is an appropriate index to be used in this study to compare the pure impact of factors with varying units and numeric ranges. The elasticity calculation used the average freight train delay per 100 train-miles predicted by the regression model as each single factor was varied +/- 15 percent from a baseline operating condition. The index values of the baseline operating condition were set to the medium values in Table 3.3.

The result of the point elasticity calculation for the numeric factors is displayed in a tornado chart (Figure 3.4) that also summarizes the baseline operating condition for the elasticity analysis. The positive and negative elasticity in Figure 3.4 are related to a 15 percent increase or decrease in the value of each numeric factor. Large point elasticity of an index indicates that the train delay response is sensitive to the change of the index.

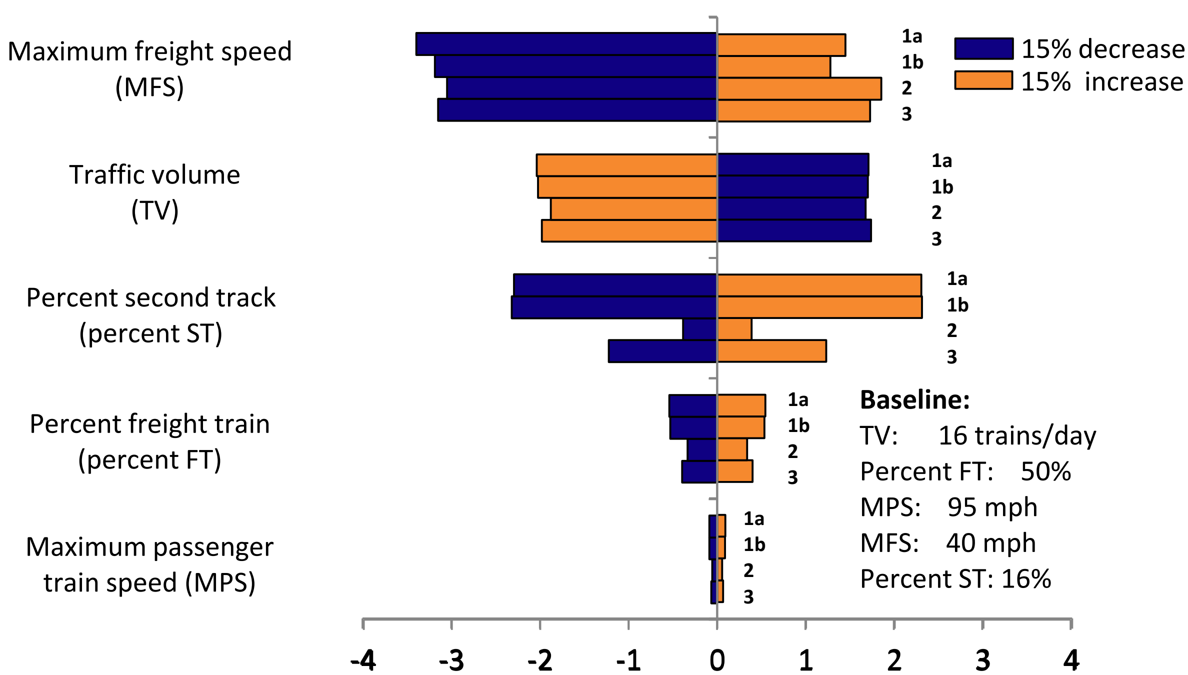


Figure 3.4. Point elasticity of indices under different alternatives

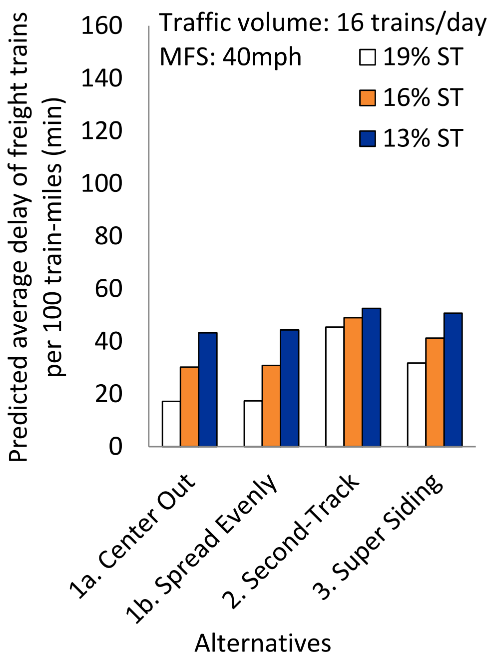
The magnitude of the calculated elasticity shows that the maximum freight speed (MFS), the traffic volume (TV) and the percent second track (percent ST) have the largest impacts on train delay. The MFS, TV, and percent ST were selected for the interaction analysis since they have larger impact. Including the percent ST in the interaction analysis can help visualize the comparison of the average train delays between different alternatives strategies.

The elasticity of percent ST to train delay also shows the effectiveness of each alternative in improving capacity. From the elasticity analysis, Alternatives 1a and 1b are the most efficient method for reducing delay, Alternative 3 is second and Alternative 2 only manages to slightly reduce delay when its level of percent ST increases. Moreover, the elasticity of MFS suggests that increasing this parameter will increase capacity. This factor could be important on single-track lines with sparse sidings that are experiencing increasing traffic. The maximum speed of the passenger train (MPS) has little effect, consistent with previous research regarding the operating behavior of single-track trains (Sogin et al., 2013a).

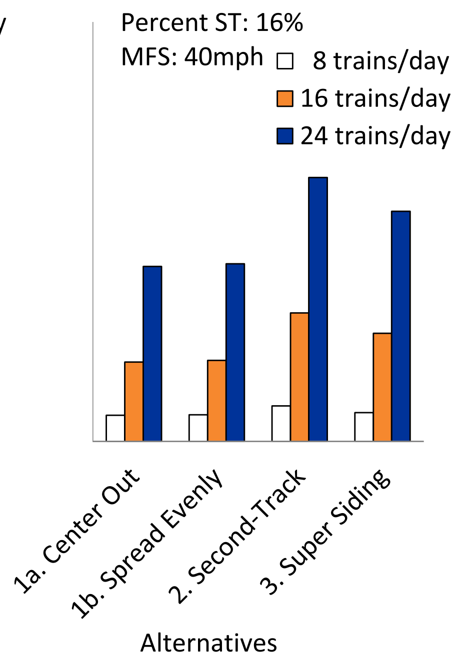
3.3.6. Interaction Analysis

To compare the performance of alternative strategies under different operating environments, the interactions between the major factors and alternatives were analyzed (Figure 3.5).

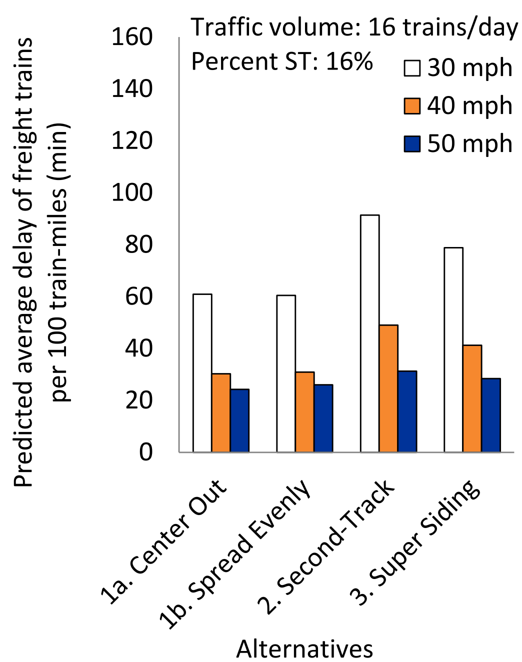
According to the values of average freight train delay per 100 train-miles for different amounts of second track (Figure 3.5a), traffic volumes (Figure 3.5b), and maximum freight speeds (Figure 3.5c), Alternative 1a and 1b have the lowest average delay compared to the other alternatives. Alternative 2, where a single long section of double track is created, consistently performs worse than the other alternatives. This indicates that Alternative 1a and 1b may



(a)



(b)



(c)

Figure 3.5. Result of the interaction analysis under different (a) percent second track (percent ST), (b) traffic volumes (TV), and (c) maximum freight speeds (MFS)

generally be the best potential candidates for implementation because they have the lowest average train delay. The values of average freight train delay per 100 train-miles for different amounts of second track, traffic volumes, and maximum freight speeds are listed in Table 3.7.

Table 3.7. Result of the interaction analysis under different (a) percent second track (percent ST), (b) traffic volume (TV), and (c) maximum freight speed (MFS)

(a)

		Average predicted freight train delay per 100 train-miles of each alternative (min)			
Index	Value	1a	1b	2	3
Percent ST (%)	13	43.3	44.3	52.6	50.8
	16	30.3	30.9	49.0	41.3
	19	17.2	17.4	45.5	31.8

* Traffic volume: 16 trains/day, MFS: 40mph

(b)

		Average predicted freight train delay per 100 train-miles of each alternative (min)			
Index	Value	1a	1b	2	3
Traffic volume	8	10.0	10.2	13.5	11.0
(trains/day)	16	30.3	30.9	49.0	41.3
	24	66.7	67.7	100.7	87.8

*Percent ST: 16%, MFS: 40mph

(c)

		Average predicted freight train delay per 100 train-miles of each alternative (min)			
Index	Value	1a	1b	2	3
MFS (mph)	30	60.9	60.5	91.4	78.8
	40	30.3	30.9	49.0	41.3
	50	24.3	26.0	31.3	28.4

* Traffic volume: 16 trains/day, Percent ST: 16%

Since Alternative 1a and 1b have equivalent train delay, their point elasticity indices reflect their robustness to baseline traffic conditions. For the MFS, Alternative 1a seems to be a bit more fragile to the uncertainty of MFS than Alternative 1b. For TV, they have equivalent robustness since the magnitudes of elasticities are close. Besides the sensitive traffic indices, Alternative 1a and 1b are similarly robust at handling fluctuations in the percent FT and MPS.

If MFS increases, the delay reduction due to the incremental addition of percent ST is reduced (Figure 3.6). This result implies that the higher the freight train operating speed, the less effective additional sidings are at mitigating congestion. Thus there exists a trade-off between investing in track speed improvement projects and infrastructure expansion projects.

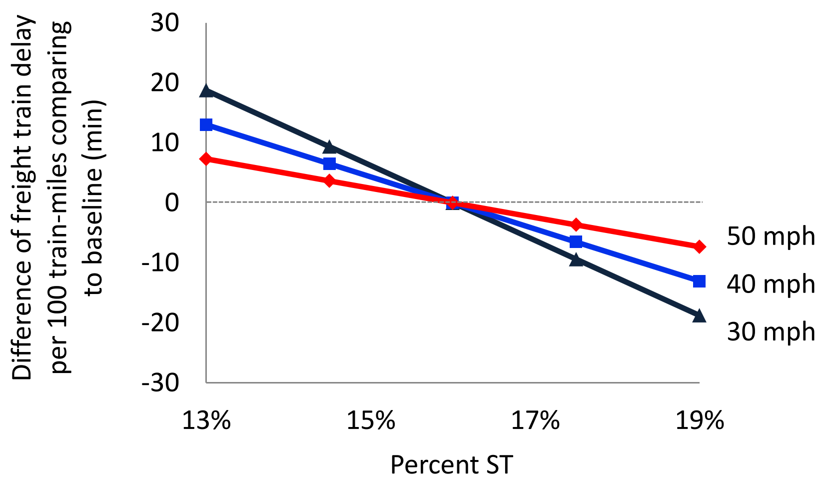


Figure 3.6. Interaction between maximum freight speed and percent ST under Alternative 1a

3.3.7. Reliability Analysis

In the interaction analysis, Alternatives 1a and b were found to have the lowest average train delay and both strategies appear to have nearly equal average values of freight train delay. Equal average freight train delay does not necessarily lead to equivalent performance, since this single value does not capture the variability in freight train delay.

The distribution of freight train delay for each scenario was chosen as an index to measure the reliability of an alternative to handle traffic under different percent ST (Figure 3.7).

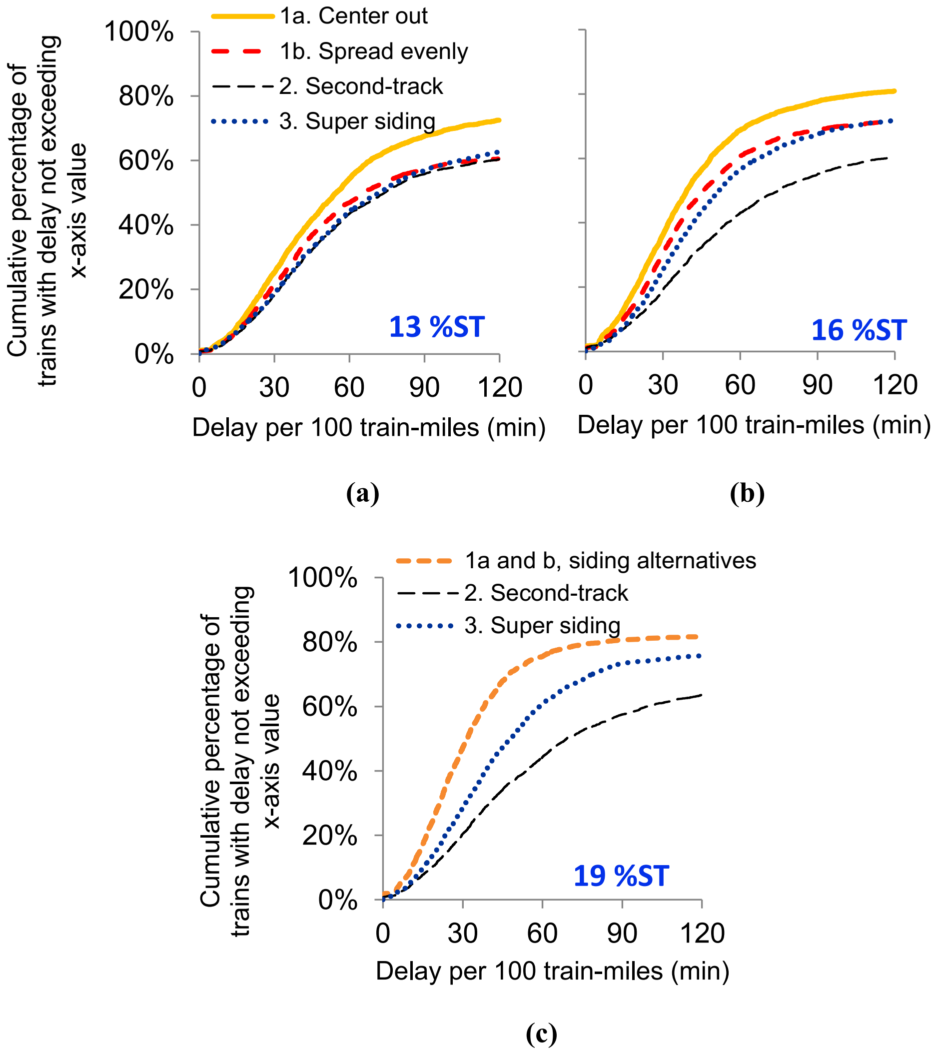


Figure 3.7. Output of the reliability evaluation based on (a) 13 percent, (b) 16 percent, and (c) 19 percent second track

The y-axis is the cumulative percentage of trains delayed less than the corresponding delay on the x-axis.

To allow direct comparison of the reliability of each alternative strategy, the same set of baseline conditions were simulated for 32 runs at each of the three percent ST levels to obtain a series of train delay distributions for each alternative strategy. The number of replicates for each scenario was increased from six to 32 in order to increase the randomness involved in the experiment and to provide a more robust test of the reliability of each alternative.

For example, 20 percent of freight trains have less than 30 minutes of delay in Alternative 2 with 16 percent ST (Figure 3.7b). Additionally, alternatives 1a and b lead to the same dense single-track line so they share the same train delay distribution at 19 percent ST.

Alternative 1a has the greatest reliability because it consistently has the highest percentage of lower-delay trains compared with the other alternatives. Although Alternatives 1a and 1b begin and end with the same track configuration (Figure 3.7c) and delay distribution, the intermediate steps show different delay characteristics. More specifically, despite having equal average train delay values, Alternative 1b (where the new siding projects are distributed evenly over the route) consistently presents a larger percentage of high-delay trains as compared with Alternative 1a (in which the new sidings are grouped together toward the middle of the route). This finding suggests that the exact order and pattern of passing siding additions may influence the reliability of a rail corridor.

Overall, even though Alternative 1a is a bit more fragile to fluctuation in the MFS than 1b, Alternative 1a has the best performance in terms of both efficiency and reliability. This suggests that it may be the preferred capacity expansion strategy for single-track lines with sparse sidings under the conditions considered in this case study.

3.4. Extension of Case Study to the Incremental Benefit of Second Main Track

Developing the relationship between percent ST and average freight train delay was a secondary objective of the case study research of this chapter. This study covers the range of percent ST between 10 percent and 19 percent, whereas the range of partial second track (referred to as partial double-track) examined by Sogin et al. (2013a; 2013b) was between 19 percent and 100 percent. The two studies combined offer a wider understanding of the relationship between percent ST and average freight train delay per 100 train miles.

To develop this relationship, a high-resolution experiment was conducted containing seven different levels of percent ST (10 percent, 11.5 percent, 13 percent, 14.5 percent, 16 percent, 17.5 percent, and 19 percent) and eight levels of homogeneous freight traffic volume (8, 12, 16, 20, 24, 28, 32, 36 freight trains per day at a 50 mph maximum speed). Homogeneous traffic was used instead of heterogeneous traffic to maintain consistency between the output of this case study and the relationship obtained by Sogin et al. (2013a; 2013b). Each combination of percent ST and traffic volume was simulated according to the Alternative 1a expansion strategy with six replicates to obtain 30 days of traffic for each combination.

The simulation results were fit to both linear and polynomial regression values and an R-square test of both methods was used to select an appropriate regression model. The R-square value of the second-order polynomial model was better suited to the results than the value from the linear model, but the polynomial exhibited over-fitting problems. Some polynomial regression lines are convex and inconsistent with other regression lines that curve downward when the percent ST is low. Moreover, the R-square values of the linear models range from 0.855 to 0.972. The precision of the linear model and overfitting characteristics of the polynomial model indicate that the linear function is a better method for describing the relationship between average train delay and percent ST (Figure 3.8). This finding is consistent with the study by Sogin et al (2013a; 2013b), where the relationship between percent ST and delay was also linear in the range of 19 to 100 percent ST.

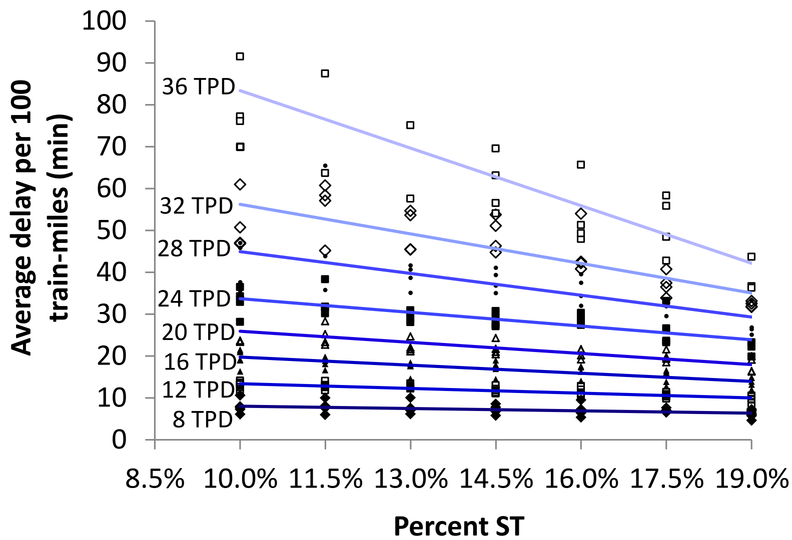


Figure 3.8. Linear relationship between average train delay, percent ST and traffic volume

Although average freight train delay is a good index for evaluating capacity, translating this value into the maximum train throughput per day provides a more straightforward and communicable index for practical use. Sogin et al. (2013b) proposed a method to transform average train delay into train throughput capacity (trains per day). The relationship between percent ST and capacity under different LOS (defined by a maximum allowable train delay per 100 train-miles) is convex but very close to linear (Figure 3.9).

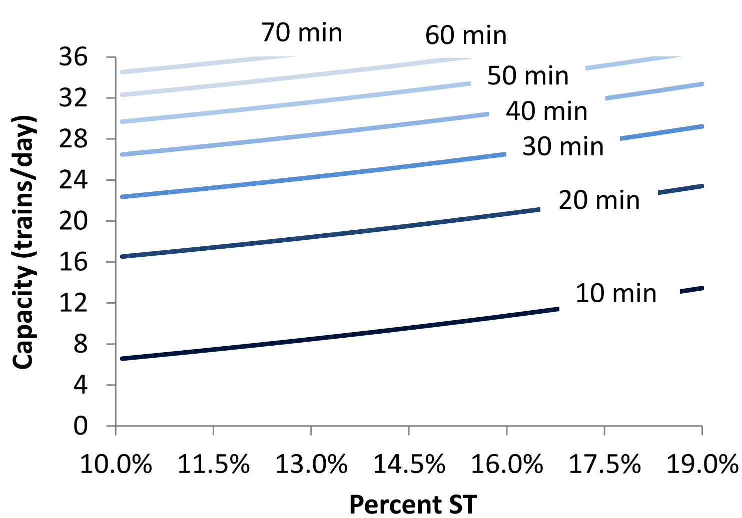


Figure 3.9. Relationship between line capacity, percent ST and maximum allowable delay

Sogin et al. also showed that the capacity versus percent ST curve in the lower range of percent ST above 19 percent is also close to linear. The relative magnitude and slope of the contours of this study compared with those of Sogin et al. show good agreement at the dense single-track network (19 percent ST) interface common to both studies. The linear relationship between percent ST and capacity implies that the bottleneck of single-track lines with sparse sidings, which requires a large investment to increase capacity, needs to be carefully considered to ensure the cost-effectiveness of the engineering option selected.

This portion of the case study involves homogeneous freight traffic. For heterogeneous traffic, the capacity evaluation process proposed in Chapter 2 can be used for analyzing the capacity defined by train-type-specific LOS. The Base Train Equivalent method proposed by Lai et al. (2012) could also be used to transform the heterogeneous traffic into an equivalent number of freight trains used to develop the illustrated relationship.

3.5. Discussion and Conclusion

The objective in this chapter was to find the best capacity expansion strategy for single-track lines with sparse sidings. To select the best alternative strategy, point elasticity, interaction and reliability analyses were used to evaluate the performance of alternatives according to RTC simulation data and resulting regression models. For the specific rail line in the case study, the three analyses determined that concentrating passing siding projects toward the middle of a sparse single-track corridor is the best-performing strategy to increase line capacity when the amount of a second main track is in the range of 10 to 19 percent.

The point elasticity and interaction analyses indicate that both infrastructure improvements and operating strategies associated with increases in the maximum speed of freight trains can be used to increase line capacity. For example, investments to increase FRA track class or reduce civil speed restrictions on a single-track line may be investigated as feasible options for increasing line capacity without adding additional track. The economics of this trade-off on lines with low traffic levels and sparse sidings should be studied further. The point elasticity analysis also provides practitioners with information on the robustness of the infrastructure alternative to the assumed future traffic conditions

The result obtained from this study also expands understanding of the transition process from a single-track line to a full double-track line. The relationship between average freight train delay (capacity) and the percent ST under the preferred alternative strategy was plotted according to the results of additional simulations. The output can be used to understand the relationship between capital infrastructure investment and delay after the percent ST axis is converted to the construction cost appropriate for a particular line. The results presented here and those of Sogin et al. (2013b) combine to further demonstrate the linear relationship between percent ST and average train delay.

Nevertheless, a number of questions related to the transition processes remain unanswered. According to Lindfeldt (2012b), adding new sidings is not the best alternative to increase line capacity under the scenario of hybrid lines that contain both passing sidings and longer segments of partial second main track. He found that extending the length of a second main track can provide more flexibility for various types of timetables and improves practical capacity more than additional sidings. Since the percent ST in Lindfeldt's study is higher than in all the cases used in this study, there might be a level of percent ST where the scenario of adding sidings is no longer the most effective alternative and instead extension of second main track is. Knowing the particular conditions and levels of percent ST where siding projects perform better and where siding connection and double-track extension projects perform better should be the subject of further study.

At a higher level, this chapter formalized a performance evaluation process for capacity expansion project alternatives and alternative expansion strategies. The process involves identification of infrastructure and traffic scenarios, experiment design, simulation, regression and performance analysis The performance analysis considers three different comparisons that examine the average train delay, train delay distribution and sensitivity of train delay to changes in operating parameters. The process to assess the performance of alternative capacity expansion projects can be used by practitioners on any rail line, including single-track lines with sparse sidings. The final output of the three analyses can be weighed by practitioners to recommend preferred expansion alternatives on lines that are experiencing, or expected to experience, traffic congestion.

CHAPTER 4: OPTIMIZATION OF SIDING LOCATION FOR SINGLE-TRACK LINES UNDER STRUCTURED OPERATIONS

An earlier version of this research appears in:

Shih, M.C., Y.C. Lai, C. Dick, and M.H. Wu. 2014b. Optimization of siding location for single-track lines. Transportation Research Record: Journal of the Transportation Research Board, 2448: 71-79.

On single-track rail lines, proper allocation of passing siding locations improves operational efficiency. Too many or too few sidings results in excessive or insufficient line capacity, respectively. Railroad mainlines may be hundreds of miles long with uneven distribution of siding locations, numerous speed restrictions, and a heterogeneous traffic pattern with a varying number and timing of train departures each day. These complexities make it difficult to select the best locations for new passing sidings analytically. Poor decisions on siding placement leads to inefficiency and train delay. Simulation models are capable of incorporating these complexities, but doing so is data and resource intensive, making it difficult to consider all possible alternatives. Use of simulation alone cannot guarantee an optimal solution will be found unless all alternatives are considered, which will often be infeasible. In this chapter I develop an optimization model to determine the number and location of passing sidings on single-track lines with sparse sidings under the special case of structured operations.

4.1. Overview of the Current Status

Railroads usually rely on experienced personnel and established recommended practices (AREMA, 2013) to determine new siding locations during the process of infrastructure upgrades (Vantuono, 2005; BNSF, 2012; Wanek-Libman, 2013). Experienced railroaders often identify good solutions; however, this method does not guarantee that all suitable alternatives have been evaluated or that the best one is implemented (Abril et al., 2008; Lai et al., 2010b). Petersen and Taylor (1987) used simulation analysis to determine the optimal positions of sidings for a line with homogeneous traffic. Pawar (2011) used an analytical model to investigate the relationship between siding length and meet delays. These two studies focused on the effect of siding length and location but did not consider the siding planning problem with heterogeneous traffic.

Lai and Barkan (2011) built a model to select capacity expansion projects in a freight rail network. Lai and Shih (2013) proposed a model to evaluate the strategic capacity planning problem with the consideration of demand fluctuation. However, these models did not consider a detailed expansion plan for the mainline.

Higgins et al. (1997) developed an optimization model to determine optimal siding locations at the mainline scale. The Higgins et al. model is more theoretical than practical as it makes numerous simplifications in determining the number and locations of sidings. It does not include factors such as siding capacity constraints, construction costs, or the existing pattern of passing sidings. In order to offer practical utility, a siding-placement optimization model should account for these factors as well as construction location constraints due to bridges, grade crossings, tunnels, and narrow rights-of-way in urban areas.

In this research I develop an optimal siding location model (OSLM) that considers infrastructure, construction cost, and traffic characteristics to determine the optimal number and location of passing sidings on a single-track route. Railroads can use this tool to assist their siding planning process. It can also be used as a prototype for railroads and researchers to build their own models that are customized for specific infrastructure and business scenarios.

4.2. Methodology

The siding planning problem focuses on determining the optimal number and locations of additional sidings to be constructed on a single-track railway line. Although part of the problem is similar to a capacity planning problem, the solution also requires an approach to establish conflict-free traffic flow on the line, especially for lines with heterogeneous traffic. As a result, the siding planning problem incorporates the ideas of both capacity planning and train dispatching through a series of constraints (Higgins et al., 1997). The first type of constraint guarantees the necessary headway between two adjacent trains to avoid conflicts (Ahuja et al., 1993; Törnquist, 2006; Harrod, 2009; Lamorgese and Mannino, 2013). The length and the capacity of the sidings need to be considered to avoid conflicts on sidings (Qiang and Kozan, 2009; Jaumard et al., 2013). The effect of train characteristics, composition, and commercial schedule must also be taken into account to capture the impact of traffic heterogeneity (Lai et al., 2010a).

In addition to these operational constraints, those related to infrastructure changes must be considered. The possible number and location of prospective sidings must first be identified according to the existing track configuration. The properties of the current track configuration, such as the location of existing sidings and stations, must be considered along with variation in construction cost in order to obtain a practical result. The rail industry's usual method accounts for only a subset of the concepts just mentioned, and thus may be inadequate in generating the most effective siding location plan to increase line capacity. The OSLM was developed to assist in the siding planning process by factoring in a wide range of related parameters that ultimately generate an optimal siding location plan.

4.2.1. Modeling Approach

There are two previous types of mathematical models that can be used as the basis for developing the train dispatching mechanism required by the OSLM: the network-based model (Ahuja et al., 1993; Cordeau et al., 1998; Harrod, 2009) (Figure 4.1a) and the job-shop model (Higgins et al., 1996; Qiang and Kozan, 2009; Liu and Kozan, 2011) (Figure 4.1b).

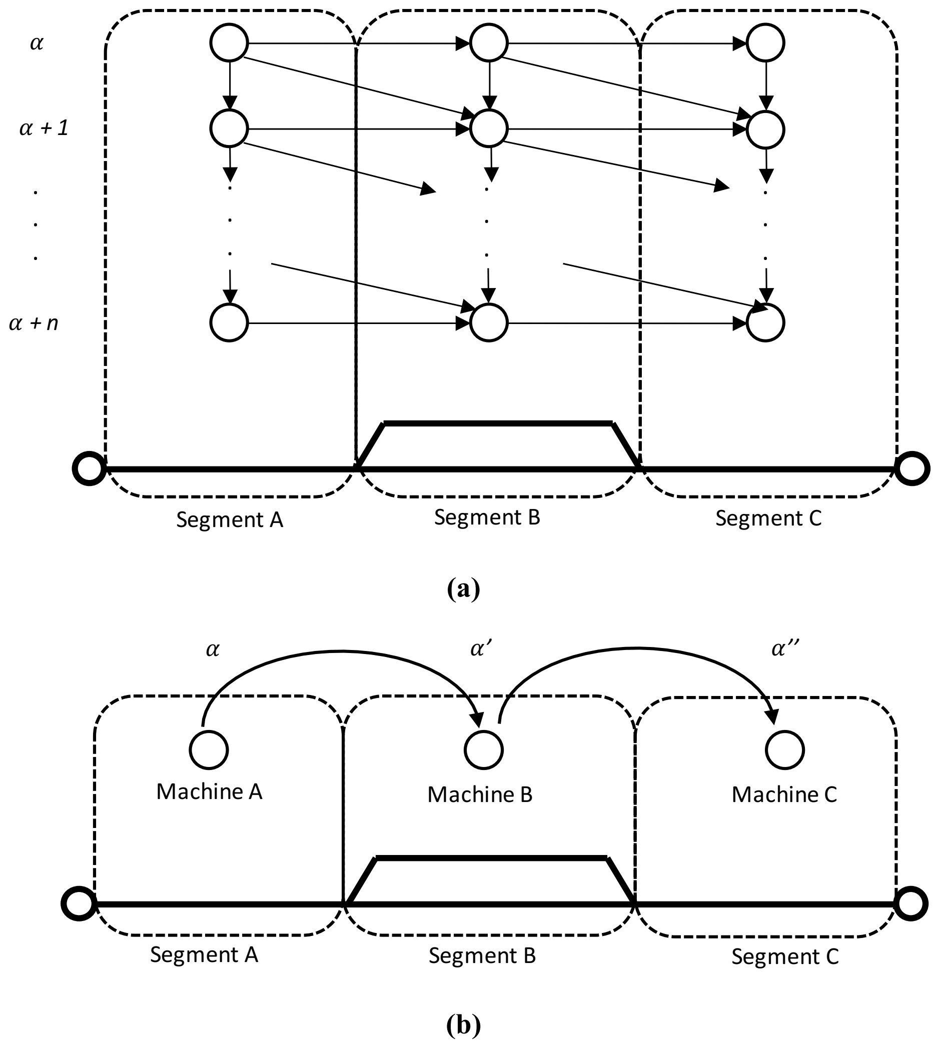


Figure 4.1. Two types of train dispatching models (a) network-based model (b) job-shop model for train dispatching

The network-based model (Figure 4.1a) regards time and distance as discrete units so the scheduling of trains can be represented by the multiple commodities on a hypergraph. Each node in the hypergraph represents a specific time (n evenly divided time units from time α to $\alpha+n$: α, $\alpha+1$, …, $\alpha+n$) and space (segment) unit with each link representing a movement in time and distance space. Train paths of traffic can be derived from the nodes and links (black arrows) that a train has passed.

The job-shop model (Figure 4.1b) originated as a machine scheduling model in the field of industrial engineering. The machine scheduling problem focuses on the sequence and time to route different raw materials to machines in order to produce a final product. This model can also be used to solve train dispatching problems since the segments of a rail line can be regarded as machines, the trains as materials, and the movement of trains across each line segment (black arrows) as sequential steps in the production process completed by each machine (Liu and Kozan, 2011). The train path can be obtained from the input and output time (α, α', α'') of a train to a segment. In a job-shop model, time and distance are treated as continuous variables. This characteristic provides more flexibility for determining the location of additional sidings and makes the job-shop model the preferred approach for this study.

The train dispatching model developed by Higgins et al., (1996) was different from the original job-shop model, because it only focused on start, end, and siding segments, instead of all parts of a line (Figure 4.2). The effect of signal blocks on the single track between passing sidings was accounted for by introducing a minimum train headway constraint between two adjacent trains. This modified model contains fewer variables and parameters than the original job-shop model and therefore can exhibit improved solution efficiency while maintaining the soundness of the model output.

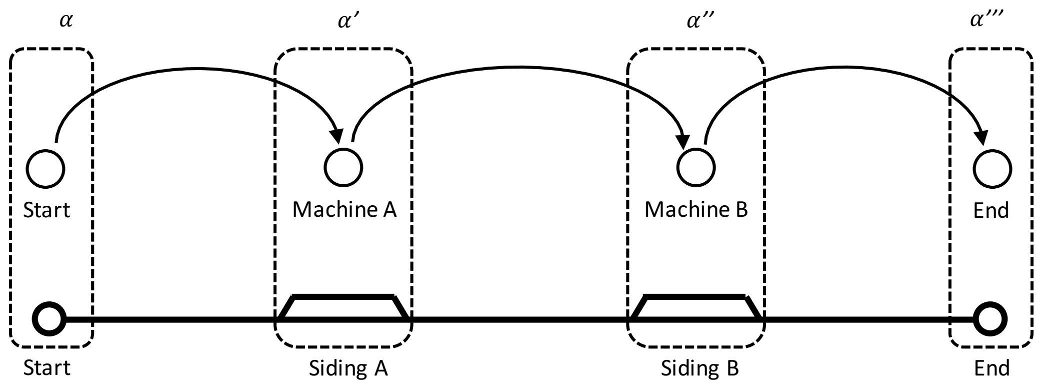


Figure 4.2. Modified job-shop model for train dispatching

In the Higgins et al. (1997) siding planning model, additional variables representing the number and locations of sidings were added to solve the siding planning problem. OSLM follows a similar structure but improves upon Higgins et al.'s siding model by modifying the formulation to make it more applicable to actual siding location problems encountered by practitioners. The details of the OSLM formulation will be introduced in section 4.2.4.

There are two reasons to adopt the structure of Higgins et al.'s modified model. First, since it only considers segments related to sidings, the train dispatching mechanism in Higgins et al.'s model is more efficient than detailed train dispatching models that consider each individual signal block (Liu and Kozan, 2011). Second, unlike the discretized time and distance units used in the network-based model (Figure 4.1a), the time and distance variables used in Higgins et al.'s model are continuous. This allows the model to generate a more precise output compared to the network-based model. Although the precision of the network-based model can be increased by using smaller distance and time units, this change will increase the size of model, making it less preferable than the job-shop model.

4.2.2. Optimization Model Framework

Traffic characteristics, track infrastructure properties, and operational parameters are used as inputs to the OSLM (Figure 4.3). On the basis of these input parameters, the optimization framework generates two types of output - train paths and an optimal siding location plan - that minimize the total of three cost categories: equivalent capital investment cost, meet and pass delay cost, and late departure cost.

Figure 4.3. Conceptual diagram of OSLM

An optimization model for the siding planning problem needs to deal with the siding location and train dispatching problem at the same time. Consequently, a combination of capacity planning and train dispatching constraints are used as the basic structure of the model. The models developed in previous studies typically provided either an optimal siding plan for a fixed schedule or an optimal schedule for a fixed set of siding locations, but were incapable of solving the complete problem by optimizing both simultaneously. The OSLM is able to generate an optimal siding location plan and a set of train paths to minimize total cost (including capital investment cost, delay cost, and late departure cost) without violating a set of practical constraints (e.g., train separation, construction cost and siding capacity).

4.2.3. Data Preprocessing

Most of the detailed input data required by OSLM (Table 4.1) can be used directly by the model but the infrastructure inputs need to be preprocessed into nodes and segments (Figure 4.4). In the processed infrastructure input, q represents nodes or sidings and the n in q_n is the index number of nodes or sidings along the line under study. The notation p_n represents the segments between each pair of adjacent nodes (sidings, stations, and yards) on the line, and n in p_n is the index number of each segment along the line. The notation c_n stands for construction zones, and n is the index number of each construction zone along the line. From the number and location of existing nodes, the maximum number and relative location of prospective sidings can be determined.

Table 4.1. Input data to OSLM

Traffic characteristics	Infrastructure properties	Operational parameters
-Maximum train speed of different train type toward each direction	- Zones related to different construction cost (milepost and USD)	- Priority of trains (delay cost in USD per hr)
- Number and direction of each type of trains (trains/per day)	- Length of existing sidings, prospective sidings and the line (mile)	- Turnout switching time (hr)
- Scheduled departure time for trains (hr)	- Speed limit of the sidings (mph)	
- Lost time per acceleration and deceleration (hr)	- Minimum siding spacing (mile)	
- Safety headway for adjacent trains (hr)	- Location of existing sidings and stations (milepost)	
- Commercial schedule for passenger trains (hr)	- Average speed limit of a line (mph)	

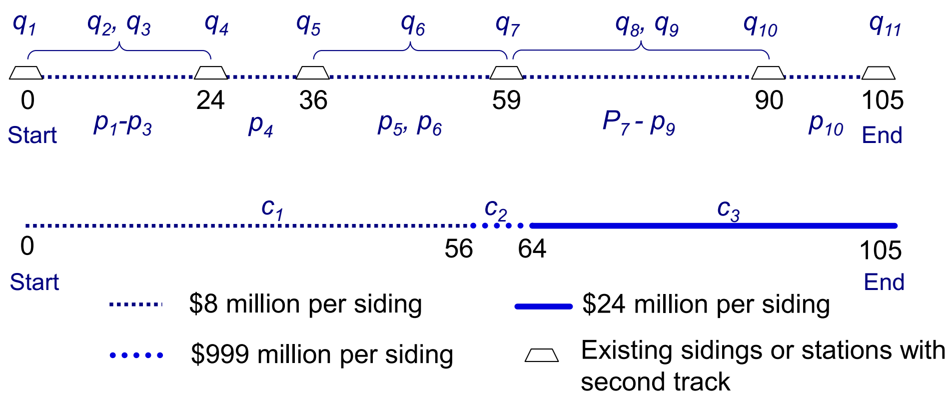


Figure 4.4. Example of preprocessing for infrastructure data

The maximum number of possible sidings between two existing sidings can be calculated by $\lfloor d/g \rfloor - 1$, where d is the segment length between two adjacent sidings and g is the minimum siding spacing. For example, in Figure 4.4 the spacing between the first existing siding and the starting node is 24 miles. Since the minimum siding spacing in this study is assumed to be 8 miles, the maximum number of prospective sidings is $\lfloor 24/8 \rfloor - 1 = 2$. Therefore, two possible sidings, q_1 and q_2, are identified between the starting node and the first siding. Both q_1 and q_2 can be built anywhere between the two existing sidings if the minimum siding spacing constraint is not violated. This holds for all sidings q_1- q_n throughout the model.

Following this process, several possible sidings (q_2, q_3, q_6, q_8, q_9) were identified and labeled in the example network (Figure 4.4). Moreover, the boundaries of each construction zone and the associated cost of siding construction can be either referenced from similar projects on other lines or obtained with high-level estimation methods. If there are particular locations where siding construction is undesirable (e.g. sections with multiple grade crossings or a narrow right-of-way), an arbitrarily high construction cost can be assigned to these inappropriate sites, much like the c_2 zone illustrated in Figure 4.4, that was given a $999 million cost of construction.

If available, more realistic construction cost values should be used for these undesired construction zones since using unnecessarily large arbitrary penalty values in a mixed integer program model can potentially increase the solution time.

4.2.4. Model Formulation

The OSLM uses the concept of mixed integer programming (Ahuja et al., 1993; Lai et al., 2010a; Lai et al., 2010b) and job-shop modeling (Qiang and Kozan, 2009). It is similar to some rail scheduling or tactical planning models (Crainic et al., 1984; Cordeau et al., 1998). The following paragraphs present the OSLM formulation.

There are three different types of OSLM decision variables: time variables, infrastructure variables, and train dispatching variables. Time variables indicate the arrival and departure time of trains at each node. The value of time variables can be used to construct the train paths.

D_i^q: departure time of train i at node q, $D_i^q \geq 0$

A_i^q: arrival time of train i at node q, $A_i^q \geq 0$

The infrastructure variables determine the need and location of additional sidings. An optimal siding plan can be obtained from the infrastructure variables.

d_p: positive variable, length of segment p, $d_p \geq 0$

z_c^q: equal to 1 if siding q exists in construction zone c, 0 otherwise, $z_c^q \in \{0,1\}$

Train dispatching variables are included in OSLM to ensure the headway between trains and avoid the potential conflicts between trains.

x_{ij}^p: equal to 1 if train i passing through segment p before train j, 0 otherwise, $x_{ij}^p \in \{0,1\}$

o_i^q: equal to 1 if train i stays on siding q to meet or pass another train during the dispatching period, 0 otherwise, $o_i^q \epsilon\{0,1\}$

θ_{ij}: equal to 1 if and only if train i stays on siding q to meet or pass before train j stays on the same siding, 0 otherwise, $\theta_{ij}^q \epsilon\{0,1\}$

In addition to the decision variables, OSLM uses many other indices (Table 4.2), sets (Table 4.3) and parameters (Table 4.4). Equation 4.1 is the OSLM objective function. The objective function of the Higgins et al. (1997) model only considered total train delay. This model formulation covers a more comprehensive set of related costs. OSLM aims to minimize

Table 4.2. Indices used in OSLM

Index	Description
(i, j) N	Indices referring to trains running on the line
(p, r) P	Indices representing sections of the line
(q, s) Q	Indices for sidings and stations (nodes)
c C	Index referring to order of construction zones

Table 4.3. Sets used in OSLM

Set	Description
b^+	Set of any two trains with same direction
b^-	Set of any two trains with opposite direction
κ	Set of existing and prospective siding nodes
ε_i	Set of origin for train i
η^+	Set of prospective sidings
η^-	Set of existing sidings and stations
k_i	Set of destination for the train i
δ_p	Set composed of all section p and adjacent node q to enter the section
θ_p	Set composed of all section p and adjacent nodes (q, s)
π	Set of origins and their adjacent sections

Table 4.4. Parameters used in OSLM

Parameter	Description
$v_M^{\,i}$	Average train speed (mph)
β	Equivalent coefficient for investment cost
$t_i^{\,q}$	Extra travel time for train i to cross siding q than a parallel section on mainline (hr)
$\tau_i^{\,q}$	Scheduled dwell time for passenger train i on station q
g	Minimum siding spacing (mile)
f_i	Lost time due to acceleration and a deceleration of train i (hr)
σ^c	Boundary of construction cost zone c (milepost)
U^c	Cost per siding in construction cost zone c (USD)
φ^q	Location of existing siding q (milepost)
$e_i^{\,+}$	Earliest possible departure time of train i (hr)
$e_i^{\,-}$	Latest possible departure time of train i (hr)
$\lambda_q^{\,i+}$	Earliest allowable arrival time for train i at station q
$\lambda_q^{\,i-}$	Latest allowable arrival time for train i at station q
$h_{ij}^{\,p}$	Safe headway between adjacent train i and j on section p (hr)
ς	Turnout processing time (hr)
$L_i^{\,q}$	Ability for siding q to accommodate train i, if the length of siding q is longer than the length of train i, then $L_i^{\,q}$ =1, otherwise 0
W^i	Delay cost, the cost generated by an idling train-hour, it also reflects the priority of train i
M	An arbitrary large number
E	Total dispatching duration (hr)
B	Available budget (USD)

the total cost during the planning horizon, defined by the summation of equivalent capital investment cost, meet and pass delay cost, and late departure cost. The coefficient β for equivalent capital investment cost can be obtained by the method proposed by Lai and Barkan (2011). Since W^i is the delay cost for different types of trains, this objective function reflects the business objectives of North American railroads (Lovett et al., 2015).

Objective: $$Min\ \beta \sum_{c \in C} \sum_{q \in \eta^{+}} U^{c} z_{c}^{q} + \sum_{i \in N} \sum_{q \in \kappa} W^{i} (D_{i}^{q} - A_{i}^{q}) + \sum_{i \in N} \sum_{q \in \varepsilon_{i}} W^{i} (D_{i}^{q} - e^{+}{}_{i}) \qquad (4.1)$$

This objective is subject to a set of constraints, including constraints on train dispatching, train schedule, siding capacity, construction cost, track configuration, and other operational parameters. The constraints listed in Equations 4.2 to 4.7 ensure the accuracy of the dispatching process. The basic principle is to ensure two adjacent trains at each node have a reasonable headway. Equations 4.2 and 4.4 maintain an appropriate headway between the departure times of any adjacent trains traveling in the same direction, and Equations 4.3 and 4.5 maintain a safe headway between the arrival times of any two adjacent trains. Equations 4.6 and 4.7 guarantee the headway between two adjacent trains in opposite directions.

$$M(1 - x_{ij}^{p}) + D_{j}^{q} \geq D_{i}^{q} + h_{ji}^{p} + o_{j}^{q} \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.2)$$

$$M(1 - x_{ij}^{p}) + A_{j}^{q} \geq A_{i}^{q} + h_{ji}^{p} + o_{j}^{q} \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.3)$$

$$M x_{ij}^{p} + D_{i}^{q} \geq D_{j}^{q} + h_{ij}^{p} + o_{i}^{q} \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.4)$$

$$M x_{ij}^{p} + A_{i}^{q} \geq A_{j}^{q} + h_{ij}^{p} + o_{i}^{q} \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.5)$$

$$M(1 - x_{ij}^{p}) + D_{j}^{q} \geq A_{i}^{q} + h_{ji}^{p} + \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.6)$$

$$M x_{ij}^{p} + D_{i}^{q} \geq A_{j}^{q} + h_{ij}^{p} + \varsigma \qquad \forall (i,j) \in b^{+},\ i \neq j,\ q \in \delta_{p},\ p \in P \qquad (4.7)$$

Equations 4.8 and 4.9 are train schedule constraints that consider the effect of traffic pattern and demand. Equation 4.8 forces trains to depart from their origin within a given time range. Additionally, Equation 4.9 ensures that all passenger trains arrive at stations within an acceptable interval.

$$e_i^{+} \leq D_i^q \leq e_i^{-} \qquad \forall i \in N,\ q \in \pi \tag{4.8}$$

$$\lambda_q^{i+} \leq A_i^q \leq \lambda_q^{i-} \qquad \forall i \in N,\ q \in \kappa \tag{4.9}$$

Equations 4.10 through 4.15 are siding capacity constraints. This set of constraints is one of the improvements made to the Higgins et al. (1997) model. It did not have constraints to ensure that the length of a train dwelling on a siding is shorter than the siding length. Also, the Higgins et al. model did not have constraints to avoid conflicts between two trains using the same siding simultaneously. Equation 4.10 links the train dwell variable o_q^i with the train meet and passing delay. Equations 4.11 and 4.12 identify the sequence of trains passing each siding. Equation 4.13 prevents two trains from occupying the same siding. This equation works together with Equation 4.9, to maintain the stopping pattern of passenger trains. Equation 4.14 forbids a train from using a siding if the length of the train is longer than the siding itself. Equation 4.15 is the arrival time constraint. This equation also captures the extra travel time experienced by trains due to acceleration, deceleration, siding speed limit, and turnout switching time if a train takes sidings. Equations 4.10 and 4.15 are also part of the schedule constraints. The notation τ_i^q in Equations 4.10 and 4.15 ensure the minimum dwell time for passenger trains at stations.

$$Mo_i^q \geq D_i^q - A_i^q - \tau_i^q \qquad \forall i \in N,\ q \in Q \tag{4.10}$$

$$\theta_{ij}^q \geq o_i^q + o_j^q + x_{ij}^p - 2 \qquad \forall i \in N,\ j \in N,\ i \neq j,\ q \in \delta_p,\ p \in P \tag{4.11}$$

$$3\theta_{ij}^q \leq o_i^q + o_j^q + x_{ij}^p \qquad \forall i \in N,\ j \in N,\ i \neq j,\ q \in \delta_p,\ p \in P \tag{4.12}$$

$$A_j^q \geq D_i^q + \varsigma + h_{ij}^p - M(1-\theta_{ij}^q) \qquad \forall i \in N,\ j \in N,\ i \neq j,\ q \in \{\kappa \cap \delta_p\},\ p \in P \tag{4.13}$$

$$o_i^q \leq L_i^q \qquad \forall i \in N,\ q \in \kappa \tag{4.14}$$

$$D_i^q \geq A_i^q + o_i^q(f_i + t_i^q + \varsigma) + \tau_i^q \qquad \forall i \in N,\ q \in Q \tag{4.15}$$

The variation in siding construction cost is taken into account by Equation 4.16. It links the construction cost zones with the location of sidings to determine how much capital investment is required to implement an additional siding. This cost was not considered by the Higgins et al. model and represents another enhancement made in OSLM. The construction cost constraint can be neglected if there is little variation in siding construction cost along the mainline under study.

$$\sum_{c\in C}\sigma^{c-1}z_c^q - M(1-\sum_{c\in C}z_c^q) \le \sum_{r\in\{r\le p\}} d_r \le \sum_{c\in C}\sigma^c z_c^q + M(1-\sum_{c\in C}z_c^q) \quad \forall q\in\{\kappa\cap\delta_p\},\ p\in P \quad (4.16)$$

Track configuration constraints are as follows: Equation 4.17 ensures minimum siding spacing is maintained, Equation 4.18 keeps the location of existing sidings, Equation 4.19 prevents trains from meeting or passing at a node without an existing siding, Equation 4.20 ensures that a siding can only exist in a valid construction zone, and Equation 4.21 ensures that the model selects all existing sidings. The track configuration constraints were created to improve the Higgins et al. model by maintaining the existing infrastructure layout.

$$d_p \ge g - M(1-\sum_{c\in C}\sum_{q\in\psi^+} z_c^q) \qquad \forall p\in P \qquad (4.17)$$

$$\sum_{r\in\{r\le p\}} d_r = \varphi^q \qquad \forall q\in\{\eta^-\cap\delta_p\},\ p\in P \qquad (4.18)$$

$$\sum_{i\in N} o_i^q \le M\sum_{c\in C} z_c^q \qquad \forall q\in Q \qquad (4.19)$$

$$\sum_{c\in C} z_c^q \le 1 \qquad \forall q\in\eta^+ \qquad (4.20)$$

$$\sum_{c\in C} z_c^q = 1 \qquad \forall q\in\eta^- \qquad (4.21)$$

Equation 4.22 is the budget constraint and Equation 4.23 ensures that the OSLM completes the dispatching process within a given time period. Equation 4.24 sets the train running time between any two adjacent nodes as the average running time between them. The average running time can be obtained from simulations (Leilich, 1998) or analytical models (Chen and Harker, 1990; Higgins and Kozan, 1998).

$$\sum_{c \in C} \sum_{q \in \eta^{+}} U^{c} z_{c}^{q} \leq B \qquad (4.22)$$

$$A_{i}^{q} \leq E \qquad \forall i \in N,\ q \in k_{i} \qquad (4.23)$$

$$A_{i}^{q} - D_{i}^{s} = d_{p} \big/ v_{M}^{i} \qquad \forall i \in N,\ (q,s) \in \vartheta_{p},\ p \in P \qquad (4.24)$$

4.3. Case Study

To demonstrate the function of OSLM, a hypothetical single-track line with a length of 105 miles (Figure 4.5) and two usable intermediate passing sidings was considered with three train types, passenger, intermodal and bulk unit trains (Table 4.5). The original traffic volume is estimated to be 14 trains per day by using the Canadian National Railway parametric model and the given route characteristics (Krueger, 1999). The future demand is assumed to be 20 trains per day at the end of the 5-year planning horizon. The question then becomes how to effectively add new sidings to accommodate the new demand.

4.3.1. Case Study Inputs

For rail traffic, a predetermined train schedule indicating the departure time and departure flexibility of each train and the stop schedule of passenger trains was used as input. The departure times of trains are set to be evenly distributed during the day without fleeting; that is,

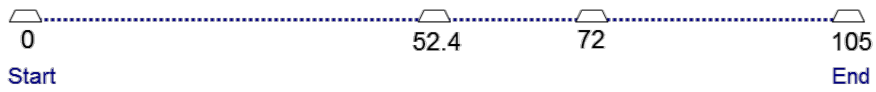


Figure 4.5. Line used in the case study

Table 4.5. Important parameters used by OSLM

Parameter	Value
Fixed maximum train speed (mph)	Passenger: 70 mph Intermodal: 55 mph Bulk: 35 mph
Number of each type of trains (trains/per day)	Passenger: 6 trains/day Intermodal: 8 trains /day Bulk: 6 trains/day
Direction of trains (eastbound/westbound)	Eastbound: 10 trains/day Westbound: 10 trains/day
Priority of trains (delay cost per hr)	Passenger: $3,000/delay hr Intermodal: $1,392/delay hr Bulk: $586/delay hr
Safety headway between two trains	6 min
Planning horizon	5 years

no adjacent trains are the same type. Each train type was dispatched using their corresponding travel time and delay cost.

The case study route infrastructure was preprocessed for use by the optimization model (Figure 4.6). The possible locations of prospective sidings (q) are identified and the locations of higher construction cost zones are labeled. For this case study, the zones with higher construction cost are associated with urban areas. In these locations, the cost of sidings is estimated on the basis of the summation of siding construction, grade separation, and land acquisition costs. Based on typical estimated construction costs for these components, the siding construction cost in an urban area is three times that of a siding in a rural area. The construction cost of a typical rural siding is $8 million, and an urban siding is $24 million.

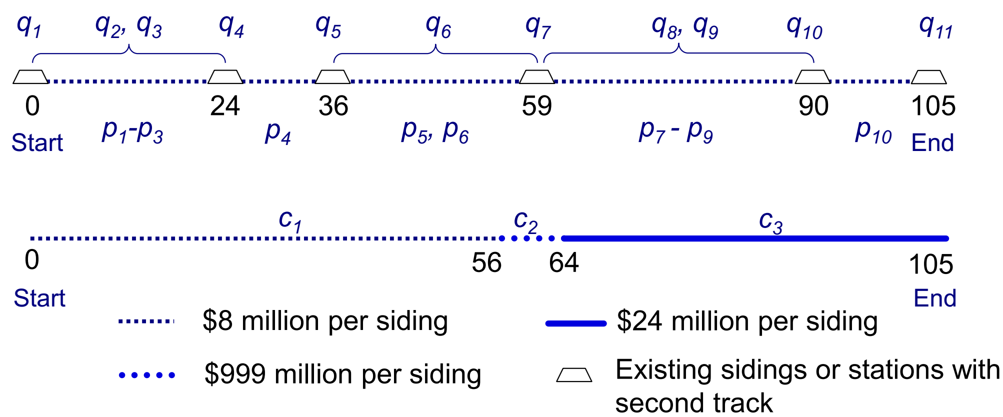


Figure 4.6. Infrastructure data after preprocessing

To demonstrate the importance of variation in construction cost to siding planning, OSLM was applied to the case study route to generate an optimal capacity expansion plan with variation in siding cost and without variation in siding cost. There are two scenarios, Scenario 1 is set to have variation in siding cost on the case study route, and Scenario 2 is set to have no variation in siding cost.

4.3.2. Case Study Results

OSLM was coded into AIMMS (Paragon Decision Technology, 2006) and solved by CPLEX. This model is a large-scale optimization problem with 9,586 variables and 32,812 equations. The solution time ranges from 1 to 8 hours depending on the construction budget available.

OSLM delivers two types of outputs, the train dispatching result and the optimal siding location plan. The string chart derived from the train dispatching result demonstrates that the OSLM constraints provide reasonable train dispatching decisions (Figure 4.7). The optimal siding location plan indicates the number and the locations of additional sidings. The siding plans of Scenario 1 are visualized in Figure 4.8, and Scenario 2 in Figure 4.9. The siding plans

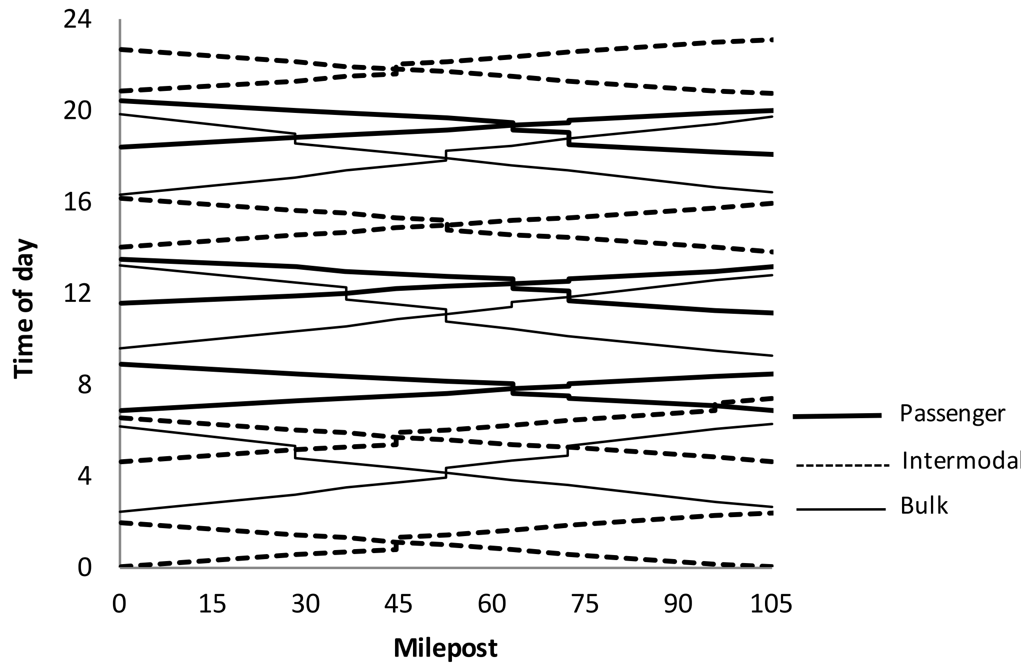


Figure 4.7. Example of string chart based on train dispatching mechanism of OSLM

shown in both figures are not progressions from the original line to the improved line. Instead, they display the optimal final siding plans for an ultimate build-out to the specified budget level. If the sidings are to be phased in over time, additional analysis is required to determine the optimal order of construction. For expansion programs with a longer time frame, the model could be run iteratively to develop a progression of siding projects.

In some cases, siding constructions could occur in zones with higher cost if the benefit is larger than the equivalent construction cost. In this case study, however, they do not. In Scenario 1 in which there is variation in siding construction cost, OSLM avoided constructing sidings at locations in in zones with higher construction cost. In contrast, siding locations in Scenario 2 were only restricted by the minimum siding spacing rule.

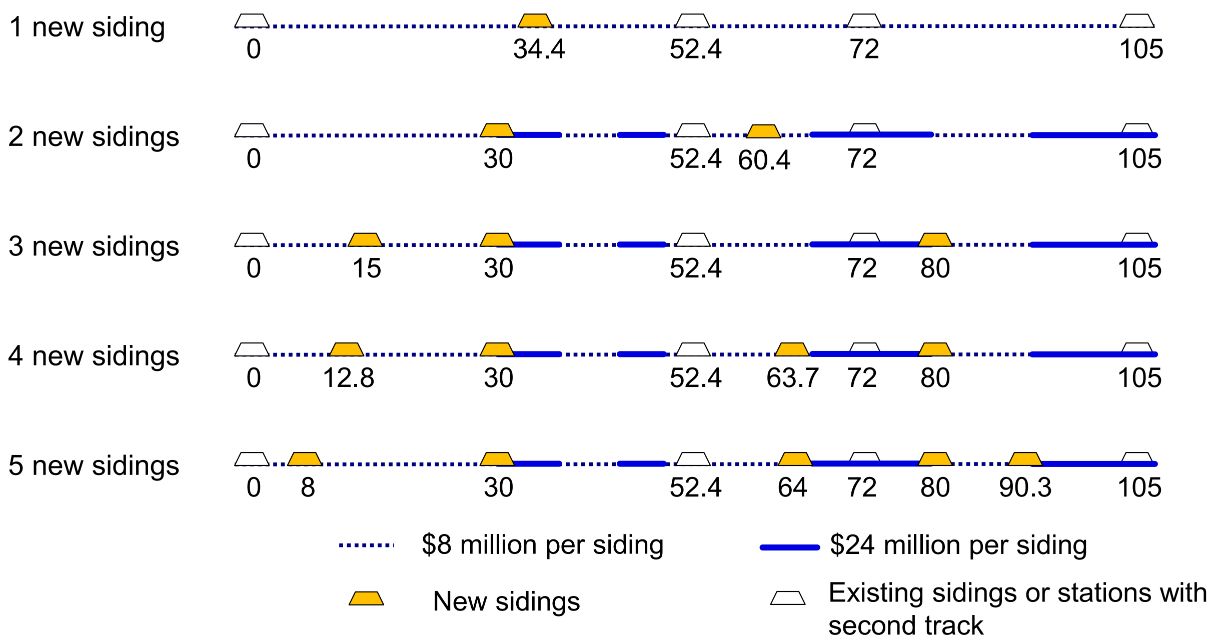


Figure 4.8. Optimal siding expansion plans of Scenario 1: variation in construction cost

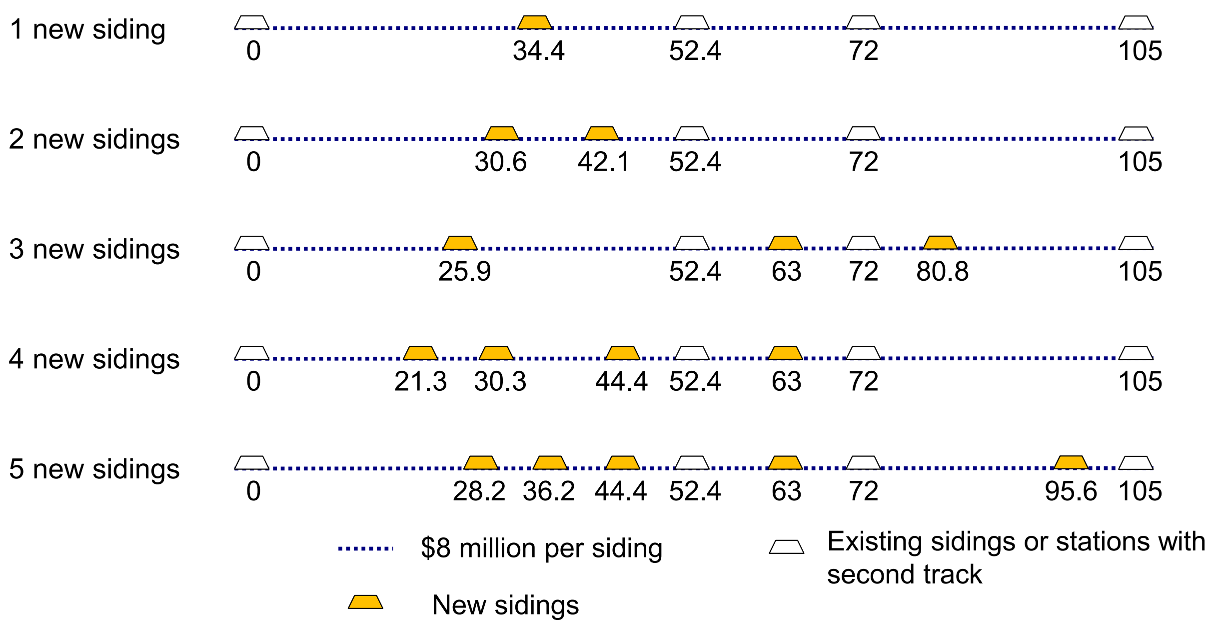


Figure 4.9. Optimal siding expansion plans of Scenario 2: uniform construction cost

The difference between the capacity expansion plans for each number of added passing sidings leads to different relationships of total costs over the five-year planning period and number of added passing sidings (Figure 4.10). In both scenarios, additional sidings increase the construction portion of total cost but reduce the delay and late departure portion of total cost.

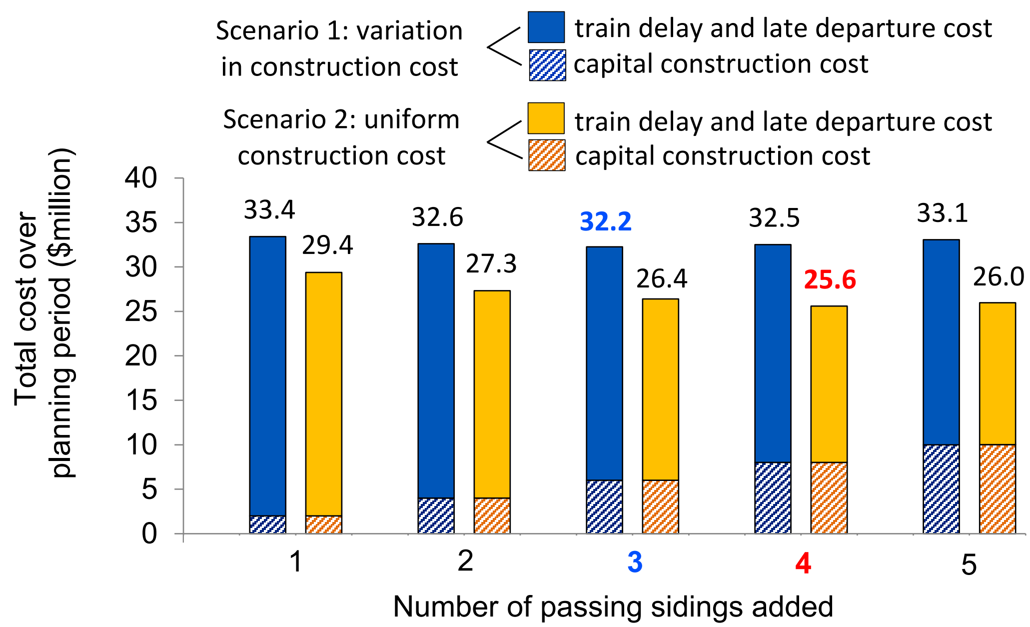


Figure 4.10. Comparison between scenarios with and without variation in construction cost

The total cost of Scenario 1 is relatively constant and higher than Scenario 2. In Scenario 1 the higher-cost zones prevent the additional sidings from being optimally located (Figure 4.8). The selected locations outside the high-cost zones limit the ability of the sidings to reduce train delay costs. The delay cost reduction facilitated by each added siding is approximately equal to the equivalent capital construction cost of that siding, causing the total cost of Scenario 1 to be similar regardless of the number of sidings added. With greater flexibility in selecting optimal siding locations, Scenario 2 obtains a greater reduction in delay costs for each siding added. However, after four sidings are added, Scenario 2 also experiences diminishing returns; the fifth siding does not reduce train delay enough to cover its equivalent capital construction cost, resulting in an increase in total cost over the planning period.

The blue and red numbers in Figure 4.10 show the lowest total costs and the corresponding number of added sidings for Scenario 1 and Scenario 2, respectively. The presence or absence of variation in construction cost has an impact on the number of additional sidings required to minimize total cost over the planning period.

OSLM has also been extended to suggest new passing siding construction and siding length extension projects on routes where trains lengths are being increased (Shih et al., 2015b), and to determine the optimal siding locations on a mainline with significant speed variation (Shih et al., 2015c). The extensions increase the adaptability of OSLM to different scenarios.

4.4. Discussion and Conclusion

When single-track lines with low traffic density experience growth in traffic, they may reach the limits of practical capacity. This study developed the OSLM to help determine the optimal number and locations of additional sidings to aid railways in planning capacity expansion projects. The model provides an optimal siding location plan under structured operation that can be used by railroad infrastructure planners. This model can help railroads maximize their return on investment and improve service quality. OSLM extends Higgins et al. (1997) model by introducing practical engineering cost constraints. The basic model has also been successfully adapted to studies of train length and passing siding extensions (Shih et al., 2015b). OSLM can serve as a basis for other researchers developing modified forms to address other railway capacity and service design questions.

In the case study, OSLM was used to select siding projects for a hypothetical single-track line with sparse sidings under structured operation. The output suggested that the presence of variation in construction cost has a substantial impact on the number of sidings required to achieve minimum total cost over the planning period. It also showed that the existence of higher cost zones tends to restrict the location of siding projects.

4.4.1. Future Study

The OSLM has practical constraints that allow it to generate a reasonable optimal siding location plan under a certain schedule and budget; however, the output cannot be directly applied to single-track lines with multiple train schedules or flexible operations. On passenger rail mainlines, there usually exist multiple schedules, such as peak-hour, off-peak, weekday and weekend, as well as the normal, emergency, and recovery operations. To obtain a more robust siding plan, it is necessary to consider all these schedule variations for the line. The current structure of OSLM can only consider one schedule at a time. Although OSLM improves upon the Higgins et al. (1997) model, it does not generate a robust output for the single-track lines with multiple or flexible train schedules.

The stochastic optimization approach (Heyman and Sobel, 2003; Lai and Shih, 2013) could possibly be used to transform the current OSLM into a new structure that can consider multiple schedules. The uncertainty considered by the stochastic OSLM is the probability of each schedule used, which resulted in different train dispatching results. However, model size after the transformation may be much larger than the original model. Several types of solution algorithms are suggested here to solve the stochastic version of OSLM, in particular the train dispatching mechanism.

The most frequently used family of algorithms for solving large-scale optimization problems is heuristic algorithms. In the field of mathematical optimization, “heuristic” refers to the techniques that are designed to obtain an approximate solution to a complex problem in a faster and more efficient fashion than traditional methods. Heuristics such as the genetic algorithm (Chaudhry and Luo, 2005), Lagrangian heuristic (Brännlund et al.,1998), nearest neighborhood search (Zhao et al., 2010), and simulated annealing (Kirkpatrick et al., 1983) can potentially be applied to a modified OSLM to obtain an approximate optimal siding location plan for multiple train schedules. Meta-heuristics (Voß et al., 2012) might be a useful approach for future research since the algorithms were developed for very large optimization problems.

Decomposition methods (Conejo et al., 2006) for mixed integer programming offer another possible research direction. Decomposition methods group variables into sets, and solve a sub-problem for each set repetitively. These translations are done because solving binary acyclic problems is more tractable than solving the original problem. Bender's Decomposition (Costa, 2005) and Column Generation (Wilhelm, 2001) are the most common decomposition methods. They have been applied to solve rolling stock, crew and locomotive planning and scheduling problems (Cordeau et al., 1998). Bender's Decomposition was also used to solve stochastic programming models related to railway operations and planning (Birge, 1985; Sherali and Fraticelli, 2002; Lai and Shih, 2013), especially for the one with multiple demand and/or schedule scenarios. Column Generation can be applied to the scenarios with a larger number of prospective siding plans. The algorithm incrementally increases the number of siding project combinations considered in an efficient way to reduce the computational resources and time required.

Besides an improved version of OSLM that can consider multiple train schedules directly, another approach to solve the siding location problem for the case of multiple of flexible schedules is to iteratively solve OSLM for each different schedule. Iteratively solving OSLM for different schedules would generate multiple siding location plans (each optimal for that schedule) that could be compared for commonalities. Siding locations that appear in multiple plans, or in the plans associated with the most common train schedules, could be prioritized over those locations that are only optimal for rare schedules and operating conditions. Weighting or filtering algorithms could be developed to aid in developing a final siding plan based on the numerous plans generated by each instance of OSLM. The major drawback to this concept is the 1 to 8-hour solution time of the OSLM for each fixed train schedule. Iteratively generating plans for numerous schedules would be computationally intensive.

An alternative to the lengthy process of repeatedly solving OSLM is the screening tool developed by Shih et al. (2016a). The screening tool was developed based on the concept of traffic conflict analysis that was originally inspired by the "root cause analysis" proposed by White (2005). The screening tool can be used to prioritize the infrastructure capacity projects on a single-track mainline based on the distribution of unresolved traffic conflicts. The calculation of traffic conflicts can consider the impact of multiple schedules and the departure-time flexibility associated with each train through a Monte Carlo simulation process. This process is effectively similar to repeatedly solving OSLM for a series of different train schedules. Since the screening tool does not resolve the train conflicts, each iteration of the Monte Carlo process requires far less computation time compared to OSLM. Hundreds of random schedules can be considered by the screening tool in the time required to solve one scenario of the OSLM. By considering a large number of flexible or predefined schedules, the screening tool may potentially provide a more optimized siding location result for mainlines with multiple schedules, or mainlines under mixed operation, compared to OSLM with a single train schedule. More details of a possible screening tool will be discussed in the future research section of the last chapter.

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH

Infrastructure and the corresponding capacity of different portions of the North American rail network vary widely. Each segment is closely tuned to the type and volume of traffic it handles. Changes in rail traffic patterns, in particular growth on some segments, means that network capacity must be selectively expanded. Operating characteristics of the different types of trains needed to accommodate this growth vary depending on their traffic characteristics and demands. The mix of train types on any given line affects both capacity and the expansion strategy. Single-track lines with sparse sidings comprise a portion of the network and pose particular questions regarding the most effective strategies for expanding their capacity. This dissertation developed models to assess the impact of traffic heterogeneity, including operating style, on capacity and train delay performance of these single-track lines. Beyond that, this research expands our understanding of the effect of traffic heterogeneity on freight railroad operations and capacity.

5.1. Conclusion

Properly matching railway line capacity to traffic demand can avoid unnecessary expense and use of resources. Existing processes, models, and tools developed to assist the planning of rail capacity are not well-suited to address the current changes in traffic mixture and characteristics on some single-track lines with historically low traffic density and infrequent passing sidings. In this dissertation I introduce the concept of "operating style" as a new dimension of traffic heterogeneity, and then build processes and tools for capacity evaluation or planning of single-track lines under representative operating styles.

The concept of schedule flexibility and operating style proposed in Chapter 1 was used to extend and redefine the terms “structured operation” and “improvised operation” introduced by Martland (2010). Operating style is defined as the variation of schedule flexibility among the various trains operating on a given rail corridor. Schedule flexibility is a property of a single train comprised of its departure and trip time flexibility (Dick and Mussanov, 2016; Shih et al., 2016a; 2016b). The importance of considering the combined impact of operating style and variability in train priority and speed when evaluating the capacity of a single-track line was also emphasized in the introduction. By considering all three factors, practitioners and researchers can better capture the characteristics of heterogeneous operations on shared-trackage corridors and gain insight to the capacity constraints experienced by freight and passenger rail transportation in North America.

In this dissertation I introduce two new approaches to evaluate and compare line capacity and performance. I used the capacity evaluation technique developed in Chapter 2 to measure the extra capacity demand due to variation in priority, speed and LOS across multiple train types. A strategy and project alternative comparison process was formalized in Chapter 3 and demonstrated in the context of planning capacity expansion of single-track lines. Through case studies, the two approaches developed through this research were used to further understand the fundamental relationships between mixed or flexible operation and railway traffic performance.

The results presented in Chapter 2 suggest that for a fixed volume of traffic in mixed or flexible operation, the capacity available to handle additional trains will vary with the traffic mixture. Depending on the properties of the route, there exists an optimal traffic mixture that maximizes available capacity. While previous research on Base Train Equivalents (Lai et al., 2012) demonstrated that different train types consume different amounts of capacity, this research further demonstrated that the capacity consumed by each incremental train added to the route is also a function of the current traffic mixture. It is important to consider this when evaluating the incremental impact of plans to introduce additional traffic volume on a route.

By definition, structured operation aims to adhere closely to a preplanned schedule so the locations of traffic conflicts are more stable than under mixed or flexible operations. This stability allows an optimal number and locations of passing sidings to be determined for a given train schedule. In Chapter 4, I present an optimization model to identify the location and number of additional sidings on a single-track line. Railroads can use this model to assist with decisions on investments in new and expanded passing sidings. The model can also be used as a prototype for development of other optimal infrastructure location and dispatching models; however, it does not consider departure or travel time randomness. Consequently it is less applicable to mixed corridors where freight trains share track with passenger trains. This model is also unable to suggest an infrastructure expansion plan for a mainline with multiple schedules. The discussion section of Chapter 4 suggests several methods for constructing an improved, optimal siding-location model that could consider multiple schedules.

To better understand and improve approaches to current rail line capacity problems in North America, I propose the concept of operating style and consider its impact on the capacity of single-track mainlines along with priority and speed variation. I developed and tested techniques to gain a more comprehensive understanding of interactions between operating style, line capacity and variability in train priority and speed. These approaches can help improve the quality of operations and capacity planning on North American railroads. Overall, this dissertation advances the understanding of rail traffic heterogeneity and the number of practical tools to assist railroads in improving the quality and efficiency of their service and infrastructure planning.

5.2. Future Study

In Chapter 2, a set of indices was suggested for quantifying the impact of train heterogeneity on train delay and line capacity. Traffic conflict analysis is another potential way to quantify the impact of traffic heterogeneity (including variability in priority, speed, and schedule flexibility) on train delay and line capacity. Traffic conflicts can be viewed from the perspective of a specific location on the rail corridor, or of a specific train. From the rail corridor perspective, locations where traffic conflicts accumulate require more track infrastructure to resolve the train conflicts and support fluid train operations (White, 2005; Williams, 2011). From the perspective of a specific train, the number of conflicts a train encounters during its trip has been shown to be positively correlated with the delay experienced by that train (Gorman, 2009).

These two perspectives suggest approaches for evaluating capacity expansion projects or predicting train delay through traffic conflict analysis. Location-based traffic conflict analysis has the potential to evaluate capacity expansion projects by analyzing the distribution of traffic conflicts along the mainline (Shih et al. 2016a). Train-based traffic conflict analysis has the potential to use the number and type of traffic conflicts encountered by a train along its route to predict its delay and arrival time distribution (Shih et al. 2016b). The following subsections introduce these two theoretical approaches in more detail.

5.2.1. Location-based Traffic Conflict Analysis

Location-based traffic conflict analysis may be able to be used to detect capacity constraints on a single-track line. A "root cause analysis" (White, 2005; Lee et al., 2016) of a single traffic conflict along a single-track mainline (Figure 5.1) can demonstrate this concept. Simulation approaches such as RTC calculate train delays that tend to accumulate where trains wait at passing sidings on single track (Figure 5.1a). Although the delay indicates capacity is constrained, it does not indicate where additional track infrastructure is required. The root cause analysis can be used to examine the original unresolved train paths (Figure 5.1b) to determine the actual conflict location where additional infrastructure is needed to resolve the conflict with a minimum of delay. When multiple trains are considered, zones with a larger cumulative number of traffic conflicts can be identified as candidate locations for capacity expansion projects.

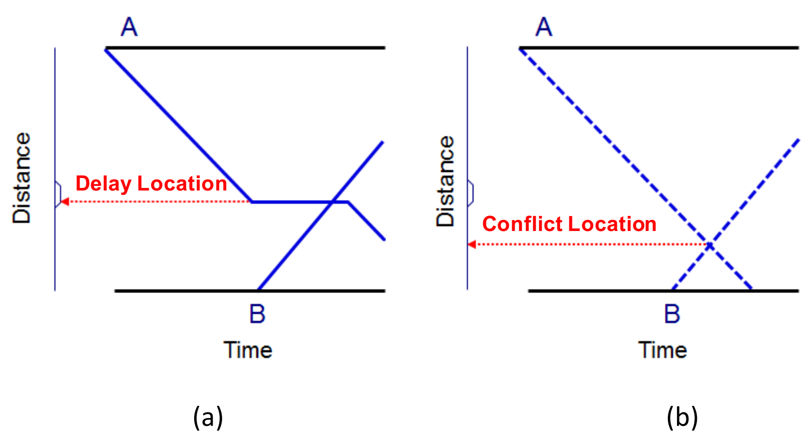


Figure 5.1. Comparison between identifying (a) delay locations or (b) conflict locations

Expanding on this idea, a conceptual framework is proposed for a screening tool based on the concept of traffic conflict analysis (Figure 5.2). The process requires a predetermined train operating plan with or without schedule flexibility. The user would define zones along the mainline under study, then, based on the inputs, the screening tool uses a Monte-Carlo process to calculate the cumulative traffic conflicts per zone during a set period of train operation (Shih et al., 2016a).

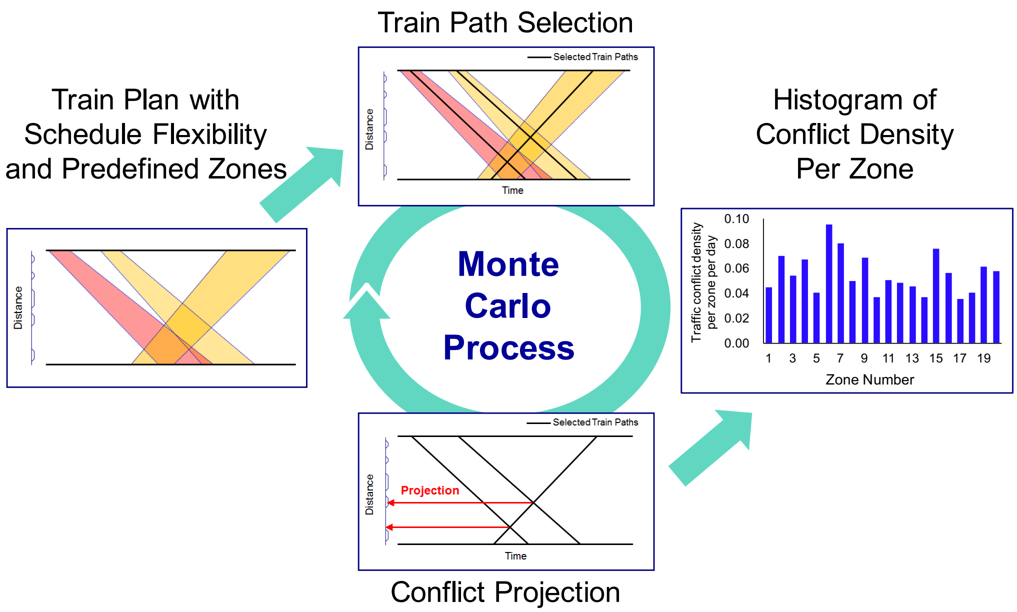


Figure 5.2. Flowchart of the Capacity Screening Tool

Preliminary investigation of a hypothetical single-track line with the screening tool suggested that the capacity expansion plan generated by the screening tool had equivalent performance (in terms of average train delay) compared to a detailed simulation method. A more comprehensive comparison of outputs from the screening tool and the detailed simulation method under different single-track layouts is required to validate the performance of the screening tool and confirm it can be applied as a more general approach for capacity planning.

A similar Monte Carlo process is also proposed as a supplement to OSLM in Chapter 4 to determine the locations of additional sidings when there are multiple major train schedules. The Monte Carlo method is hypothesized to return a near-optimal suggested infrastructure location plan if the probability of each schedule is precisely known and quantified.

5.2.2. Train-based Traffic Conflict Analysis

Train-based traffic conflict analysis is proposed to quantify the characteristics of heterogeneous traffic under different operating styles and describe the relationship between heterogeneity and train delay. The concept of using train conflicts (also referred to as traffic conflicts) to better predict train performance was formalized by Gorman (2009). He used historical delay data from ten single-track freight lines to test the relationship between train running time and various operating and infrastructure factors. Gorman found that traffic conflicts, represented by the number of meets, passes and overtakes, significantly affects train delay.

Gorman did not connect the indices he used to traffic heterogeneity. Based on Gorman's findings and preliminary investigations conducted in parallel with the main elements of this dissertation, three indices are proposed to capture the impact of traffic heterogeneity on the train delay distribution:

- Total Conflicts (TC) considers all of the potential conflicts a train could encounter during its trip. A larger number of traffic conflicts increases the difficulty of train dispatching. This index is also an analog to traffic volume since higher train volumes usually lead to more train conflicts. To simplify calculations, TC does not include potential conflicts with additional trains due to trip time flexibility.
- Adjusted Train Priority (ATP) quantifies the actual priority of a train within the given traffic mixture on the route. ATP is calculated for a target train by the summation of inferior conflicts (target train has inferior priority relative to the conflicting train) and half of equal conflicts (target train has equal priority to the conflicting train). In

previous studies, the assigned priority of a train was a static ordinal value based on its train type. The actual priority of a train should be a dynamic value since it varies with the traffic mixture. For example, the actual priority of an intermodal train within traffic composed of 80 percent inferior trains should be higher than the relative priority of the same train within traffic composed of only 20 percent inferior trains. The physical interpretation of ATP as a delay mechanic is the percentage of conflicts where the target train will need to stop and wait for the other conflicting train.

- Inferior Pass (IP) represents the impact of train speed heterogeneity on train conflicts and delay. IP calculates the expected number of inferior passes (target train has inferior priority to passing train). When IP is high there is a greater diversity in train speed and meets make up a smaller share of train conflicts. This is in comparison to cases where speed is homogeneous and all train conflicts are meets. The physical meaning behind IP is the expected number of passes that will cause the target train to stop or encounter delay. Delays for passes are assumed to be the origin of extra delay caused by train speed heterogeneity.

Most of the current capacity evaluation tools only predict average train delay. Using average train delay to predict train trip and cycle time as part of freight operations planning does not consider the full impact of train operation randomness. A plan based on a cycle time calculated with average train delay may frequently fail as it is fragile to the stochastic railroad operating environment. A model to predict the distribution of train delay can improve the reliability of freight operating plans and increase the stability of the system.

Quantile regression (Koenker and Bassett, 1978; Koenker, 2005) may be an appropriate technique for developing relationships to predict the distribution of individual train delays using the three indices introduced earlier in this section. Preliminary work conducted in parallel with this dissertation suggests that a quantile regression model can successfully predict the distribution of train delays but its accuracy is limited to lower quantiles of train delay. The ability of the model to predict extremely long train delays still needs to be improved. Investigating the types of conflicts experienced by trains with significant delay may help identify new indices that better capture the characteristics of trains with extreme delay and improve model performance. Methods other than quantile regression can also be tested for their ability to capture the distribution of train delays.

5.2.3. Other Possible Directions

The conduct of this dissertation research suggests other possible future directions. The transformation process used in Chapter 2 can be applied to different regression models. For example, it can be used to extend the study of Dick and Mussanov (2016) to investigate the interaction between schedule flexibility, traffic mixture, and LOS (Mussanov et al., 2017). For the reliability analysis in Chapter 3, the application of quantile regression can help construct a more systematic reliability analysis since it can provide more comprehensive information related to train delay. Several potential methods for improving OSLM were already addressed in the future study section of Chapter 4.

Finally, the concept of traffic conflict analysis shows promise to be the basic concept behind the development of processes, tools, or models to investigate the complex relationships between variability in priority, speed, schedule flexibility, and line capacity. While two ideas were proposed earlier in this chapter, additional concepts based on traffic conflict analysis are likely to be developed in the future.

REFERENCES

Abril, M., F. Barber, L. Ingolotti, and M. Salido. 2008. An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, 44(5): 774-806.

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

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

Association of American Railroads (AAR). 2015. *Railroad Facts 2015*. AAR, Washington, DC, USA.

American Railway Engineering and Maintenance-of-Way Association (AREMA). 2013. *Manual for Railway Engineering*, Chapter 16, Part 1: Railway Location. The American Railway Engineering and Maintenance-of-Way Association, Lanham, MD, USA.

Atanassov, I., C.T. Dick, and C.P.L. Barkan. 2014. Siding spacing and the incremental capacity of the transition from single to double track. In: *Proceedings of the ASME 2014 Joint Rail Conference*. American Society of Mechanical Engineers (ASME). Colorado Springs, CO, USA.

Atanassov, I.H. 2015. *Influence of Track Arrangement on Expanding Rail Corridor Capacity and Operations*. Master's Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.

Atanassov, I. and C.T. Dick. 2015. Capacity of single-track railway lines with short sidings to support operation of long freight trains. *Transportation Research Record: Journal of the Transportation Research Board*, 2475: 95-101.

Bing, A.J., E.W. Beshers, M. Chazvez, and D.P. Simpson. 2010. *Guidebook for Implementing Passenger Rail Service on Shared Passenger and Freight Corridors*. National Cooperative Highway Research Program, NCHRP Report 657, ISSN 0077-5614. Washington, DC, USA.

Birge, J.R. 1985. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33(5): 989-1007.

BNSF. 2012. *Railway*. BNSF, Fort Worth, TX, USA.

BNSF. 2016. *BNSF Railway Network Update*. BNSF, Fort Worth, TX, USA.

Bonsra, K., and J. Harbolovic. 2012. *Estimation of Run Time in a Freight Rail Transportation Network*. The MIT Press, Cambridge, MA, USA.

Box, G.E.P., and K.B. Wilson. 1951. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society Series B*, 13(1): 1-45.

Box, G.E.P., and B. Soren. 1987. The scientific context of quality improvement. *Quality Progress*, 20(6): 54-61.

Brännlund, U., P.O. Lindberg, A. Nou, and J.E. Nilsson. 1998. Railway timetabling using Lagrangian relaxation. *Transportation Science*, 32(4): 358-369.

Bronzini, M.S., and D.B. Clarke. 1985. Estimating rail line capacity and delay by computer simulation. *Journal of the Transportation Research Forum*, 2(1): 5-11.

Burdett, R.L., and E. Kozan. 2006. Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological*, 40(8): 616–632.

Burnham, K.P., and D. Anderson. 2003. *Model Selection and Multi-model Inference: A Practical Information-theoretic Approach*. Springer Science and Business Media, New York, NY, USA.

Burnham, K. P., and D.R. Anderson. 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2): 261-304.

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

Carey, M. 1999. Ex ante heuristic measures of schedule reliability. *Transportation Research Part B*, 33(7): 473-494

Chaudhry, S.S., and W. Luo. 2005. Application of genetic algorithms in production and operations management: a review. *International Journal of Production Research*, 43(19): 4083-4101.

Chen, B., and P.T. Harker. 1990. Two moments estimation of the delay on single-track rail lines with schedule traffic. *Transportation Science*, 24(4): 261-275.

Conejo, A.J., E. Castillo, R. Minguez, and R. Garcia-Bertrand. 2006. *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Springer Science and Business Media, New York, NY, USA.

Cordeau, J.F., P. Toth, and D. Vigo. 1998. A survey of optimization models for train routing and scheduling. *Transportation Science*, 32(4): 380-404.

Costa, A.M. 2005. A survey on benders decomposition applied to fixed-charge network design problems. *Computers and Operations Research*, 32(6): 1429-1450.

Crainic, T., J.A. Ferland, and J.M. Rousseau. 1984. A tactical planning model for rail freight transportation. *Transportation Science*, 18(2): 165-184.

Dick, C.T. and D. Mussanov. 2016. Operational schedule flexibility and infrastructure investment: capacity trade-off on single-track railways. *Transportation Research Record: Journal of the Transportation Research Board*, 2546: 1-8.

Dingler, M., Y.C. Lai, and C.P.L. Barkan. 2009. Impact of train type heterogeneity on single-track railway capacity. *Transportation Research Record: Journal of the Transportation Research Board*, 2117: 41-49.

Dingler, M.H., Y.C. Lai, and C.P.L. Barkan. 2013. Mitigating train-type heterogeneity on a single-track line. *Journal of Rail and Rapid Transit*, 227(2): 140-147.

Ekman, J. 2004. Capacity estimation of new infrastructure based on a discrete event model of train path. In: J. Allan, C.A. Brebbia, R.J. Hill, G. Sciutto, S. Sone, (Eds.) *Computers in Railways IX: Computer Aided Design, Manufacture and Operation in the Railway and Other Advanced Mass Transit Systems*. The WIT Press, Ashurst, UK, pp. 593-548.

Fransoo, C., and J. Bertranda. 2000. Aggregate capacity estimation model for the evaluation of railroad passing constructions. *Transportation Research Part A: Policy and Practice*, 34(1): 35-49.

Gorman, M.F. 2009. Statistical estimation of railroad congestion delay. *Transportation Research Part E: Logistics and Transportation Review*, 45(3): 446-456.

Harrod, S. 2009. Capacity factors of a mixed speed railway network. *Transportation Research Part E: Logistics and Transportation Review*, 45(5): 830-841.

HDR, and Transit Safety Management. 2006. *Statewide Rail Capacity and System Needs Study Task 3-Rail Capacity Needs and Constraints*. HDR, Inc., Omaha, NE, USA.

Heyman, D.P., and M.J. Sobel. 2003. *Stochastic Models in Operations Research: Stochastic Optimization*. Courier Corporation, North Chelmsford, MA, USA.

Higgins, A., E. Kozan, and L. Ferreira. 1996. Optimal scheduling of trains on a single line track. *Transportation Research Part B: Methodological*, 30(2): 147-161.

Higgins, A., E. Kozan, and L. Ferreira. 1997. Modeling the number and location of sidings on a single line railway. *Computers and Operations Research*, 24(3): 209-220.

Higgins, A., and E. Kozan. 1998. Modeling train delays in urban networks. *Transportation Science*, 32(4): 346-357.

Jaumard, B., T.H. Le, H. Tian, A. Akgunduz, and P. Finnie. 2013. An enhanced optimization model for scheduling freight trains. In: *Proceedings of the ASME 2013 Joint Rail Conference*. American Society of Mechanical Engineers (ASME). Knoxville, TN, USA.

Jensen, L.W., A. Landex, O.A. Nielsen, L.G. Kroon, and M. Schmidt. 2017. Strategic assessment of capacity consumption in railway networks: framework and model. *Transportation Research Part C: Emerging Technologies*, 74:126-149.

Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi. 1983. Optimization by simulated annealing. *Science*, 220(4598): 671-680.

Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica*, 46: 33-50.

Koenker, R., 2005. *Quantile Regression*. Cambridge university press, Cambridge, UK.

Krueger, H. 1999. Parametric modeling in rail capacity planning. In: *Proceedings of the 1999 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers (IEEE). Phoenix, AZ, USA, pp. 1,194-1,200.

Lai, Y.C. 2008. *Increasing Railway Efficiency and Capacity Through Improved Operations, Control and Planning*. Doctoral Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.

Lai, Y.C., M.H. Dingler, C.E. Hsu, and P.C. Chiang. 2010a. Optimizing train network routing with heterogeneous traffic. *Transportation Research Record: Journal of the Transportation Research Board*, 2159: 69-76.

Lai, Y.C., M.C. Shih, and J.C. Jong. 2010b. Railway capacity model and decision support process for strategic capacity planning. *Transportation Research Record: Journal of the Transportation Research Board*, 2197: 19-28.

Lai, Y.C., and C.P.L. Barkan. 2011. Comprehensive decision support framework for strategic railway capacity planning. *Journal of Transportation Engineering*, 137(10): 738-749.

Lai, Y.C., Y.H. Liu, and T.Y. Lin. 2012. Development of base train equivalents to standardize trains for capacity analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 2289: 119-125.

Lai, Y.C. and M.C. Shih. 2013. A stochastic multi-period investment selection model to optimize strategic railway capacity planning. *Journal of Advanced Transportation*, 47(3): 281-296.

Lamorgese, L., and C. Mannino. 2013. The track formulation for the train dispatching problem. *Electronic Notes in Discrete Mathematics*, 41: 559-566.

Landex, A., A.H. Kaas, E.M. Jacobsen, and J. Schneider-Tilli. 2007. The UIC 406 capacity method used on single track sections. In: *Proceeding of the International Association of Railway Operations Research (IAROR) 2nd International Seminar on Railway Operations Modelling and Analysis*. Hannover, Germany.

Landex, A. 2008. *Methods to Estimate Railway Capacity and Passenger Delays*. Doctoral Thesis, Technical University of Denmark, Department of Transport. Lyngby, Denmark.

Lee, W.H., L.H. Yen, and C.M. Chou. 2016. A delay root cause discovery and timetable adjustment model for enhancing the punctuality of railway services. *Transportation Research Part C: Emerging Technologies*, 73: 49-64.

Leilich, R.H. 1998. Application of simulation models in capacity constrained rail corridors. In: *Proceedings of 30th Winter Simulation Conference*. Los Alamito, CA, USA, pp. 1,125-1,133.

Lindner, T. 2011. Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. *Journal of Rail Transport Planning and Management*, 1(1): 49-57.

Lindfeldt, O. 2007. *Quality on Single-Track Railway Lines With Passenger Traffic: Analytical Model for Evaluation of Crossing Stations and Partial Double-Tracks*. Licentiate Thesis. KTH Royal Institute of Technology, Department of Transport and Economics. Stockholm, Sweden.

Lindfeldt, O. 2009. Validation of a simulation model for mixed traffic on a Swedish double-track railway line. In: *Proceedings of Railway Engineering 10th International Conference and Exhibition*. London, UK.

Lindfeldt, A. 2012a. *Congested Railways: Influence of Infrastructure and Timetable Properties on Delay Propagation*. Licentiate Thesis. KTH Royal Institute of Technology, Department of Transport and Economics. Stockholm, Sweden.

Lindfeldt, O. 2012b. From single to double track: effects of alternative extension measure. In: A. Brebbia, N. Tomii, J.M. Mera, B. Ning, P. Tzieropoulos (Eds.) *Computers in Railways XIII: Computer System Design and Operation in the Railway and Other Transit Systems*. WIT Press, Ashurst, UK, pp. 313-334.

Liu, S.Q., and E. Kozan. 2011. Optimising a coal rail network under capacity constraints. *Flexible Services and Manufacturing Journal*, 23(2): 90-110.

Lovett, A.H., C.T. Dick, and C.P.L. Barkan. 2015. Determining freight train delay costs on railroad lines in North America. In: *Proceedings of the International Association of Railway*

Operations Research (IAROR) 6th International Conference on Railway Operations Modelling and Analysis. Tokyo, Japan.

Luethi, M., A. Nash, F. Laube, R. Wuest, and U. Weidmann. 2007. *Increasing Railway Capacity and reliability through integrated real-time rescheduling*. URL https://www.researchgate.net/profile/Andrew_Nash/publication/268297119_Increasing_Railway_Capacity_and_Reliability_through_Integrated_Real-Time_Rescheduling/links/54d1ef5d0cf28370d0e171eb.pdf. Accessed: 2016-03-02.

Marin, A., and J. Salmerón. 1996. Tactical design of rail freight networks. Part I: Exact and heuristic methods. *European Journal of Operational Research*, 90(1): 26-44.

Martland, C.D. 2010. Improving on-time performance for long-distance passenger trains operating on freight routes. *Journal of the Transportation Research Forum*, 47(4): 63-80.

Mattsson, L.G. 2007. *Railway Capacity and Train Delay Relationships*. Springer, Berlin, Germany.

Mitra, S., and D. Tolliver. 2010. Estimation of railroad capacity using parametric methods. *Journal of the Transportation Research Forum*, 49(2): 111-126.

Montgomery, D.C. 1984. *Design and Analysis of Experiments*, 2nd ed. John Wiley and Sons, New York, NY, USA.

Murali, P., M. Dessouky, F. Ordonez, K. Palmer. 2010. A delay estimation technique for single and double-track railroads. *Transportation Research Part E: Logistics and Transportation Review*, 46: 483-495.

Mussanov, D., N. Nishio and C.T. Dick. 2017. Delay performance of different train types under combinations of structured and flexible operations on single-track railway lines in North America. In: *Proceedings of the International Association of Railway Operations Research (IAROR) 7th International Conference on Railway Operations Modelling and Analysis*, Lille, France, 2017.

Norio, T., T. Yoshiaki, T. Noriyuki, H. Chikara, and M. Kunimitsu. 2005. Train rescheduling algorithm which minimizes passengers' dissatisfaction. *Innovations in Applied Artificial Intelligence*, 3533: 829-838.

Paragon Decision Technology. 2006. *The AIMMS User's Guide*. Paragon Decision Technology, Bellevue, WA, USA.

Parkinson, T., and I. Fisher. 1996. *TCRP Report 13*. National Academy Press, Washington, DC, USA.

Pawar, S.P.S. 2011. *An Analysis of Single Track High Speed Rail Operation*. Master's Thesis, University of Birmingham, Department of Railway Systems Engineering and Integration. Birmingham, UK.

Petersen, E.R. 1982. A structured model for rail line simulation. *Transportation Science*, 8(1): 192-206.

Petersen, E., and A. Taylor. 1987. Design of single-track rail line for high-speed trains. *Transportation Research Part A: General*, 21(1): 47-57.

Pouryousef, H., P. Lautala, and T. White. 2013. Review of capacity measurement methodologies; similarities and differences in the U.S. and European railroads. In: *Proceedings of the 92nd Annual Meeting of the Transportation Research Board of the National Academies*, Washington, DC, USA.

Pouryousef, H., P. Lautala, and T. White. 2015. Railroad capacity tools and methodologies in the US and Europe. *Journal of Modern Transportation*, 23(1): 30-42.

Prokopy, J.C., and R.B. Rubin. 1975. *Parametric Analysis of Railway Line Capacity*. Federal Railroad Administration (FRA), FRA Report OPPD-75-1. Washington, DC, USA.

Qiang, L.S., and E. Kozan. 2009. Scheduling trains as a blocking parallel-machine job shop scheduling problem. *Computers and Operations Research,* 36(10): 2,840-2,852.

Salido, M.A., F. Barber, and L. Ingolotti. 2012. Robustness for a single railway line: analytical and simulation methods. *Expert Systems with Applications*, 39(18): 13,305-13,327.

SAS Institute Inc. 2007. *Start Statistics: A Guide to Statistics and Data Analysis Using JMP®*, 6th ed. SAS Institute, Cary, NC, USA.

Sherali, H.D., and B.M. Fraticelli. 2002. A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization*, *22*(4): 319-342.

Shih, M.C., C.T. Dick, S. Sogin, and C.P.L. Barkan. 2014a. Comparison of capacity expansion strategies for single-track railway lines with sparse sidings. *Transportation Research Record: Journal of the Transportation Research Board*, 2448: 53-61.

Shih, M.C., Y.C. Lai, C. Dick, and M.H. Wu. 2014b. Optimization of siding location for single-track lines. *Transportation Research Record: Journal of the Transportation Research Board*, 2448: 71-79.

Shih, M.C., C.T. Dick, and C.P.L. Barkan. 2015a. Impact of passenger train capacity and level of service on shared rail corridors with multiple types of freight trains. *Transportation Research Record: Journal of the Transportation Research Board*, 2475: 63-71.

Shih, M.-C., C.T. Dick, and Y.-C. Lai. 2015b. Optimizing location and length of passing sidings on single-track lines for long heavy-haul freight trains. In: *Proceedings of the 11th International Heavy Haul Association (IHHA) Conference*, Perth, Australia.

Shih, M.-C., C.T. Dick, and Y.-C. Lai. 2015c. Optimization of siding location for single-track lines with non-uniform track speed. In: *Proceedings of the International Association of Railway Operations Research (IAROR) 6th International Conference on Railway Operations Modelling and Analysis*, Tokyo, Japan.

Shih, M.-C., P.Y. Liao, and C.T. Dick. 2016a. A screening tool to identifying mainline capacity constraint under mixed heterogeneous rail operation. In: *Proceedings of the ASME 2016 Joint Rail Conference*. American Society of Mechanical Engineers (ASME), Columbia, SC, USA.

Shih, M.-C., C.T. Dick, and C.P.L. Barkan. 2016b. A parametric model of the train delay distribution based on meet and pass conflicts. In: *Proceedings of INFORMS Nashville 2016 Annual Meeting*, Nashville, TN, USA.

Sipilä, H. 2015. *Simulation of Rail Traffic: Methods for Timetable Construction, Delay Modeling and Infrastructure Evaluation*. Doctoral Thesis, KTH Royal Institute of Technology, Department of Transport and Economics. Stockholm, Sweden.

Sogin, S.L. 2013. *Simulations of Mixed Use Rail Corridors: How Infrastructure Affects Interactions Among Train Types*. Master's Thesis, University of Illinois at Urbana-Champaign, Department of Civil and Environmental Engineering. Urbana, IL, USA.

Sogin, S.L., Y.C. Lai, C. Dick, and C.P.L. Barkan. 2013a. Comparison of capacity of single- and double-track rail lines. *Transportation Research Record: Journal of the Transportation Research Board*, 2374: 111-118.

Sogin, S., C.T. Dick, Y-C. Lai and C.P.L. Barkan. 2013b. Analyzing the progression from single to double track networks. In: *Proceedings of the ASME 2013 Joint Rail Conference*. American Society of Mechanical Engineers (ASME). Knoxville, TN, USA.

Stenstrom, C., A. Parida, D. Galar, and U. Kumar. 2013. Link and effect model for performance improvement of railway infrastructure. *Journal of Rail and Rapid Transit*, 227(4): 392-402.

Törnquist, J. 2006. Computer-based decision support for railway traffic scheduling and dispatching: A review of models and algorithms. In: *Proceedings of the OASIcs-OpenAccess Series in Informatics*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, Wadern, Germany.

Union Internationale des Chemins de fer/International Union of Railways (UIC). 2004. *UIC CODE 406R-Capacity 1st ed*. UIC, Paris, France.

Union Internationale des Chemins de fer/International Union of Railways (UIC). 2013. *UIC CODE 406R-Capacity 2nd ed*. UIC, Paris, France.

Vantuono, W.C. 2005. Capacity Is Where You Find It: How BNSF Balances Infrastructure and Operations. *Railway Age*, 206(6):17-24.

Voß, S., S. Martello, I.H. Osman, and C. Roucairol. 2012. *Meta-heuristics: Advances and Trends in Local Search Paradigms for Optimization*. Springer Science and Business Media, Berlin, German.

Vromans, M., R. Dekker, and L.G. Kroon. 2004. Reliability and heterogeneity of railway services. *European Journal of Operational Research*, 172(2): 647-665.

Vromans, M. 2005. *Reliability of Railway Systems*. Doctoral Thesis, Erasmus University, Research Institute of Management, Rotterdam, Denmark.

Wanek-Libman, M. 2013. *Railway Track and Structure*. Simmons-Boardman Publishing Corporation, New York, NY, USA.

White, T. 2005. Alternatives for railroad traffic simulation analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 1916: 34-41.

Wilhelm, W.E. 2001. A technical review of column generation in integer programming. *Optimization and Engineering*, 2(2): 159-200.

Williams, M.K. 2011. Using Simulation to Understand Bottlenecks, Delay Accumulation, and Rail Network Flow. In: *Proceedings of the 2011 Annual AREMA Conference*. American Railway Engineering and Maintenance-of-Way Association (AREMA). Minneapolis, MN, USA.

Wilson E. 2012. *Rail Traffic Controller (RTC).* Berkeley Simulation Software, Berkeley, CA, USA.

Zhao, J., Q. Peng, C. Wen, and J. Xu. 2010. Local neighborhood search algorithm for generalized dynamic wagon-flow allocation of railway technical stations. *Journal of Southwest Jiaotong University*, 3: 30-38.