EFFECT OF TRAIN LENGTH ON RAILROAD ACCIDENTS AND A QUANTITATIVE ANALYSIS OF FACTORS AFFECTING BROKEN RAILS

BY

DARWIN H. SCHAFER II

B.S., University of Illinois at Urbana-Champaign, 2006

THESIS

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

Urbana, Illinois

Adviser:

Professor Christopher P.L. Barkan

ABSTRACT

Continuous improvement in safety is an ongoing goal of the railroad industry and a critically important element of this is reducing the number of train accidents. Accomplishing this efficiently requires understanding of the factors that contribute to accident occurrence. Train accident rates are an important metric of railroad transportation safety and risk. These rates have been the subject of a number of analyses but they have generally not considered the effect of train length. In this thesis, Federal Railroad Administration (FRA) accident data were used to develop a new quantitative metric to classify FRA accident causes as either train-mile or car-mile-related and the results were used to revise a previous classification of these causes. These reclassified causes were then incorporated into a model to calculate new train-length dependent accident rates. A sensitivity analysis was conducted to investigate the effects of changes in train length on an individual train's accident likelihood and on the system-wide accident rate.

The second major focus of this thesis was an analysis of factors related to the occurrence of broken rails. Broken rails are the leading cause of major derailments and hazardous material release accidents on U.S. Class I railroads. From 2003 through 2006 there was an average of 84 mainline broken-rail derailments per year with an average track and equipment damage cost of $525,000 per incident. In the last ten years their annual frequency has increased 18%; consequently, efforts to reduce their occurrence are increasingly important. The purpose of this study was to examine the factors that influence the occurrence of broken rails and develop a predictive tool that will enable railroads to identify locations that have a high probability of broken rail occurrence. The factors that were considered included rail characteristics, infrastructure features, maintenance activity, operational information, and rail testing results. To analyze the factors related to broken rails two modeling techniques were applied, one using statistical regression and the other employing artificial neural networks (ANN). Several variations of the logistic regression (LR) and ANN techniques were developed, including hybrid LR/ANN models. The accuracy and practicality of the models were evaluated and compared. A "practical" logistic regression model was developed that used only the top

eight related factors and was able to identify broken rail locations with approximately 65% accuracy.

Finally, to assist with decisions regarding broken rail prevention, the economic impact of broken rails was also studied. This included the associated costs of broken rail service failures and derailments, as well as the cost of typical prevention measures. A train delay calculator was developed based on railroad industry operating averages. The results and methodologies presented in this thesis are intended to help railroads better understand the factors contributing to the occurrence and severity of railroad accidents and more effectively allocate resources to improve their safety and risk management efforts.

To Jessica

ACKNOWLEDGMENTS

I would first like to thank all of those who continued to encourage me throughout the writing of this thesis and my graduate work, most especially Jessica and my parents. Jessica, you were a constant source of energy and support during my college studies and most importantly the last two years. I could have not completed this thesis without both your understanding and commitment, and I very much look forward to the rest of our lives together. I also owe my deepest thanks to my family for their unwavering support for my dreams and goals. Many times when I doubted myself and my abilities, you have been there with kind and supportive words to help me believe in myself and to keep me on track. I should also apologize and thank you for your understanding during this time, as I know I have not visited or called often enough. This would not have been possible without the love given to me from each of you.

During my studies, I have had significant assistance from University of Illinois students as well as railroad industry experts. I am grateful to current and past graduate students: Daniel Peltier, Yung-Cheng Lai, M. Rapik Saat, Mark Dingler, Robert Anderson, C. Tyler Dick, G. Avery Grimes, J. Riley Edwards and Jeremiah Dirnberger. Each of you contributed significantly to both the completion of this research and my growth as an industry professional. I would also like to thank the undergraduates who assisted me with this work, including John Zeman, Mike Wnek, and Billy Lau. A significant amount of information contained in this thesis was contributed by those at the BNSF railway, including Larry Milhon, Paul Gabler, Mike Mischke, Manal Bishr, Michael Armstrong, Dennis Morgart, Mark Stehly, and Ed Chapman. I would also like to thank Dale Mellor from the UP Railroad, David Ferryman from the CN Railway, Chuck Gullakson and Michael Hillfor from CSX Transportation, Tom Guins from TTCI, and Todd Treichel from the RSI-AAR Tank Car Safety Project for their assistance and reviews on the topics presented in this thesis. I am also grateful that my research was supported by both a CN Railway fellowship and a BNSF Railway research grant.

Finally, I would like to thank my advisor, Dr. Chris Barkan, for his continued support and guidance throughout this research project. Your knowledge and encouragement will never be forgotten.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The purpose of the research described in this thesis was to improve our understanding of factors related to railroad accidents, particularly broken rail derailments, and provide modeling tools to assist in risk analysis and accident prevention. Improving the safety of rail transportation is an ongoing objective of the railroad industry. This research examines two topics that are particularly timely with regard to railroad operating practices. The first topic is an evaluation of train accident rates and accident causes based on train length. Understanding the effect of train length on accident likelihood is important because as railroad freight traffic increases railroads must either run more trains or longer trains, or both. This research provides insight into the safety implications of both of these practices. The second topic is focused on accident reduction by preventing broken rails. Broken rails are the leading cause of major train accidents and the frequency of broken rail derailments has increased 18% in the last ten years (FRA 2007a). This thesis presents several modeling techniques for the prediction of broken rail locations and an evaluation of costs associated with broken rails.

1.1 Objective and Scope

The first topic considered is the relationship between train length and train accident causes. Accident data and causes from the Federal Railroad Administration (FRA) Office of Safety accident database were analyzed for a 16 year period from 1990 through 2005. During this period, U.S. Class I freight railroads experienced 13,181 reportable mainline and siding accidents (FRA 2007a). Evaluating the average length of trains involved in specific causes provided the ability to develop conclusions in regards to the relationship between a train's length and likelihood of an accident. The results of this study were used to calculate new train-length dependent accident rates.

The second topic, and primary focus, of this thesis is a study to understand the factors related to broken rail service failures and derailments. Broken rails are an increasing concern to the railroad industry due to both their frequency and severity. A service failure in this thesis is defined as a broken rail that does not result in a derailment. Better understanding the factors that contribute to broken rails is necessary for efficient prevention of their occurrence. The analysis of broken rails was divided into three main areas. The first section is an evaluation of previous work as well as a presentation of new predictive modeling techniques. The second section presents a predictive model based on recent service failure data. Finally, the third section summarizes the economic impact to railroads from broken rail service failures, derailments, and prevention measures.

I used data on broken rail service failures and derailments for a four-year period from 2003 through 2006. Class I railroads in the U.S. experienced an average of 84 mainline broken rail derailments per year. The average equipment and track damage cost of a broken rail derailment during the study period was approximately $525,000 (FRA 2007a).

1.2 Organization

This thesis is divided into seven chapters, an introduction, a literature review, four chapters describing the research, and a conclusion. The majority of the research chapters have been presented, published, or are being prepared for publication in various engineering conference proceedings and journals as discussed below. Also included in this thesis is an appendix with further information regarding the prediction models presented. Following is a summary of each of the five main chapters.

Chapter 2:

In this chapter I present a review of previous literature on the topics presented in this thesis. This literature includes work both directly and indirectly related to this research. Some of the references in this section are also cited elsewhere in this thesis; however, this chapter provides a more detailed review of previously published work. The topics surveyed include accident causes, accident rates, fracture defect growth, factors influencing broken rails, statistical modeling techniques, neural network modeling applications, and railroad economic research.

Chapter 3:

Accepted for publication in the Transportation Research Record

This chapter presents an analysis of train accident causes and rates based on train length. Train accident causes were separated into two distinct categories, those related to the train's length (car-mile-related) and those independent of train length (train-mile-related). The decision to dispatch the same number of shipments in fewer longer trains versus more, shorter trains will affect the overall network accident rate. Since some accident causes are correlated with car-miles and others with train-miles, accurate classification of the causes is important to correctly determine the effect of changes on accident rates. In a previous study, all FRA accident causes were combined into 51 unique accident cause groups and classified as either car or train-mile-related. I developed a new metric to quantitatively evaluate each cause group based on accident data. Use of the metric led to a reclassification of 11 cause groups. The new classification was found to be more representative of car and train-mile expectations. Mainline car-mile and train-mile-related accident rates were calculated for Class I freight railroads. These rates were used in a sensitivity analysis to illustrate the effect of changes in train length on accident rate.

Chapter 4:

Presented in part at INFORMS 2007 annual conference and SRA 2007 annual meeting

This chapter is an introduction to the analysis of broken rail service failures and derailments. Broken rail derailments are the second leading cause of train accidents; only grade crossing collisions occur more frequently. This chapter presents initial work evaluating factors leading to broken rails in order to develop a prediction model. A previous study by Dick (2001) used service failure data from the BNSF Railway to develop a logistic regression prediction model. I evaluated the robustness of Dick's model using unseen service failure data. I also developed three new prediction models using artificial neural networks (ANN). The models developed were a stand-alone ANN and two hybrid ANN-logistic regression (ANN/LR) models. I determined that all four models have similar predictive abilities and are robust for unseen data.

Chapter 5:

Presented in part at the Joint Rail Conference in 2008 and the 8th World Congress on Railway Research in 2008

Currently being prepared for submission to the journal Accident Analysis & Prevention

This chapter extends the work presented in the previous chapter using a more recent and comprehensive dataset to attempt to better understand and quantify the factors affecting broken rails. Chapter 4 used service failure data and information on train movement and track information for a two-year period. In this new study I used a four-year period of recent data as well as an expanded dataset to include more possible factors that lead to service failures. The factors that were considered included rail characteristics, infrastructure features, maintenance activity, train operation data, and on-track testing results. Multiple predictive models, using both logistic regression and artificial neural network techniques, were evaluated to determine the factors related to service failures. A "practical" prediction model with a limited number of input parameters was constructed that was both understandable and useable. The models were also tested against unseen data and found to be robust. Finally, the practical prediction model was applied to a hypothetical case study to illustrate the potential use of the model as a maintenance planning tool.

Chapter 6:

Currently being prepared to be presented in part at the AREMA 2008 Annual Conference

This chapter examines the economic impact of broken rails. The purpose is to quantify the cost of broken rails to assist with decision making regarding their prevention. Average track and equipment damages for broken rail derailments for U.S. Class I freight railroads were examined using available FRA accident data. Additionally, average service failure repair cost and broken rail accident clean-up time and cost were determined from data provided by railroad industry experts. Train delay cost was evaluated based on car cost, locomotive cost, fuel cost, labor cost, and traffic density of the line. Based on this information, combined with FRA data, a new, more complete estimate of the cost of broken rails was developed. Finally, the cost of preventive maintenance techniques, such as rail grinding, rail replacement, and track surfacing, were summarized.

CHAPTER 2: LITERATURE REVIEW

In this chapter I present a survey of literature describing previous work completed on the topics presented in this thesis. Many of the references presented here are also cited elsewhere in this thesis; however, this chapter presents a more in-depth description of the most important previous work pertinent to my research. This chapter is divided into four specific sections: accident causes and rates, factors related to broken rails, statistical and neural network modeling techniques, and railroad economic research.

The literature reviewed regarding accident causes and rates is the framework for the analysis conducted in Chapter 3. The previous research conducted on the factors leading to broken rails revealed that crack growth is linked to a number of variables but is also somewhat unpredictable. The literature sources reviewed regarding modeling techniques are the basis for the advanced logistic regression and artificial neural network models presented in Chapters 4 and 5. Finally, the previous work examined on railroad economic research is used as background information for the methodology used to calculate the cost of broken rails in Chapter 6. The literature reviewed in each section is presented starting with the most significant contribution to the work completed in this thesis.

2.1 Accident Causes and Rates

The literature presented in this section was used to develop the premise that the likelihood of a train accident is dependent on train length. ADL (1996) introduced the concept that train accident causes can be separated into two groups, those related to the train length and those independent of train length. Anderson (2005) and Anderson & Barkan (2005a) considered some of the implications of this. I conducted a similar analysis in Chapter 3 of this thesis to analyze how accident causes are affected by train length.

Anderson & Barkan (2005a) conducted an in-depth examination of train accident rates on U.S. freight railroads. With regard to train-mile and car-mile-related causes, they stated,

> "The likelihood that a train will be involved in an accident is a function of both train-miles (TM) and car-miles (CM) operated. Car-mile related causes are those for which the likelihood of an accident is proportional to the number of car-miles operated. These include most equipment failures for which accident likelihood is directly proportional to the number of components (e.g. bearing failure) and also include most track component failures for which accident likelihood is proportional to the number of load cycles imposed on the track (e.g. broken rails or welds). Train-mile related causes are those for which accident likelihood is proportional to the number of train miles operated. These include most human error failures for which accident likelihood is independent of train length and depends only on exposure (e.g. grade crossing collisions)…The probability that an accident will occur is then a summation of the number of train-miles multiplied by the train-mile accident rate and the number of car-miles multiplied by the car-mile accident rate. Thus, it follows that longer trains have an increased likelihood of having an accident due to a larger number of car-miles of exposure."

Anderson & Barkan (2005a) grouped accidents by track class, which they used as a proxy for train speed. The frequency of accidents and the average number of cars derailed were examined for each group. They found that the likelihood of a train accident varies by track class. The effects of train length and the position of derailed cars were also examined. The authors concluded that the position of a car within a train's consist affects its probability of derailment. They also found that the number of cars derailed, or the severity of an accident, is dependent on train length. Therefore, the probability that a particular car will be derailed in a derailment is largely a function of train length, train speed, and position within the consist.

Anderson & Barkan (2004) examined railroad accident rates for U.S. freight railroads based on different FRA track classes. Mainline accident rates were calculated for each FRA track class and reported in terms of both accidents per train-mile and per car-mile. They found a difference of two orders of magnitude between accident rates for the lowest and highest track classes. They concluded that incorporating a track class term in the calculation of derailment probabilities and risk increases the accuracy and usefulness of the results.

Arthur D. Little, Inc (ADL) (1996) examined the risk involved with transportation of hazardous materials by rail. ADL examined each FRA accident cause and grouped similar accident causes together. Each cause group was classified as either car-mile-related or train-mile-related based on industry expert opinion. The number of accidents for each cause group and track class was determined and accident rates per car-mile and per train-mile were calculated for each FRA track class.

Saccomanno & El-Hage (1989 & 1991) completed studies evaluating how the placement of dangerous commodity cars within a train consist affect the probability of their derailment. The authors determined that the number of cars derailing is a function of the cause of derailment, train speed, and the residual train length. It was also found that the point of derailment (POD) was strongly affected by the cause of derailment and train length. The authors demonstrated that the derailment probability of each car in a train can be calculated based on train length. The authors concluded that effective marshalling strategies may reduce the number of derailments in which hazardous materials are released.

Transport Canada (2006) conducted an evaluation of risk associated with stationary dangerous goods (DG) railway cars. They calculated the probability of a derailment of trains moving on tracks adjacent to stationary DG cars. The expected number of freight train derailments per million freight train miles was calculated based only on factors for traffic density and track class. The work concluded,

> "Derailments due to certain types of causes were found to be more influenced by these factors [traffic density and track class]. The relationship of higher derailment rates on low-density lines was anticipated for track-related causes. The track quality of low density lines often reflects a lower capital investment and there are less stringent tolerances in the maintenance standards for these lines, as they would often be uneconomical to keep operating otherwise."

Barkan et al. (2003) examined railroad derailment factors in the context of hazardous materials transportation risk. This study determined that the speed of a train and the number of cars derailed was significantly related to hazardous materials release probability. Certain accident causes were found to be more likely than others to create accident conditions in which hazardous materials may be released. In particular, these causes were broken rails or welds, buckled track, improper train handling, and broken wheels. The authors determined that different derailment causes resulted in different accident severities and distributions of cars derailed.

Dennis (2002) performed an analysis on the decline of accident rates for railroads since deregulation in 1980. Dennis developed a model that evaluated changes in the rail industry from 1983 to 1994 to understand the effect of these changes on accident rates. He concluded that federal regulation, whether measured by defect rates, violation rates, or inspection rates, had a statistically insignificant effect on the rate of track accidents during the period. However, Dennis determined that the investment in railroad track by railways did have a statistically significant effect on the decline in track accidents.

2.2 Factors Related to Broken Rails

The literature reviewed in this section was used as a framework for the analyses presented on service failure prediction in Chapters 4 and 5. The logistic regression service failure prediction model presented in Dick et al. (2003) was examined and evaluated for recent service failure data. In my analysis I expanded on the logistic regression technique examined in their study. The model developed to predict broken rail locations by Sourget & Riollet (2006) led to the conclusion that inclusion of additional factors such as maintenance activities, could be used to create a more accurate service failure prediction model. Additional literature sources reviewed here focused on the mechanistic analysis of broken rails. Kim & Kim (2002), da Silva et al. (2003), Skyttebol et al. (2005), and Aglan & Gan (2001) all conducted studies examining fatigue crack growth in rail. Their analyses were used to understand the factors and some of the unexplained variance in the service failure prediction models that I developed.

2.2.1 Statistical Prediction of Broken Rails

Dick et al. (2003) developed a service failure prediction model based on factors related to broken rails. They determined that broken rails were the most frequent cause of severe train accidents on U.S. freight railroads. A prediction model was developed that evaluated track and traffic characteristics. The factors examined included rail age, rail weight, degree of curvature, speed, average tons per car, average dynamic tons per car, percent grade, annual gross tonnage, annual wheel passes, presence of insulated joints, and presence of mainline turnouts. The prediction model was developed using the step-wise logistic regression technique. The data that were used in this analysis were service failures for a two-year period from the BNSF Railway. The retrospective model they developed was found to be 87.4% accurate for predicting service failures. The model terms found to be significant for service failure prediction were rail age, degree of curvature, annual traffic, rail weight, annual number of wheel passes, average dynamic wheel load, presence of a turnout, tons per car, and track speed. They concluded that the model could be used to provide probabilistic estimates of the likelihood of service failure occurrence on the basis of engineering and operational input parameters.

Dick (2001) completed research focused on evaluating the factors affecting broken rail service failures and derailments. The first objective of the study was to examine the importance of broken rail derailments for freight railroads. He determined that broken rail derailments were far above the average in terms of frequency and consequence. The second objective was to determine possible predictive factors for service failures. Each possible predictive factor was evaluated by the means of a single variable statistical analysis. The final objective was to complete a multivariate analysis of predictor variables, as described in Dick et al (2003) and to show how the model can be used to reduce broken rail derailments.

Sourget & Riollet (2006) developed a statistical tool, called PROBARAIL, to assist railroads with decision making regarding the optimal trade-off between maintenance cost and the damage cost of broken rails. A predictive model using logistic regression, based on Dick et al. (2003), was developed to identify the probability of broken rail failure at specific locations and compare that to an acceptable threshold level for failure. The model also allowed for different weights to be associated with certain portions of the rail network according to the seriousness of the consequences inherent to a failure in that particular location. The logistic regression model that was developed takes into account the rail age, rail weight, rail profile, curve locations, track cant, gradient, track speed, traffic levels, and maintenance activities.

Sourget & Riollet (2006) developed two models for prediction of broken rails: logistic regression and decision trees. The models were developed by creating a "learning" sample of data and comparing that to a validation sample. The authors first experimented with developing the models for a 20-year period; however, they found that changes in railway operations resulted in an inaccurate model. The authors state, "The analyses carried out have showed that the quality of prediction is decreasing with the length of the history." Therefore, the models were re-created using only five years of data. It was determined that the use of a logistic regression equation for data from 1999 through 2003 led to an accuracy level of approximately 75%. The authors concluded that there was a positive correlation between the failure probability and the traffic tonnage, rail weight, and maximum allowable track speed.

Sourget & Riollet (2006) also developed a severity model of a broken rail failure to assist with maintenance decisions. Segments with high probability of rail failure were examined in the context of two impact classes. The first class was the impact on the railroad's guaranteed level of safety. The second class is the financial impact for the railroad, such as maintenance costs, train delay costs, and derailment costs. Based on a number of factors, including sleeper characteristics, track circuit type, bridges, speed, and the closest maintenance center, the model determines the level of severity for the broken rail. The user of this model has access to both the failure prediction model information as well as the failure severity model information. Users can also examine an occurrence/severity matrix to compare different track sections on the network. The authors concluded that these tools allow for better decisions regarding maintenance work for preventing broken rails.

Shry & Ben-Akiva (1996) developed a model that established a relationship between fatigue failures of rail and factors affecting fatigue. The research examined discrete usage periods for multiple types of rail defects. A Weibull distribution was used to include variables for the dynamic operation of the rail line and changing maintenance conditions. The authors developed both a survival function and a hazard function for the condition of the rail. They concluded that the defect rate for a specific rail only depends on current conditions and not historical data.

Zarembski & Palese (2005) completed a study regarding rail transportation risk due to broken rails. The authors stated, “The risk of broken rail derailments is directly related to the rate of rail defect development and the associated relationship between service defects and detected defects.” The authors used a statistical analysis to evaluate the relationship between service defects and derailments. They found a correlation between broken rail derailments and service defects, with approximately 1 derailment per 125 service defects for main line track under current axle loading. Additionally, they determined that if a greater number of detected defects were found then the risk from broken rail derailments would decrease. Therefore, the authors examined the effect of improved inspection techniques. The research found that the use of risk-based ultrasonic inspection scheduling techniques will reduce the risk of broken rail derailments, due to the increase in detected defects. The authors stated that the use of risk based scheduling reduced the rate of broken rails and service defects by 30% or more.

Palese & Zarembski (2001) described the risk-based ultrasonic inspection program currently implemented by the BNSF railway. They considered a risk-based approach to scheduling inspections based on three factors, defect initiation, defect growth, and detection reliability. The authors state, “Combining the knowledge that not every defect will be found during a given test with the understanding of how defects initiate and propagate allows for a better understanding of how often ultrasonic tests must be conducted to increase the chance of finding these rail defects.” Some of the risk factors developed for specific BNSF track segments were passenger-carrying-miles, dark territory, single-track territory, and BNSF-defined key routes. The authors determined that both the service failure rate and the service-failure-to-detected-defect ratio have decreased significantly with use of the risk-based inspection scheduling. The authors concluded that more defects were being found by the detector cars as opposed to being found as service failures, thereby reducing the risk associated with broken rails.

Zhao et al. (2007) studied the risk of derailment of railway vehicles due to rail defects and broken rails. The risk of a derailment was measured by the expected number of broken rails multiplied by the severity of the broken rail event. Four models were developed and combined to predict the number of broken rails over a particular segment. The four models that were developed were breaks due to: thermit weld defects, imperfect inspections, fatigue defects, and the impact of grinding on reducing defects. The combination of the models was used to calculate the risk of broken rails.

2.2.2 Mechanistic Analysis of Broken Rails

In Hay's (1982) textbook on railroad engineering he reviews the types of rail defects that lead to broken rail service failures and derailments. Defects are divided into four groups: longitudinal defects, transverse defects, base defects, and other defects. He stated that the most dangerous type of rail defects are transverse fissures that initiate inside the rail head due to minute shatter cracks and then expand across the rail head due to cyclic loading until the rail breaks, often with little or no prior indication of the weakened condition. He also discussed different defect detection techniques used, such as induction systems, residual magnetic systems, and ultrasonic inspection systems. He also examined surface defects that arise from contact and shearing stresses. These include head checks, spalling, flaking, and shelly defects.

Sperry Rail Service (1999) published a guide to assist railroads with rail defect management and identification. The manual is divided into sections explaining each type of possible rail defect. The categories of rail defects that the authors defined are transverse defects, longitudinal defects, web defects, base defects, damaged rail, nicked rail, surface defects, and miscellaneous defects. The guide stated,

> "The growth of a rail defect depends on a great many variables. The chemical composition of the rail and the amount of rail flexing are factors which must be considered. The type of rolling stock (freight, passenger, or motive power), its weight, and its condition of repair are important, as well as the frequency of these loads. The conditions of the roadbed and weather changes which result in track movement also affect growth. With so many variables contributing to development, it is impossible to predict accurately the growth of any defect."

Smith (2005) completed an overview of railway wheel and rail fatigue failures. Smith stated that the quality of steel manufacturing has improved over the last 30 years, thus eliminating many fatigue failures from internal defects in the rail head; instead a large proportion of rail failures are now occurring at weld locations. Smith also found that the life of a rail is principally determined by wear at the railhead. The wear can lead to shape change along the length of the rail, which in turn produces greater stresses in the wheel-rail contact. The author stated that rail grinding can be used to remove corrugations and to restore the accurate rail head profiles that are essential for controlling these stresses.

Zarembski et al. (2005) and Zarembski (2005) examined the effectiveness and use of rail grinding. Traditional rail grinding objectives and uses are explored, as well as applications of profile rail grinding. The authors examined how to monitor the effectiveness of rail grinding by developing grinding quality indices. These indices can be used to determine if rail grinding is improving the quality of the track from pre-grind to post-grind. Additionally, the authors determined that for high rail in curves, rail grinding was 76% effective for improving rail profile, while rail grinding on the low rail in curves was only 46% effective and 61% effective in tangent track sections. The authors also conducted an economic study of rail grinding and found that the savings due to extended rail life alone pays for the cost of rail grinding. In addition there are a number of other benefits from rail grinding, such as reduced fuel consumption, reduced track geometry degradation, reduced tie-fastener degradation, reduced damage to rolling stock, and reduced noise and vibration.

da Silva et al. (2003) performed tests on newly manufactured rail to determine fatigue crack growth rate. Four different European rail manufacturers' steels were each tested under stress with temperature and humidity maintained at a laboratory level. To determine fatigue crack growth, manual measurements of surface cracks were recorded at regular intervals. The tests showed that regression analysis could be used to model the crack growth in each of the specimens. Additionally, by examining the regression, three different states (stage I, stage II, and stage III) of crack growth can be identified in steel. The authors found that crack growth is difficult to predict, because in stage III, crack growth can accelerate, remain steady, or slow down. The authors concluded there was no significant difference in the crack growth rates in the samples from the four different manufactures.

Skyttebol et al. (2005) studied the effect of residual stresses on fatigue crack growth in rail welds. The authors used finite element analysis and fracture mechanics to calculate residual stresses in a flash-butt welded rail. The authors varied a number of parameters in this test, including axle load, crack location, crack size, and rail temperature. The authors concluded that, “the analyses show that typical crack sizes that can be found in a weld may grow to failure in a very short time if the residual stress fields interact with the axle load.” The authors also found that, fatigue is strongly dependent on ambient temperature, time to failure depends on axle load, and that surface cracks are more dangerous than an embedded crack in the rail.

Kim & Kim (2002) completed a study examining the fatigue behavior of rail steel under mixed loading levels, such as what is experienced by typical railway steel. To simulate the affect of mixed loadings in the laboratory, the fatigue crack growth behavior was evaluated using various comparative stress intensity ranges. The results of this analysis were compared to the testing completed under constant stress. Specifically, the authors examined the transition from shelling to a transverse crack under mixed mode loadings. Finite element modeling was used to analyze the effects from the wheel/rail interface. The authors determined that internal cracks first grew in the longitudinal plane and turned into a transverse crack. The authors also concluded that fatigue crack growth rate under mixed loading conditions was slower than that under constant load.

Aglan & Gan (2001) examined the fatigue crack growth behavior of head hardened premium rail steel under load. The authors used the modified crack layer theory to model fatigue crack growth behavior. They recorded the crack length and number of cycles of loading to determine the crack speed and the energy release rate of the steel. Three distinct stages of crack growth were observed, crack initiation, stable crack growth, and unstable crack growth. The results of this study showed that a microscopic examination of the crack reveals microcracks, inter-granular separation, and plastic deformation of the material which lead to a deceleration of crack growth in the second stage. The authors also found that cleavage facets initiated from the grain boundaries led to unstable growth in the third stage of crack growth.

Fletcher et al. (2004) completed an examination of rails in which large rolling contact fatigue cracks had developed. The study focused on the interaction between adjacent long cracks, at least 10mm in length, that are at the beginning of their bending-stress-driven propagation phase. The authors developed a model based on the boundary element technique for the growth of adjacent long cracks. The results of the analysis were shown in a series of plots of stress intensity factors around crack fronts for both single and multiple-crack situations. The conclusions reached in their study were that the previous models of single-crack growth are misleading when dealing with a rail containing multiple adjacent cracks.

Fischer et al. (2006) studied the growth and behavior of surface cracks on railway track under load. The research considered the situation of a shallow angle surface crack that may propagate either parallel or perpendicular to the surface of the rail. This situation was tested for various strain and stress states. The authors found that small surface cracks are common but generally experience slow crack growth. However, some surface cracks were found to grow up to a few millimeters in length, and then change their direction towards the rail surface. The authors concluded that this deviation occurs at a crack length where the stress intensity ranges reach a threshold where the tensile residual stress are sufficiently large to change the direction of the crack.

Zumpano & Meo (2006) and Bouteiller et al. (2006) were studying the possibility of new detection techniques for rail damage and broken rails. Currently, the majority of rail inspection is completed externally with geometric measurements and internally with ultrasonic inspection. Zumpano & Meo (2006) were studying a new technique using wave propagation phenomena to identify discrepancies, or damage, to the rail. Bouteiller el al. (2006) were investigating the use of a voltage application to project a high frequency wave through the rail structure to detect broken rails. Both new developments work under the same principle: an improper returning wave indicates damage on or within the rail.

2.3 Statistical and Artificial Neural Network Modeling Techniques

The literature reviewed in this section serves as an introduction to the statistical and artificial neural network models developed in Chapters 4 and 5. Hocking's (1976) work established the four main variable selection techniques I used to determine the factors related to service failures. The research conducted by Dougherty (1995) and Tu

(1996) serve as background information on how artificial neural networks (ANN) are formed as well as their advantages and disadvantages compared to logistic regression. Finally, techniques developed by Yim & Mitchell (2003) on hybrid artificial neural network / logistic regression (ANN/LR) models were used to develop my ANN/LR hybrid service failure prediction models.

2.3.1 Statistical Models

Hocking (1976) completed an analysis reviewing the different variable selection techniques in linear regression models. The research described various computational procedures for the inclusion or removal of variables from a logistic regression model. The first computational technique described is the method to include all possible predictors for regression and compute the best regression model based on those parameters. Some of the more advanced techniques included forward selection, backward elimination, and step-wise. The forward selection technique starts with zero variables in the model and adds one variable at a time until a specified threshold level is achieved. The backward selection technique performs the opposite procedure where all variables are initially included in the model and variables are removed until a given threshold is reached. The hybrid of these two models, step-wise selection, begins as forward selection, but after each variable is added, the backward selection technique is then performed. In conclusion the author stated, "neither forward, backward, nor step-wise selection will assure that the 'best' subset is revealed." The author recommended that all three techniques should be performed in the hope of seeing agreement between the developed models.

Lei & Jing-feng (2006) developed a logistic regression model for determining landslide susceptibility. The logistic regression method is used to analyze the relationship between the binary response variable of a landslide occurrence and the continuous or binary explanatory variables. The first logistic regression equation developed showed that elevation, proximity to a road, river, and residential area are main factors triggering landslide occurrences in the study area. The predicted accuracy of the landslide susceptibility map was shown to be approximately 80%. In order to improve the accuracy, the authors developed a second logistic regression equation which was used only on areas with high susceptibility to landslides and used only engineering and geological condition data. The second logistic regression model yielded a higher level of accuracy for these locations. The authors concluded that using the double logistic regression modeling technique improved the predictive ability of the model.

2.3.2 Artificial Neural Network Models

Dougherty (1995) completed a review of the use of artificial neural networks (ANN) for transportation applications. The author reviewed the different learning techniques that can be used to form neural networks. The most traditional method, supervised learning, is a network that is constructed based on the inputs, the computed output, and the actual output value. A global error function is created that computes the difference between the calculated output value and what actually occurred. The neural network adds neurons and adjusts the connection weights in order to minimize the global error function. Some other types of learning explored in the paper include, reinforcement learning, self-organizing networks, and combined networks. The author also explored some of the possible applications in transportation such as predicting driver behavior, pavement maintenance, vehicle classification, traffic pattern analysis, etc. The author concluded that many of the problems in transportation systems are highly non-linear and the use of ANNs for these applications may prove to be a useful tool.

Tu (1996) completed a study to evaluate the advantages and disadvantages of using ANNs versus logistic regression. This study specifically looked at the application of ANNs and logistic regression in the context of predicting medical outcomes. Tu described multiple advantages of ANNs over traditional statistical methods: the neural network models require less formal statistical training to develop, ANN models can implicitly detect complex nonlinear relationships between independent and dependent variables, ANNs have the ability to detect all possible interactions between predictor variables, and ANNs can be developed using multiple learning algorithms. Tu also considered the disadvantages of ANNs compared to statistical methods: ANNs are referred to as “black box” model due to their limited ability to explicitly identify variable relationships, ANN models may be more difficult to use in field applications, ANN modeling requires greater computational resources, ANN models are prone to over-fitting, and ANN model development is empirical and many methodological issues remain unsolved. Tu concluded that logistic regression remains the clear choice when the primary goal of model development is to look for possible causal relationships and when the modeler wishes to understand the effect of predictor variables on the outcome. Tu also concluded that a hybrid version of both ANNs and logistic regression may lead to the best possible prediction model.

Yim & Mitchell (2003) completed a study that compared ANNs, logistic regression models, discriminant analysis (DA), and hybrid models for the use of predicting corporate firm failure in Australia. The authors applied traditional statistical methods and neural network techniques, but also developed two hybrid models for prediction. The first hybrid model, "Logit-ANN", used logistic regression to preselect the significant variables for prediction. Only these significant variables were then used as input variables for the ANN. The second hybrid model, "PLogit-ANN", used the output value (probability of failure) from the logistic regression technique as a new input to the neural network. ANNs in this study were constructed using backpropagation. The results of this analysis showed that stand-alone ANNs outperformed the traditional statistical techniques. Yim & Mitchell also found that of all the models examined in the study, the PLogit-ANN hybrid model performed the best. They concluded that the use of ANNs and hybrid models are a valuable tool for event prediction.

Odom & Sharda (1990) developed a neural network model for the prediction of firm bankruptcy occurrence by examining past financial data. The authors compared the predictive power of ANNs to the predictions made by a multivariate discriminant analysis. The results for the training sample were that the neural networks classified 100% of the cases correctly and discriminant analysis classified 96% correctly. However, when the models were tested against "unseen", validation data, the neural network only classified 59% correctly and the discriminant analysis classified 81% correctly, indicating that the neural networks had over-fit the data in this study.

Fanning & Cogger (1994) and Towell & Shavlik (1994) conducted studies using advanced neural network modeling techniques. Fanning & Cogger (1994) examined the use of a generalized adaptive neural network algorithm (GANNA) processor in comparison to traditional backpropagation neural nets and logistic regression techniques.

The advantages of a GANNA are that they do not require the size of the network to be predefined, but instead layers are added by measuring the additional performance and generalization of the model. Fanning & Cogger concluded that both ANN and GANNA performance was similar to the traditional statistical methods. Towell & Shavlik (1994) completed a study developing knowledge-based ANNs (KBANN). KBANN is described by the authors as a hybrid learning system built by mapping specific "domain theories" into neural networks and then refining the neural network using backpropagation. Towell & Shavlik concluded that KBANN networks generalized better than other traditional ANN models.

2.4 Railroad Economic Research

The literature reviewed in this section is used as background information for my economic impact study of broken rails presented in Chapter 6. Cannon et al. (2003) determined the different types of costs associated with broken rails. Their breakdown of broken rail costs is used as the framework for my economic impact analysis. The study by Zhao et al. (2006) was used to understand how the economic life of rail is affected by frequency of detected rail defects. Finally, Poole's (1962) study provided several tools to assist with calculating costs associated with railroad operations. His methods were used in the development of my train delay cost calculator.

Cannon et al. (2003) completed an overview of rail defects in railway track. One specific section of his analysis examined the cost of rail failures. The authors stated that the costs of rail failures include the following: track inspection, train delay, remedial treatments, pre-emptive treatments, derailments, and loss of business. Track inspection cost, referring to both visual and ultrasonic testing, highly depends on the frequency of inspection. Train delay cost includes a railroad's own equipment as well as required payouts as penalties to other railroads due to operating agreements. Remedial treatment costs include the cost for expenses such as rail replacement and weld repair. Pre-emptive treatment costs refer to rail grinding and track surfacing. The final two costs examined by the author were the cost of derailments from broken rails and the corresponding loss of business due to loss of customer confidence and support. The authors concluded that some costs associated with rail failures can be stated with certainty, while many others cannot.

Zhao et al. (2006) conducted research assessing the economic life of rail based on the frequency of rail defects. The authors examined the occurrence of rail defects and broken rails as they relate to maintenance activities. Specifically, the research evaluated the impact of rail grinding on the occurrence of rail defects. Additionally, the authors attempted to model the chance of imperfect track inspections. The authors determined that the life-cycle cost of rail is calculated by accounting for the cost of track inspections, rail repairs, train accidents, and rail renewal. The authors developed a model to determine the economically optimal amount of time rail should remain in service based on a given set of circumstances.

Poole (1962) examined and developed tools to assist with calculating the costs associated with railroad operations. For freight transportation costs, Poole examined the costs of maintenance of way and structures, maintenance equipment, depreciation and interest costs, cost of car repairs, and other transportation expenses. Poole also developed a method to determine the number of meets and passes and associated costs from diverting or rerouting traffic to another line. Other topics he explored in regards to track operating costs were the economics of faster train speeds, abandonment of alternative routes, and cost of transporting heavier cars with more capacity.

CHAPTER 3: THE RELATIONSHIP BETWEEN TRAIN LENGTH AND ACCIDENT CAUSES AND RATES

Train accident rates are a critical metric of railroad transportation safety and risk performance. Understanding factors affecting accident rates is also important for evaluating the effectiveness of various accident prevention measures. Accident rates have been the subject of a number of analyses but these have generally not considered the effect of train length on accident rate. Accident causes can be classified into two groups, those dependent on train length and correlated with the number of cars in the train, and those independent of train length, corresponding to the number of train-miles operated. These classifications have implications for the quantitative effect of various changes in railroad operating practices on railroad safety performance. Whether an accident cause is a function of car-miles or train miles affects how safety measures that might reduce the likelihood of that cause will affect overall train accident rate. Accident causes have been classified as car or train-mile-related based on expert opinion but these classifications have not been quantitatively tested. FRA accident data were used to develop a metric to objectively classify accident causes and 11 causes were reclassified from the previous classification. Based on the results of the study a sensitivity analysis was conducted to evaluate how changes in train length affect individual trains' accident likelihood and system-wide accident rate. The concept of car-mile versus train-mile accident causes leads to the premise that, although longer trains are expected to experience more accidents than shorter trains, operation of longer trains results in a lower system-level accident rate.

3.1 Introduction to Railroad Accident Causes and Rates

Train accident rates are a critical measure of rail transportation safety and risk and understanding them is necessary to evaluate the effect of accident prevention measures. Accident rates have been calculated by various organizations and railroads and aggregated statistics for all U.S. railroads are published annually by the Federal Railroad Administration (FRA) Office of Safety (FRA 2006, FRA 2007c). Rates have been used to assess various factors such as track class, geographic location, train speed, and track type (Treichel & Barkan 1993, Anderson & Barkan 2004, Transport Canada 2006). However, these analyses have generally not considered the effect of train length on train accident rate. Train length is thought to have an effect on accident rate because more cars in a train increase the likelihood that a car or track component in or under a train may fail. Based on this premise, it has been suggested that accident causes can be classified into two types, those that are a function of the number of train-miles operated and those that are a function of car-miles operated (CCPS 1995, ADL 1996). The initial classification into these two categories was developed by Arthur D. Little Inc. (ADL) based on the opinions of railroad industry experts. These classifications have implications for the quantitative effect of various changes in practice on railroad safety performance and have been used in subsequent studies of railroad safety (STB 2002, Zhao et al. 2007, Kawprasert & Barkan 2008). Therefore, statistical evaluation of the classifications will enhance their utility and may also clarify our understanding of them. Furthermore, this classification has implications for an accurate understanding of the relationship between train length and accident rate and consequent policy implications for railroad operating practices.

I undertook a study to investigate and evaluate the ADL accident cause classifications with the goal of understanding how operating practices, such as train length, affect the likelihood of a train accident. The objectives of this analysis were:

- Present the methodology for calculating train accident rates based on car-mile and train-mile accident causes,
- Develop a metric to quantitatively evaluate the classification of accident causes as car or train-mile-related,
- Use the metric to properly classify train accident causes,
- Develop new, up-to-date, train accident rates based on train length, and
- Conduct a sensitivity analysis on the model to illustrate how changes in train length may affect train accident rate.

3.2 Train Length Based Accident Rates

Train accidents include derailments, collisions, highway-rail grade crossing accidents, and other accident types. The likelihood that a train will be involved in an accident is a function of both car-miles and train-miles operated (ADL 1996, Anderson 2005, Anderson & Barkan 2005b). The number of car-miles operated for a particular train is affected by train length; longer trains accumulate more car-miles per train mile. However, not all accident causes are directly related to the length of the train; instead, some are related only to the operation of a train, irrespective of its length. Train accident causes that are a function of car and train-miles can be defined as follows:

"Car-mile-related causes are those for which the likelihood of an accident is proportional to the number of car-miles operated. These include most equipment failures for which accident likelihood is directly proportional to the number of components (e.g. bearing failure) and also include most track component failures for which accident likelihood is proportional to the number of load cycles imposed on the track (e.g. broken rails or welds)."

"Train-mile-related causes are those for which the accident likelihood is proportional to the number of train-miles operated. These include most human error failures for which accident likelihood is independent of train length and depends only on exposure (e.g. grade crossing collisions)." (Anderson & Barkan 2005a)

3.2.1 Car vs. Train-Mile Expectations

The car-mile cause and train-mile cause definitions lead to the premise that longer trains will experience more accidents than shorter trains. This is because longer trains are more susceptible to car-mile-related accidents than shorter trains due to the additional cars in the train. Conversely, a train should experience accidents due to train-mile-related causes regardless of train length. The length of a train, referred to here and throughout the paper, is the number of cars in the train and not the linear measure of a train's actual length.

This premise leads to two expectations that should be evident when examining accident data and can be used to evaluate different train accident causes. The first expectation is that the average length of a train involved in an accident should be greater for car-mile-related causes compared to train-mile-related causes because longer trains will experience a greater proportion of car-mile-related accidents. Conversely, train-mile-related accidents are independent of train length and should not be biased towards long or short trains.

The second expectation is that the percentage of accidents for car-mile-related accidents should be an asymptotically increasing function of train length, whereas the percentage of train-mile-related accidents should be an asymptotically decreasing function of train length. Longer trains should experience a higher percentage of accidents from car-mile-related causes due to their higher percentage of car-miles per train-mile operated. Conversely, shorter trains are expected to experience a greater percentage of accidents from train-mile-related causes.

3.2.2 Accident Rate Equation

Under the premise that train accidents can be separated into two distinct groups, car-mile-related causes and train-mile-related causes, a new accident rate model that takes into account the two types of classifications can be developed. The new accident rate equation must include a factor for train length to account for accidents that are dependent on the number of car-miles operated.

To develop the new model, all FRA train accident causes were examined (FRA 2007a, FRA 2007b). A previous study by ADL classified each accident cause as either car-mile or train-mile-related (ADL 1996). The purpose of this study was to quantify the risk of hazardous material transportation by examining all accident causes. The ADL study showed that accident types should be classified as either car-mile or train-mile-related to properly quantify the car-mile and train-mile related risk. By determining the number of accidents that have occurred due to each cause, two independent and mutually exclusive accident rates can be calculated, the car-mile-accident rate and the train-mile-accident rate. The expected number of accidents that a train will be involved in is the sum of the car-mile-accident rate multiplied by number of car-miles and the train-mile-accident rate multiplied by the number of train-miles. The expected number of train accidents that will occur can be calculated as follows:

$$A_{EXP} = R_C M_C + R_T M_T \tag{3.1}$$

where,

A_{EXP} = Accidents expected

R_C = Car-mile-accident rate (accidents per car mile)

M_C = Number of car miles

R_T = Train-mile-accident rate (accidents per train mile)

M_T = Number of train miles

Under this model it is expected that longer trains will experience more train accidents. As a train's length increases, train-miles operated remains constant, but the number of car-miles increases with each additional car. Therefore, the number of expected accidents for a single train increases due to the additional car-miles (Figure 3.1a).

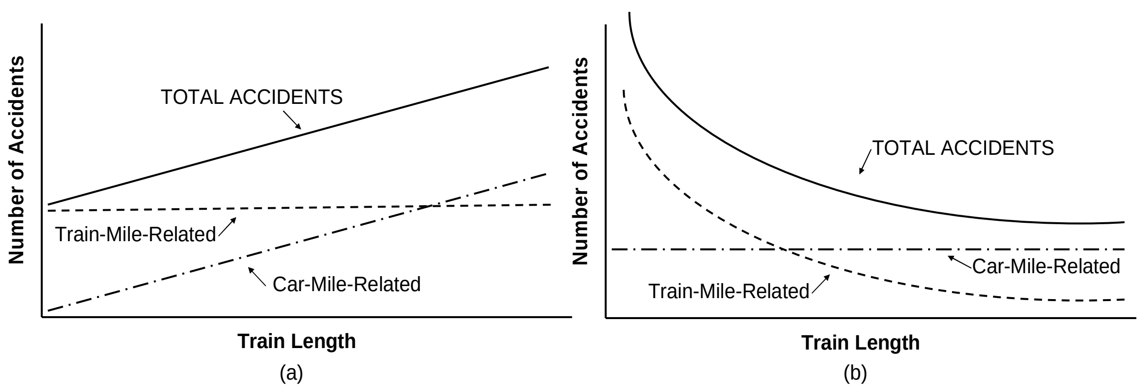

Figure 3.1 Expected Accidents from Car-Mile and Train-Mile-Related Causes as a Function of Train Length for a Single Train (a) and for a Fixed Amount of Traffic (b)

If one extends this model to any given number of cars that must be transported, it suggests the general result that operating longer trains should result in fewer accidents. As train length decreases, more trains are required to move the same number of cars thereby leading to more train-mile-related accidents. Under this simple scenario, accidents will be minimized by running the longest trains feasible given infrastructure and other constraints (Figure 3.1b).

It should be noted that there are limits to the validity of this result for very long train lengths (>150). This is because the expectations presented, as well as the data used in my analysis, apply to trains less than this length. In practice it is possible that accident rates for certain train-mile-related accidents may increase as train length becomes very long due to causes such as train handling, train braking, and other factors. The intention of this analysis is not to suggest that longer trains will necessarily improve safety; instead the purpose is to develop a better quantitative understanding of how changes that affect various accident causes, such as number of trains and train length, may affect overall accident rates.

3.3 Classification of Accident Causes

To accurately determine the car-mile and train-mile-accident rates, proper classification of each FRA accident cause is needed. The FRA accident cause classification system is very detailed and often includes several variations of one related group of causes. This is a useful attribute of the database, but is more detailed than necessary for the purpose of this analysis. Consequently, ADL combined similar accident causes into 51 unique groups, 34 of which they classified as car-mile-related (CM) and 17 as train-mile-related (TM) (Table 3.1) (ADL 1996). The FRA accident causes are separated into five main groups, mechanical, human, signal, track, and miscellaneous causes. ADL defined most track and mechanical failures as car-mile-related, while most human and signal errors were defined as train-mile-related. The various miscellaneous causes were assigned to either car-mile or train-mile-related.

I used FRA accident data, "Rail Equipment Accidents" from the FRA Office of Safety, to evaluate the ADL classification of accident causes for the period 1990 to 2005 (FRA 2007a). These data included all accidents occurring on either mainline or siding tracks for all classes of railroads. Accidents on yard and industry tracks were excluded because the average train length for these types of accidents is comparatively shorter due to yard operations. Mainline and siding accidents were combined because of similar accident causes and train length. Car and train-mile relationship predictions for each cause group were compared with the corresponding data from the FRA database. Train

Table 3.1 ADL/AAR Accident Cause Groups and Classification of FRA Accident Causes

Group	CM/TM	Cause Description	Group	CM/TM	Cause Description
01E	CM	Air Hose Defect (Car)	06H	TM	Radio Communications Error
02E	CM	Brake Rigging Defect (Car)	07H	TM	Switching Rules
03E	CM	Handbrake Defects (Car)	08H	TM	Mainline Rules
04E	CM	UDE (Car or Loco)	09H	TM	Train Handling (excl. Brakes)
05E	CM	Other Brake Defect (Car)	10H	TM	Train Speed
06E	CM	Centerplate/Carbody Defects (Car)	11H	TM	Use of Switches
07E	CM	Coupler Defects (Car)	12H	TM	Misc. Track and Structure Defects
08E	CM	Truck Structure Defects (Car)	01M	TM	Obstructions
09E	CM	Sidebearing, Suspension Defects (Car)	02M	TM	Grade Crossing Collisions
10E	CM	Bearing Failure (Car)	03M	CM	Lading Problems
11E	CM	Other Axle/Journal Defects (Car)	04M	CM	Track-Train Interaction
12E	CM	Broken Wheels (Car)	05M	TM	Other Miscellaneous
13E	CM	Other Wheel Defects (Car)	01S	TM	Signal Failures
14E	CM	TOFC/COFC Defects	01T	CM	Roadbed Defects
15E	CM	Loco Trucks/Bearings/Wheels	02T	TM	Non-Traffic, Weather Causes
16E	CM	Loco Electrical and Fires	03T	CM	Wide Gauge
17E	CM	All Other Locomotive Defects	04T	CM	Track Geometry (excl. Wide Gauge)
18E	CM	All Other Car Defects	05T	CM	Buckled Track
19E	CM	Stiff Truck (Car)	06T	CM	Rail Defects at Bolted Joint
20E	CM	Track/Train Interaction (Hunting) (Car)	07T	CM	Joint Bar Defects
21E	CM	Current Collection Equipment (Loco)	08T	CM	Broken Rails or Welds
01H	TM	Brake Operation (Main Line)	09T	CM	Other Rail and Joint Defects
02H	TM	Handbrake Operations	10T	CM	Turnout Defects-Switches
03H	TM	Brake Operations (Other)	11T	CM	Turnout Defects-Frogs
04H	TM	Employee Physical Condition	12T	CM	Misc. Track and Structure Defects
05H	TM	Failure to Obey/Display Signals			

lengths were grouped into 10-car bins and the percentage of all car-mile-related and train-mile-related accident causes was graphed versus train length (Figure 3.2).

A regression analysis was conducted in which a power function, of the form $y=ax^b$, was fitted to the data to evaluate how well they conformed to an asymptotically increasing or decreasing functional form. The critical term regarding the curve form of the power function, is the exponent, b. If $b > 0$, the data are more representative of an asymptotically increasing function (Figure 3.3a). If $b < 0$, the data are more representative of an asymptotically decreasing function (Figure 3.3b). As b approaches zero the power curve approaches linearity; whereas for larger absolute values of b, the power function curves more sharply. In the case of $b > 0$, the function will be convex for $b > 1$ or concave for $b < 1$. The residual error from the fitted power curves was also calculated as a function of train length (Figure 3.2).

The results were generally consistent with the car and train-mile premises developed. The average length of trains involved in an accident due to car-mile-related

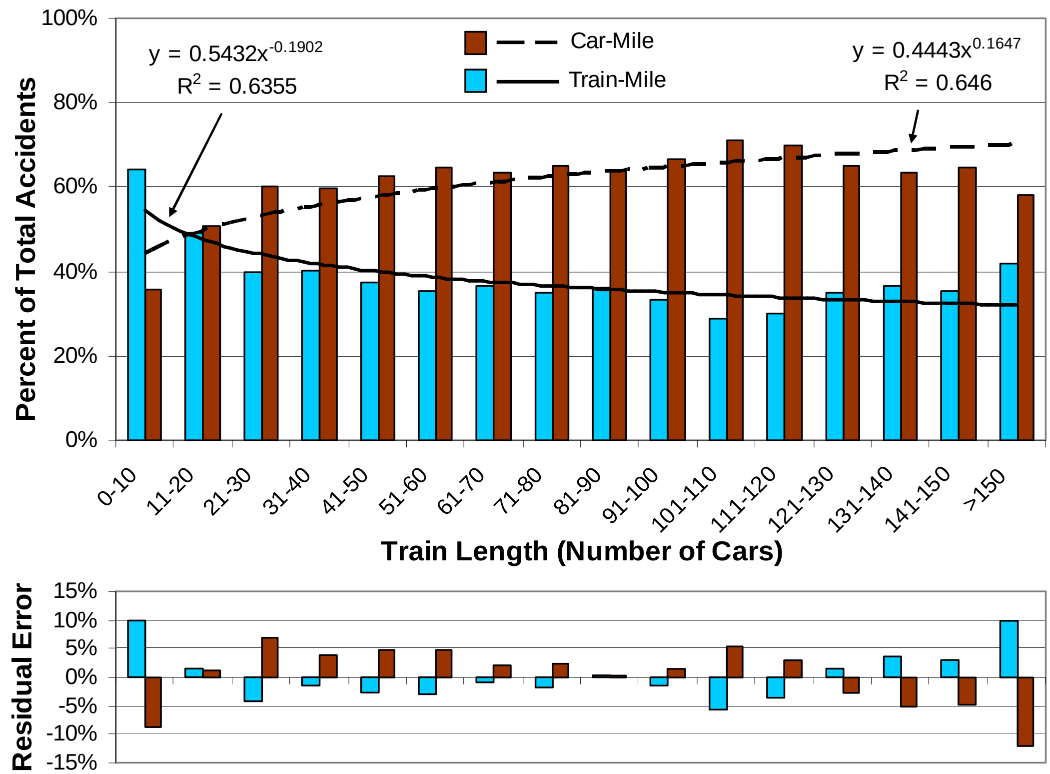

Figure 3.2 Percentage of Car and Train-Mile-Related Accidents versus Train Length using the ADL Accident Cause Classification

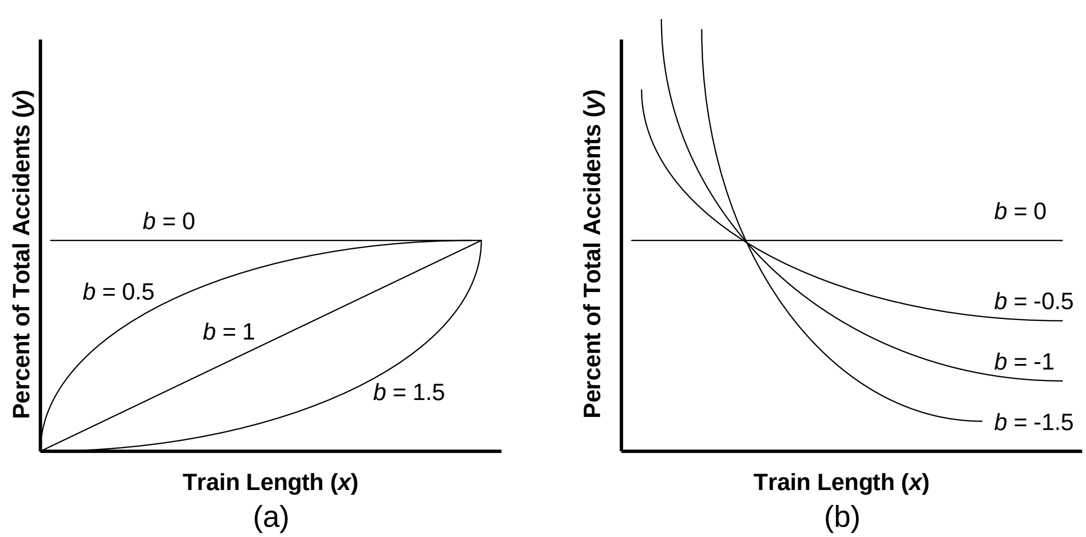

Figure 3.3 Characteristics of Exponential Term, *b*, of Power Function $y = ax^b$, where (a) Represents a Car-Mile-Related Cause and (b) Represents a Train-Mile-Related Cause

causes was 68.3 cars, whereas the average for train-mile-related causes was 52.5 cars. Also, the percentage of train-mile-related accidents declined asymptotically as a function of train length. However, although the R^2 values for the regression analysis were significant, it was evident that there were some discrepancies between the observed data and the predicted relationships, as indicated by the large residual error for the extreme train lengths (Figure 3.2).

These discrepancies suggested that the previous classification of accident causes should be re-evaluated to see if they could be improved based on newer data and analysis. Therefore, a more detailed analysis of individual accident causes was conducted. The relationships between number of accidents and percentage of accidents as a function of train length were graphed for each cause group. Although there were not enough data for accurate assessment of all the accident cause groups, many of them conformed well to the expectations for train-mile or car-mile-related causes, examples of which were grade crossing collisions and air hose defects, respectively (Figures 3.4a and 3.4b). However, examination of the data also suggested that some of the cause groups needed to be reclassified because the results were inconsistent with the car and train-mile expectations (Figures 3.4c and 3.4d).

A possible explanation exists for the cause group "train handling", which is caused by a locomotive engineer improperly handling the train and commonly attributed to excessive horsepower use. This had been previously defined as a train-mile-related cause because it is due to human error. However, accidents caused by the use of excessive horsepower are in fact more common in long trains than short trains and therefore resemble a car-mile-related cause. Conversely, the cause group "all other locomotive defects" had been classified as a car-mile cause because it is a mechanical failure. However, the number of locomotives, and therefore the likelihood of a locomotive defect, is not necessarily affected by an increased number of cars. Several discrepancies were also observed in other accident cause groups. Therefore a quantitative metric was developed to objectively classify each accident cause group as train-mile or car-mile-related.

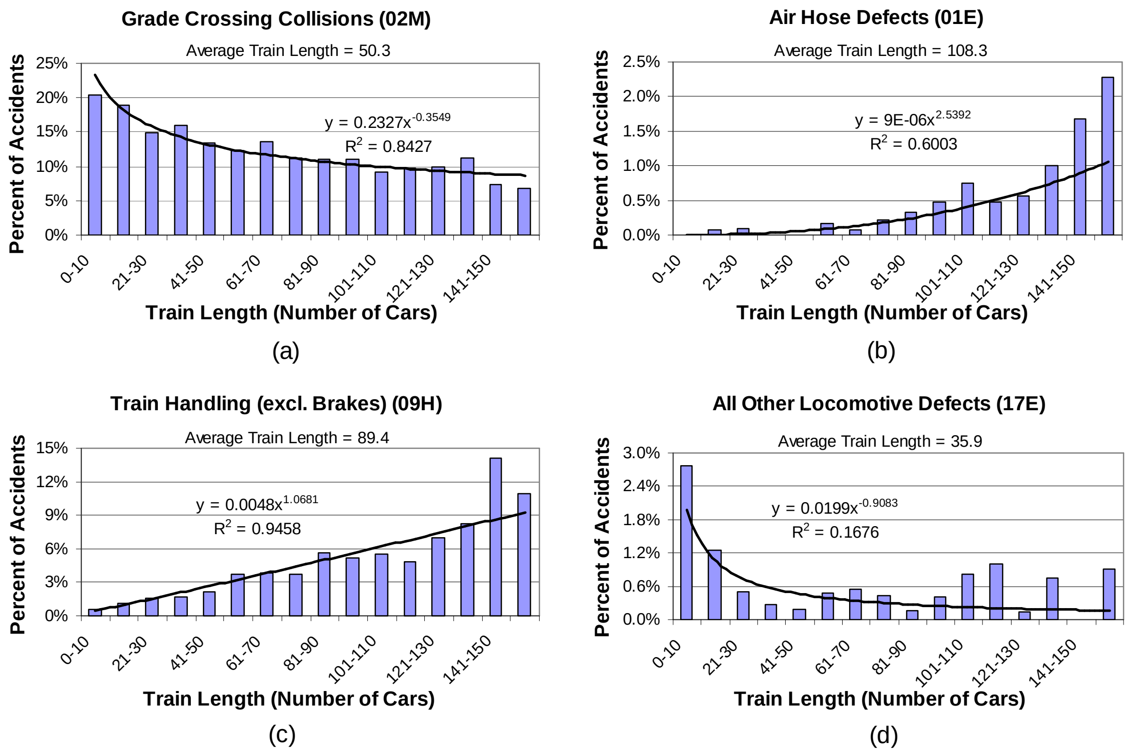

Figure 3.4 Percentage of Accidents versus Train Length for Four Example Cause Groups; Correctly Classified (a) and (b), and Incorrectly Classified (c) and (d)

3.3.1 Development of Classification Metric

I used the previously stated premise about car and train-mile-related causes to develop a quantitative metric to classify each of the ADL accident cause groups. Car-mile accidents should be more prevalent in longer trains and should be an asymptotically increasing function of the percentage of accidents as train length increases, and the reverse should be true for train-mile-related causes.

Two parameters were calculated for each accident cause to characterize them as either car-mile or train-mile-related. The first parameter is the average length of trains involved in an accident for each cause group. The second parameter is derived from the power function curve and its goodness of fit to the data for the percentage of accidents for each cause group as a function of train length. The exponent in the power function was used to assess the asymptotical increase or decrease in the data (Figure 3.3). The greater the difference between the calculated value of *b* and zero, the stronger the asymptotically increasing or decreasing function, and therefore the indication of either a car-mile or a train-mile-related cause. For example, cause group 2T, non-traffic/weather causes (b = -0.8666), showed a much stronger indication of a train-mile-related cause than 1M, obstructions (b = -0.3322).

In addition to characterizing the shape of the curves for each accident cause group, it was also important to quantify how well they fit the data. In some cases there were insufficient data to fit a curve and in others the data showed no trend. In order to assess the goodness of fit, the coefficient of determination, R^2, for each data set was calculated. R^2 values range from 0 to 1 and quantify the goodness of fit with higher values indicating a better fit. Therefore the accident causes with a high R^2 value are weighted more strongly in the metric than those with a low R^2 value. In summary, the accident metric, which was termed AM_i, incorporates three characteristics associated with each accident cause group, i: the average length of trains, l, the "shape" of the curve as a function of train length as indicated by the exponential term, b, and the goodness of fit of the data to the curve, as indicated by the R^2 value and is expressed as follows:

$$AM_i = \frac{l_i}{L} + (b_i R_i^2) \tag{3.2}$$

where,

AM_i = Accident cause metric for cause group i

l_i = Average train length for cause group i

L = Overall average length of trains involved in accidents in dataset = 61.79

b_i = Value of exponential term in power curve equation, $y=ax^b$, for cause group i

R_i^2 = Coefficient of Determination for a power curve fit to the data for cause group i

If the average length of trains in accidents due to cause i (l_i) is greater than L, AM_i is increased and vice versa. The greater the difference between l_i and L the more AM_i is affected. The second term of the metric is the power function exponent, b. If $b_i > 0$ for cause i it increases AM_i; and vice versa. Similarly, the greater the difference between b_i and 0 the greater the effect on AM_i. Finally, b is multiplied by R^2 to account for how well the function fits the data. If R^2 is close to 1, the second term will influence the metric more strongly. If the function is a poor fit (low R^2), b will have little effect on AM_i. Therefore, for R^2 values close to 1, AM_i will be calculated based on both average train length and b; whereas for low R^2 values AM_i will be calculated primarily based on average train length.

The metric, AM_i, was used to classify and rank the cause groups (Table 3.2). Not all cause groups included enough data to properly classify them as either car-mile or train-mile-related and these were excluded from the analysis. In particular, cause group 21E, current collection equipment, was excluded because only short passenger trains (<10 cars) were involved in this cause group with none of the accidents resulting in a derailment. The cause groups in Table 3.2 are ordered from most car-mile-related at the top, to most train-mile-related at the bottom. Cause groups with rankings in the middle are not represented strongly by either car-mile or train-mile classifications.

3.3.2 Reclassification of Accident Causes

AM_i is used to classify accident causes as either more consistent with characteristics of car-mile-related accidents or train-mile-related accidents. If $AM_i > 1$ the cause group is classified as a car-mile accident; conversely, if $AM_i < 1$ the cause group is classified as a train-mile-related accident (Table 3.2). If the classification based on the metric is different from the previous ADL classification this is indicated by a “YES” in the column heading “Change”. Using the metric reclassified 11 cause groups. Cause groups 1H, 9H, and 1S were changed from train-mile to car-mile causes. Groups 16E, 17E, 18E, 19E, 1T, 3T, 4T, and 12T were changed from car-mile to train-mile causes. Cause groups 3E, 4E, 14E, 21E, 4H, and 11T were not evaluated using the metric due to the small number of accidents for each group. The highest ranked car-mile-related accident cause is 1E, air hose defect, with a score of 3.277; whereas the highest ranked train-mile related-accident cause is 02H, handbrake operations, with a score of -0.0275.

As discussed above, there were instances where the accuracy of the initial classification based on the characteristics of the car and train-mile premise could be improved. Using the calculated values for AM_i I reexamined the overall train-mile and car-mile-related causes for comparison to the ADL classification. After reclassifying the data, the values were now more clearly representative of car-mile and train-mile-related

Table 3.2 Classification, Score, and Rank of Accident Cause Groups Using Metric

CAR-MILE-CAUSES		*Trendline $y=ax^b$*			*Distribution*		*Metric*		
Cause	**Description**	***a***	***b***	***R^2***	**Cases**	**Avg. Length**	**Score**	**Rank**	**Change**
01E	Air Hose Defect (Car)	0.000	2.539	0.600	50	108.30	3.2770	1	--
12E	Broken Wheels (Car)	0.001	1.631	0.942	372	96.90	3.1054	2	--
10E	Bearing Failure (Car)	0.002	1.409	0.893	780	89.24	2.7025	3	--
11E	Other Axle/Journal Defects (Car)	0.001	1.218	0.863	156	95.81	2.6022	4	--
09H	Train Handling (excl. Brakes)	0.005	1.068	0.946	647	89.34	2.4561	5	YES
01H	Brake Operation (Main Line)	0.002	1.047	0.822	209	90.43	2.3238	6	YES
07E	Coupler Defects (Car)	0.002	0.998	0.859	274	89.39	2.3043	7	--
13E	Other Wheel Defects (Car)	0.003	0.924	0.886	324	88.38	2.2486	8	--
06E	Centerplate/Carbody Defects (Car)	0.003	0.838	0.896	281	85.99	2.1423	9	--
05T	Buckled Track	0.006	0.697	0.726	438	78.95	1.7842	10	--
08E	Truck Structure Defects (Car)	0.000	0.834	0.059	61	94.66	1.5807	11	--
09T	Other Rail and Joint Defects	0.003	0.498	0.667	153	75.65	1.5562	12	--
04M	Track-Train Interaction	0.008	0.616	0.536	483	74.36	1.5337	13	--
05E	Other Brake Defect (Car)	0.002	0.517	0.320	109	77.73	1.4233	14	--
08T	Broken Rails or Welds	0.046	0.391	0.369	1798	71.66	1.3040	15	--
02E	Brake Rigging Defect (Car)	0.001	0.384	0.014	73	79.15	1.2863	16	--
20E	Track/Train Interaction (Hunting) (Car)	0.002	0.369	0.233	80	73.79	1.2799	17	--
07T	Joint Bar Defects	0.004	-0.180	0.004	115	78.44	1.2688	18	--
09E	Sidebearing, Suspension Defects (Car)	0.006	0.355	0.149	267	71.65	1.2125	19	--
06T	Rail Defects at Bolted Joint	0.004	-0.018	0.000	110	72.82	1.1785	20	--
01S	Signal Failures	0.000	0.724	0.053	64	69.27	1.1592	21	YES
10T	Turnout Defects-Switches	0.026	0.034	0.009	528	65.37	1.0583	22	--
03M	Lading Problems	0.020	0.131	0.082	469	64.60	1.0563	23	--
15E	Loco Trucks/Bearings/Wheels	0.009	-0.415	0.038	127	64.59	1.0294	24	--

TRAIN-MILE-CAUSES		*Trendline $y=ax^b$*			*Distribution*		*Metric*		
Cause	**Description**	***a***	***b***	***R^2***	**Cases**	**Avg. Length**	**Score**	**Rank**	**Change**
10H	Train Speed	0.002	0.113	0.014	64	61.67	0.9996	21	--
19E	Stiff Truck (Car)	0.021	-0.601	0.067	212	62.58	0.9728	20	YES
04T	Track Geometry (excl. Wide Gauge)	0.040	-0.796	0.113	1064	63.69	0.9405	19	YES
03H	Brake Operations (Other)	0.005	-0.122	0.060	80	58.05	0.9321	18	--
01T	Roadbed Defects	0.040	-0.796	0.113	274	55.18	0.8028	17	YES
05H	Failure to Obey/Display Signals	0.040	-1.134	0.138	213	56.79	0.7621	16	--
11H	Use of Switches	0.098	-0.901	0.124	561	53.41	0.7526	15	--
02T	Non-Traffic, Weather Causes	0.027	-0.867	0.159	155	53.28	0.7242	14	--
05M	Other Miscellaneous	0.061	-0.255	0.294	814	48.16	0.7045	13	--
18E	All Other Car Defects	0.017	-0.353	0.223	254	45.41	0.6562	12	YES
12H	Misc. Track and Structure Defects	0.018	-0.308	0.347	248	45.14	0.6237	11	--
03T	Wide Gauge	0.101	-0.480	0.407	933	49.68	0.6090	10	YES
06H	Radio Communications Error	0.015	-1.196	0.214	67	52.39	0.5915	9	--
16E	Locomotive Electrical and Fires	0.018	-0.799	0.139	161	43.12	0.5867	8	YES
01M	Obstructions	0.057	-0.332	0.626	686	46.41	0.5430	7	--
02M	Grade Crossing Collisions	0.233	-0.355	0.843	2546	50.27	0.5145	6	--
17E	All Other Locomotive Defects	0.020	-0.908	0.168	169	38.56	0.4718	5	YES
07H	Switching Rules	0.053	-0.601	0.678	411	44.72	0.3165	4	--
08H	Mainline Rules	0.026	-0.473	0.475	349	31.64	0.2873	3	--
12T	Misc. Track and Structure Defects	0.148	-1.379	0.303	569	30.30	0.0730	2	YES
02H	Handbrake Operations	0.144	-1.475	0.349	442	30.13	-0.0275	1	--

NOT EVALUATED USING METRIC		*Trendline $y=ax^b$*			*Distribution*		*Metric*		
Cause	**Description**	***a***	***b***	***R^2***	**Cases**	**Avg. Length**	**Score**	**Rank**	**Change**
04H	Employee Physical Condition				27	59.56			
11T	Turnout Defects-Frogs				25	76.00			
03E	Handbrake Defects (Car)				25	32.80			
04E	UDE (Car or Loco)				39	103.72			
14E	TOFC/COFC Defects				19	54.26			
21E	Current Collection Equipment (Loco)				86	7.62			

causes (Figure 3.5). The average train lengths for car-mile-related causes increased from 68.3 to 79.0 cars while the average train length of train-mile-related causes decreased from 52.5 to 48.4 cars. Also, *b* increased to 0.6175 and $R^2 = 0.9147$ for car-mile-related causes; whereas, *b* decreased to -0.4063 and $R^2 = 0.9201$ for train-mile-related causes. Overall, the new classification is more consistent with expectations from the stated car-mile and train-mile premise.

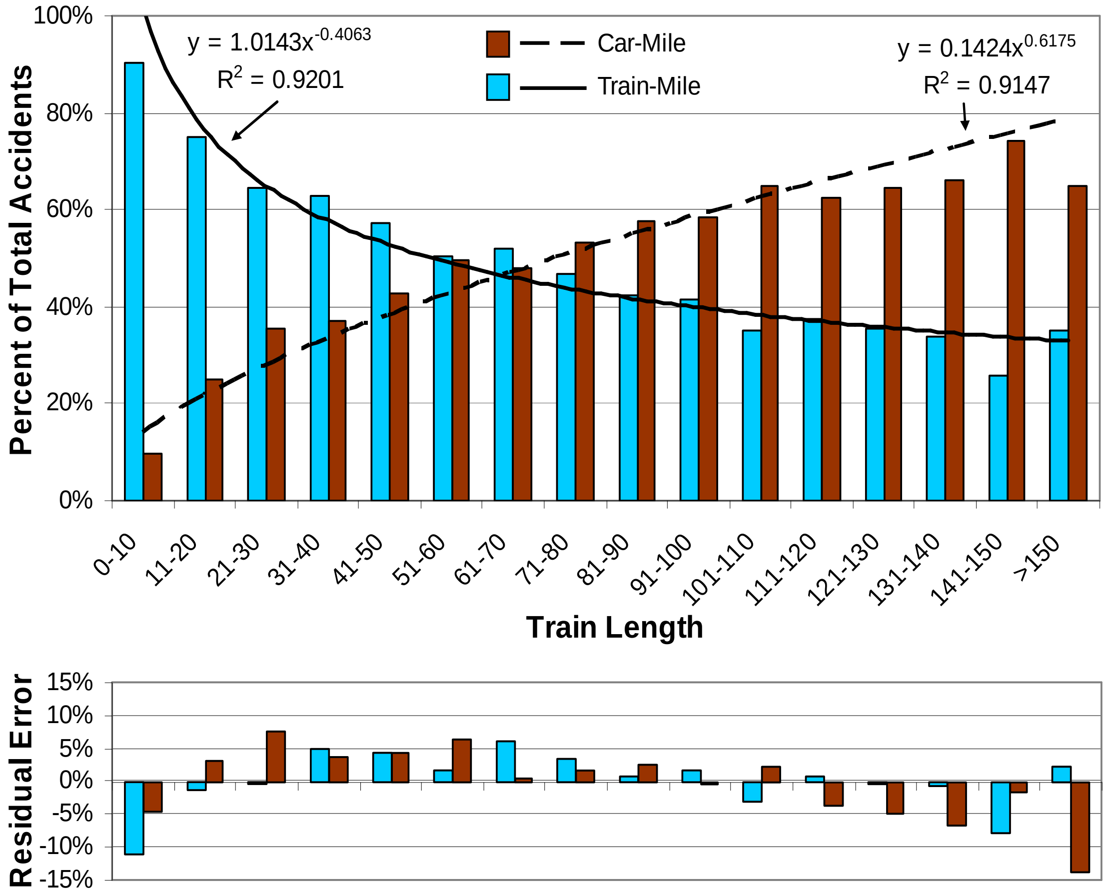

Figure 3.5 Percentage of Car and Train-Mile-Related Accidents versus Train Length using the New Accident Cause Classification

3.4 Calculation of Accident Rates

As stated earlier, train accident rates can be determined by summing the car-mile and train-mile-related rates. The two rates can be calculated using known accident data, the number of car and train-miles operated, and the new classification of accident causes.

Data on car-miles and train-miles operated are available from the AAR (AAR 2007). Car and train-miles are defined as the movement of a car or train the distance of one mile and is based on the distance run between terminals or stations. Accident information was downloaded from the FRA Office of Safety for the time period 1990-2005 (FRA 2007a). Data for all accident types for Class I railroads operating on mainline and siding tracks were used to ensure consistency with the AAR definition of car and train-miles for this portion of the analysis. The developed classification metric was used to classify each accident cause.

The car and train-mile related accident rates from 1990 to 2005 were calculated by dividing the number of accidents by the number of miles operated (Table 3.3). In 2005 the accident rate for car-mile-related causes was $1.05\text{x}10^{-8}$ or about 0.011 accidents per million car-miles and the train-mile-related accident rate was $8.62\text{x}10^{-7}$ or about 0.86 accidents per million train-miles. The expected number of train accidents, based on 2005 data, can be calculated as follows:

$$A_{EXP} = 1.05\text{x}10^{-8} M_C + 8.62\text{x}10^{-7} M_T \tag{3.3}$$

where,

A_{EXP} = Accidents expected
M_C = Number of car miles
M_T = Number of train miles

It is clear based on this equation that if the number of cars per train is increased, the consequent increase in car-miles operated leads to an increase in the accident rate for each train so affected. Similarly, an increase in the number of trains operated on a system will increase the number of train-miles operated, and thus increase the number of train-mile-caused accidents. To understand the effect of train length on accident likelihood, the accident rate equation can be expanded to include the term for train length:

$$A_{EXP} = 1.05\text{x}10^{-8}\, n\, d\, T_L + 8.62\text{x}10^{-7}\, n\, d = n\, d\, (1.05\text{x}10^{-8}\, T_L + 8.62\text{x}10^{-7}) \qquad (3.4)$$

where,

A_{EXP} = Accidents expected

n = Number of trains operated

d = Number of miles operated

T_L = Average cars per train (train length)

This equation is useful for understanding how changes in operating procedures, such as train length or number of trains operated, will affect the expected number of train accidents.

Table 3.3 Car and Train Mainline Accident Rates using the Reclassification of Accident Causes, Class I Freight Railroads, 1990-2005

Year	Car-Mile-Caused Accidents	Car-Miles Operated (Millions)	Car-Mile Accident Rate (per million car miles)	Train-Mile-Caused Accidents	Train-Miles Operated (Millions)	Train-Mile Accident Rate (per million train miles)
1990	510	26,159	0.0195	486	380	1.280
1991	479	25,628	0.0187	465	375	1.240
1992	360	26,128	0.0138	414	390	1.061
1993	370	26,883	0.0138	432	405	1.065
1994	315	28,485	0.0111	418	441	0.948
1995	362	30,383	0.0119	457	458	0.997
1996	379	31,715	0.0120	402	469	0.858
1997	343	31,660	0.0108	418	475	0.880
1998	378	32,657	0.0116	422	475	0.889
1999	367	33,851	0.0108	362	490	0.738
2000	420	34,590	0.0121	433	504	0.859
2001	400	34,243	0.0117	468	500	0.937
2002	374	34,680	0.0108	380	500	0.761
2003	392	35,555	0.0110	431	516	0.835
2004	424	37,071	0.0114	453	535	0.847
2005	395	37,712	0.0105	472	548	0.862
1990-2005	6,268	507,400	**0.0124**	6,913	7,460	**0.927**

3.5 Accident Rate Sensitivity Analysis

I conducted two simple sensitivity analyses to illustrate the effect of changes in train length on train accident rate. In the first I examined an operational choice of train length given a fixed number of car movements. The analysis parameters are intended to represent a typical high density, long distance, Class I railroad mainline with 25,000 car movements per week and a distance of 2,000 miles with train length and number of trains as the variables. The estimated number of accidents based on 2005 data is 1.05×10^{-8} accidents per car-mile plus 8.62×10^{-7} accidents per train mile as calculated by the previous reclassification of accident causes. I varied train length from 10 cars to 150 cars per train (Table 3.4).

Table 3.4 Sensitivity Analysis of the Effect of Train Length on Accident Rate

Average Train Length (T_L)	Number of Trains (n)	Probability of an Accident for each Individual Train	Total Expected Number of Accidents
10	2,500	0.00193	4.84
20	1,250	0.00214	2.68
30	833	0.00235	1.96
40	625	0.00256	1.60
50	500	0.00277	1.39
60	417	0.00298	1.24
70	357	0.00319	1.14
80	313	0.00340	1.06
90	278	0.00361	1.00
100	250	0.00382	0.96
110	227	0.00403	0.92
120	208	0.00424	0.88
130	192	0.00445	0.86
140	179	0.00466	0.83
150	167	0.00487	0.81

25,000 Carloads Shipped; 2,000 Miles; 150 Car Maximum Train Length

As train length increases, the likelihood that a train will be involved in an accident increases due to the increase in car-miles per train; however, because of the reduction in train miles, the net effect is a reduction in the total number of accidents. So all other things being equal, train accidents will be minimized when train length is maximized or the number of trains operated is minimized.

The second study examined how an increase in traffic levels may affect train accident rates. The analysis parameters are similar to those from the previous study of a 2,000 mile Class I railroad freight mainline with the same weekly traffic level of 25,000 car movements. The railroad is currently operating trains with an average length of 100 cars. The car movements are expected to increase by 10% to a new total of 27,500 movements. The operational choice in this study is either to continue operating the same number, but longer trains, or maintain the current train length and operate more trains. The traffic increase will lead to an increase in overall accidents; however, this effect can be minimized by increasing the length of trains instead of increasing the number of trains operated (Table 3.5). Again, this study suggests for this type of scenario that a railroad may be able to reduce the overall number of accidents by running fewer, longer trains as opposed to a higher number of shorter trains.

Table 3.5 Sensitivity Analysis of the Effect of Traffic Increase on Accident Rate

Number of Trains (n)	Average Train Length (T_L)	Probability of an Accident for each Individual Train	Total Expected Number of Accidents
250	100	0.00382	0.96
250	110	0.00403	1.01
275	100	0.00382	1.05

27,500 Carloads Shipped; 2,000 Miles

3.6 Conclusions

Accident rates are affected by both car-mile and train-mile-related accident causes. A consequence of this is that the length of trains affects accident likelihood. Previous research combined the FRA accident causes into 51 unique cause groups based on expert opinion. I developed a new quantitative metric to classify the causes as either car-mile or train-mile-related. Use of the new metric led to the reclassification of 11 of the cause groups and was found to be more representative of car and train-mile expectations. Therefore, using the new classification and recent accident data, updated mainline car-mile and train-mile-related accident rates were calculated for Class I freight railroads. These rates, as evaluated in a sensitivity analysis, showed that the decision to dispatch the same number of shipments in fewer longer trains versus more, shorter trains may affect the overall accident likelihood.

3.6.1 Future Work on Train Length Analysis

The analysis completed in this paper is based on a binary classification of accident causes as either train-mile or car-mile-related. However, many causes may not be purely train or car-mile-related, but instead depend on a combination of both. Future work may be possible to define a function for each cause group based on both car-miles and train-miles. Additional information, such as the distribution of trains operated by train length, would be useful in defining the linear or non-linear accident cause functions.

Future work may also be possible to evaluate and further refine the accident cause classification metric. For example, it may be possible to transform the current summation metric into a product metric using the same evaluation factors. The product metric may strengthen the analysis because it would multiply the classification terms and their effects on the metric instead of a simple summation. Further research into the classification metric might also reveal a better threshold for accident cause classification. An adjustment to the classification metric to include the average length of trains operated instead of average length of trains involved in accidents may remove potential bias.

Finally, it may also be possible to determine an optimal train length to minimize the number of cars derailed. Longer trains may be involved in fewer total accidents, but longer trains derail or damage more cars on average than shorter trains (Barkan et al. 2003).

CHAPTER 4: COMPARISON OF BROKEN RAIL PREDICTION MODELS

Broken rails are the second most frequent cause of mainline accidents in the U.S., exceeded only by grade-crossing collisions (Figure 4.1). More importantly, broken rails are the leading cause of major derailments and are the most frequent cause of hazardous materials releases. The average cost of damage to track and equipment due to mainline broken rail derailments on Class I railroads is $525,000 (FRA 2007a). Mainline broken rail derailments on U.S. Class I railroads have increased from 77 in 1997 to 91 in 2006, consequently steps to understand and prevent broken rail derailments are important.

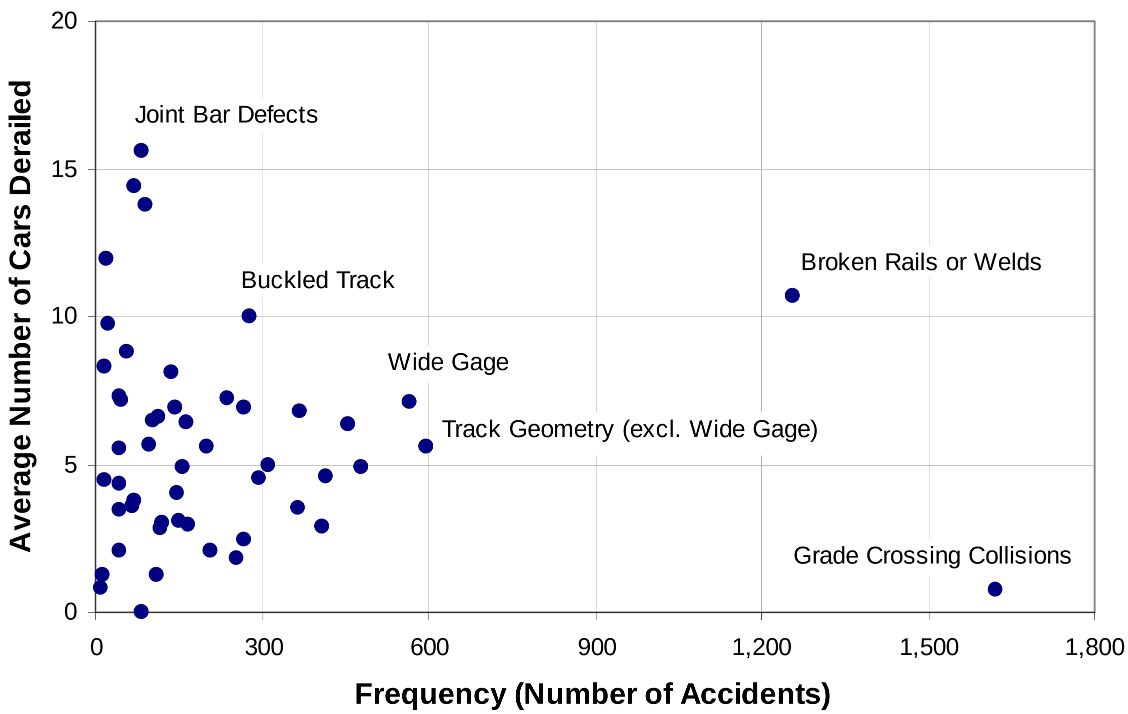

Figure 4.1 Railroad Accidents by Cause Severity vs. Frequency Graph, 1996 – 2005

One approach is to examine factors that potentially influence the occurrence of a broken rail in order improve the quantitative understanding of how they contribute to the likelihood of such an event. The objective of this study is to develop an accurate, predictive tool that will enable railroads to identify locations with high likelihood for broken rail occurrence so they can better prioritize preventive and mitigation measures.

Among the factors of interest are track and rail characteristics, traffic, maintenance activity, on-track testing results, and the presence of special infrastructure (i.e. bridges) that affect track modulus.

4.1 Introduction to Service Failure Prediction Model

Broken rail risk can be defined as the probability of a broken rail occurrence multiplied by its consequence. The consequence of a broken rail depends on the type of broken rail event that occurs. Broken rail events can be classified into two, broad categories: “service failures” and broken rail derailments. A service failure refers to the occurrence of a broken rail that does not result in a derailment. This generally occurs in situations where a broken rail is detected by the signal system or a track inspector. Under these circumstances, trains generally do not proceed onto the track section with the broken rail. The economic impact of service failures and broken rail derailments will be considered in more detail in Chapter 6.

Broken rails are caused by the growth of internal defects in the rail or surface defects on the head of the rail. Internal defects are generally caused by inherent flaws in the rail that form when the rail is manufactured. These internal defects are generally minute in size and nearly impossible to detect until they begin to grow. The growth of these rail defects is linked to a number of factors. Previous research as focused on both mechanistic analyses (Aglan & Gan 2001, Kim & Kim 2002, da Silva et al. 2003, Fletcher et al. 2004, Skyttebol et al. 2005, Smith 2005, Fischer et al. 2006) and statistical analyses (Shry & Ben-Akiva 1996, Dick 2001, Dick et al. 2003, Zarembski & Palese 2005, Sourget & Riollet 2006) in order to understand the factors that cause crack growth in rails and ultimately broken rails.

Among the previous studies was a multivariate statistical analysis of various factors affecting service failure occurrence (Dick 2001, Dick et al. 2003). This work used a discrete choice logistic regression model to determine the probability of a service failure occurrence for any given section of track. Discrete choice models have been used extensively for various classification applications. (Ben-Akiva & Lerman 1985, McCullagh & Nelder 1989).

The purpose of my analysis was to evaluate the previous work and expand on it using artificial neural networks (ANN). Different types of ANNs have been developed to predict events with promising results (Odom & Sharda 1990, Fanning & Cogger 1994). Both logistic regression analysis and ANNs have relative strengths and weaknesses, and for that reason, a hybrid model of both techniques was developed and evaluated. Previous work has shown that hybrid ANN/logistical regression models outperform purely statistical approaches in economics (Yim & Mitchell 2003), but this approach has not previously been applied to the prediction of broken rails. The objectives of this analysis were:

- Evaluate the previous statistical prediction model,
- Analyze the use of artificial neural networks as a classification tool,
- Develop a hybrid logistic regression/neural network model,
- Test all models for prediction capability of "unknown" cases, and
- Compare strengths and weaknesses for each model.

4.2 Statistical Prediction Model

The outcome of Dick et al.'s (2003) analysis was a model that used multiple parameters for specific locations in the railroad network and determined the probability of a service failure occurrence at any location. Data for service failure locations were provided by the BNSF Railway for the time period 1998 to 2000. The service failure prediction model was constructed using a logistic regression analysis.

4.2.1 Data Set Description

In Dick's (2001) study a "location" was defined as a track segment of length 0.01 miles, or approximately 53 feet. During the two-year period of May 1998 to May 2000, there were data for 1,903 service failures for the BNSF network. Of these, 1,861 segments contained sufficiently complete information to be included in the analysis. When modeling rare events, a commonly used approach is to sample all of the rare events and compare these with a similar sized sample of instances where the event did not occur (McCullagh & Nelder 1989). Therefore 1,900 locations were randomly selected from the BNSF network. Of these, 1,814 locations contained complete information and did not have a service failure in the two-year study period. Therefore, the data used in this analysis included 3,675 total locations from the BNSF network, approximately half of which had experienced a service failure and approximately half that had not.

In addition to data on the occurrence or non-occurrence of a service failure, the dataset that was developed included a large number of other parameters believed to have a possible effect on service failure occurrence. The parameters that were considered included:

- Rail age
- Rail weight
- Degree of curvature (if present)
- Track speed
- Average tons per freight car
- Average dynamic tons per freight car
- Percent grade (if present)
- Annual gross tonnage
- Annual wheel passes
- Presence of an insulated joint
- Presence of a turnout

Additionally, each case in the database contained another parameter for the occurrence of a service failure during the two-year time period. Each of the input parameters were entered as a numerical value, except for the presence of a service failure, insulated joint, or turnout, which were binary entries of either zero or one. The output, or dependent variable, for these analyses of this dataset is the occurrence of a service failure, while the remaining 11 variables are the input or independent variables.

4.2.2 Logistic Regression Model

The problem was defined as a discrete choice classification problem of either failure or non-failure. A location with a failure was defined as the occurrence of a service failure during the two-year study period. The input variables were the track, traffic, and infrastructure data available, as described previously. The service failure probability model was developed using the Statistical Analysis Software (SAS) and the

LOGISTIC procedure (SAS 2006). This procedure fits a discrete choice logistic regression model to the input data. The output of this model is an index value between zero and one corresponding to the probability of a service failure. A probability threshold value of 0.5 for classification of each track segment as either service failure or non-service failure was determined to be optimal.

The SAS LOGISTIC software has various possible regression analysis techniques. The step-wise regression technique was determined to be the optimal method for classification. Step-wise regression is a step-by-step method that selects the most important factor influencing the output value at each step until all factors are entered into the model. The procedure stops when the there are no additional factors remaining that will improve the model by at least a defined level of significance. At the beginning of each step the procedure uses a "goodness of fit" test to see how the inclusion of each factor influences the performance of the model. The factor that results in the greatest increase in fit will be the next factor added to the model. At the end of each step in the step-wise procedure, the model examines the factors already included and eliminates any that are no longer improving the model. The SAS system also has the benefit of monitoring and alerting the user of the presences of collinear variables as well any parameters that are functions of each other that affect the results of the step-wise regression.

Dick (2001) used the step-wise regression technique to produce a retrospective service failure model. The model was "retrospective" because it made predictions about past events in the database that was developed with approximately 50% failures. Therefore a "prospective model" was developed by adjusting a constant term to more appropriately represent the probability of a service failure. The prospective model was used to calculate the number of expected service failures per mile, and identify locations that had a high likelihood of experiencing a service failure.

The retrospective model was of interest in this analysis for the purpose of further evaluating its predictive ability. The prediction model that Dick developed using the logistic regression procedure is:

$$P_{\mathrm{SF2}} = \frac{e^u}{(1+e^u)} \tag{4.1}$$

$$\begin{aligned} U = Z + 0.059A + 0.025AC - 0.00008A^2C^2 + 5.101\tfrac{T}{S} \\ + 217.9\tfrac{W}{S} - 3861.6\tfrac{W^2}{S^2} + 0.897(2N-1) - 1.108\tfrac{P}{S} \end{aligned} \tag{4.2}$$

where,

P_{SF2} = probability that a service failure occurred during a two-year period

Z = -4.569, model specific constant

A = rail age (in years)

C = curvature of track (in degrees)

T = annual traffic (in million gross tons)

S = rail weight (in pounds per yard)

W = annual number of wheel passes (in millions)

P = dynamic wheel load (in tons)

N = presence of turnout (1 if present, 0 otherwise)

L = weight of car (in tons)

V = track speed (in miles per hour)

Dick’s model was limited to two-term interaction variables and second order exponential terms. More detailed interpretation of each model term can be found in Dick (2001) and Dick et al. (2003).

The fitted model was evaluated by testing the accuracy of the prediction for each case. This evaluation technique was conducted using the previous data set and did not incorporate any unseen cases of service failures. An optimal threshold value of probability was determined to be 0.5 for the greatest accuracy. Table 4.1 shows the results of the logistic regression model developed at the probability threshold of 0.5. The model accurately predicted 3,212 of the 3,675 cases (87.4%) and had almost twice as many false positives as false negatives (302 compared to 161).

Table 4.1 Classification Results of Logistic Regression Model

Model Type	Outcome of Model Classification	Cases	Percent of Total
Logistic Regression	Correct Prediction	3,212	87.4%
	False Positive	302	8.2%
	False Negative	161	4.4%

4.2.3 Evaluation of Logistic Regression Model

To evaluate if a prediction model is accurate, a model should be tested not only on its original dataset but also an unknown dataset. Dick's (2001) study did not evaluate the accuracy of classification based on validation data, or unseen cases. To determine if the previous model is robust for validation data, additional data near the time period of the study are required. However, since all service failure data for the time period were used in the construction of the model, an alternative approach was needed to test the robustness of the model. The approach used was to divide the original data into two groups. The first group was defined as the training dataset and the second group was defined as the testing dataset. Of the 3,675 available cases in the original database, 2,205 cases (60%) were randomly selected for the training sample and the remaining 1,470 cases (40%) were placed in the testing sample. This process was replicated three times to produce three random samples of both training data and testing data for analysis.

The same step-wise logistic regression analysis described above was repeated on each of the three samples of training data and three new predictive equations were developed. The equations were similar; each used the same parameters with only the coefficients changing slightly. As was previously done, each prediction equation was used with a threshold level of 0.5 to evaluate the accuracy of the training dataset. The accuracies of the training samples after completing the logistic regression on each sample were 87.03%, 87.21% and 87.89% (Figure 4.2).

The next step was to test each prediction equation on the respective sample's unseen testing dataset. The predictive equations were used to determine the probability of a service failure for each case in the testing dataset. Again, a probability threshold level of 0.5 was used to classify each case as either failure or non-failure. The predicted classification of either failure or non-failure was compared to the actual event that

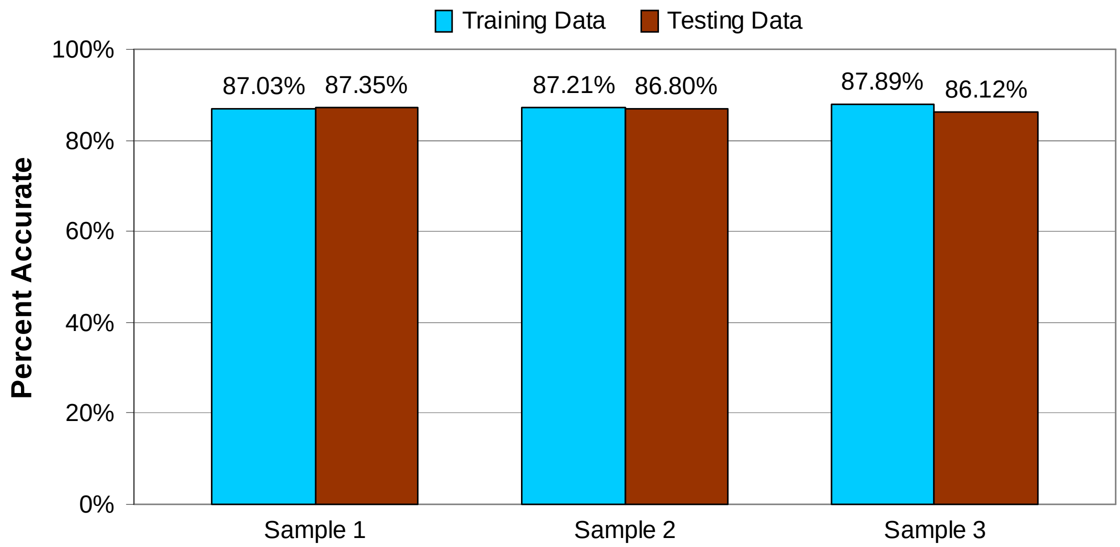

Figure 4.2 Accuracy of Training Data and Testing Data to Test the Robustness of the Logistic Regression Model

occurred in the testing dataset. The accuracy of each prediction model was found to be 87.35%, 86.80%, and 86.12% (Figure 4.2).

The prediction accuracy of each of the three models was similar for both the training dataset and the testing dataset (Figure 4.2). The difference in prediction accuracies for each sample ranged from 1.8% to -0.3%. The first sample actually had a higher level of accuracy for the testing dataset over the training dataset. Therefore, the logistic regression technique and the prediction models that were developed are robust for unseen data. These results affirmed and strengthened the validity of Dick's (2001) service failure prediction model.

The next objective of this analysis was to evaluate different prediction and classification techniques for service failures. The previous statistical methods were able to accurately classify service failures in approximately 87% of the cases in the original dataset. The remaining analysis was conducted to determine if different classification techniques could increase the prediction accuracy. One of the limitations of Dick's original logistic regression model was that it only considered linear mathematical relationships. Additionally, the statistical model was only evaluated for two term and second power interactions. In the next section artificial neural networks are explored as a possible prediction tool for service failures.

4.3 Artificial Neural Network Classification Model

The use of artificial neural networks (ANNs) is an alternative technique to statistical methods for purposes of classification and prediction. Many different types of neural network optimization procedures have been developed and more are currently being explored. Initial research into the use of neural networks as a classification technique were driven by economic and medical topics but have since been expanded to include many other fields, including engineering. Neural networks have both advantages and disadvantages when compared to statistical methods. For this reason, simple ANNs, as well as hybrid models, that combine both ANNs and logistic regression techniques, were developed and evaluated in this study.

4.3.1 Introduction to ANNs

Artificial neural networks have been used as an alternative to logistic regression in various applications. ANNs are a computational tool that can "learn" mathematical relationships between a series of input variables and their respective output values. ANNs are an interconnected group of "neurons" that have the ability to change their structure based on information that flows through the network. The development of ANNs was based on the idea of interconnected neural networks in biological systems such as animals. Artificial neural networks refer to those developed by computer systems.

The main parts of an ANN are the inputs, hidden layer, output, and node connections (Figure 4.3). The input layer of the ANN, shown on the far left, is comprised of the various input parameters into the classification problem. The inputs for this analysis were the same input parameters as the service failure model. The neurons in the hidden layer, shown in the middle of the diagram, are represented by mathematical equations and relationships that are determined by the algorithm. The arrows on the diagram represent a series of weighted connections between various nodes. The creation of node connections and their weights are determined by the ANN algorithms. Finally, the output node, shown on the far right, is connected to the hidden layer; in this case the only possible outputs are failure or non-failure.

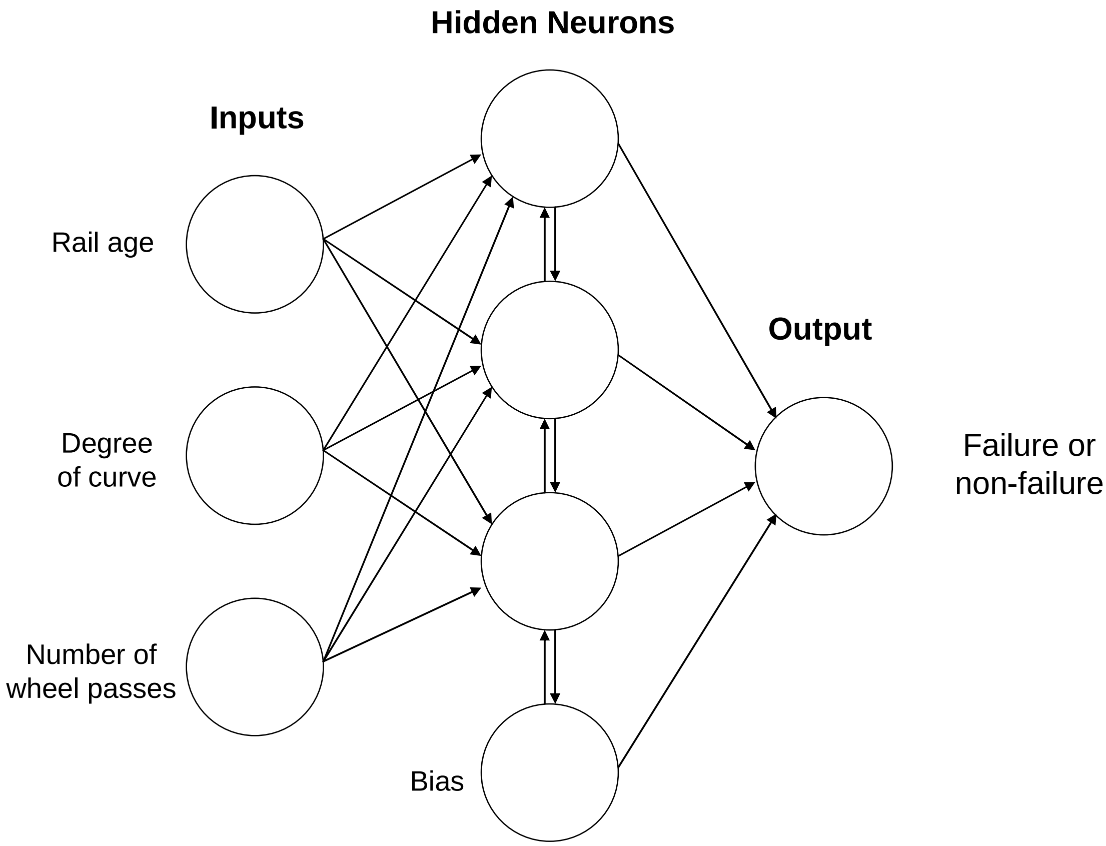

Figure 4.3 Diagram of an Artificial Neural Network

An artificial neural network is constructed by computer algorithms that add hidden neurons one by one until the optimal network is determined. An optimal network and the optimal number of hidden neurons represent a balance of model accuracy versus generalization. A network that generalizes well is one that is able to provide good results for data not used to train the neural net. In other words, the algorithm attempts to produce a neural network that is both accurate and robust for unseen cases. The software used for this analysis was "NeuroShell Classifier" developed by Ward Systems Group, Inc. (Ward Systems Group Inc., 2006).

Artificial neural networks have been used in various studies of event prediction, in particular classifying future events into either failure or non-failure. A previous study conducted used neural networks for predicting bankruptcy failure of firms based on limited financial data (Odom & Sharda 1990). The neural net developed by Odom & Sharda showed a higher level of prediction accuracy and robustness over previous statistical techniques. Another study used ANNs as well as a generalized adaptive neural network algorithm (GANNA) for the study of failing and non-failing firms (Fanning & Cogger 1994). Fanning & Cogger's study, using only three input variables, showed that ANNs, GANNA, and logistic regression models were comparatively similar in their prediction abilities.

4.3.2 Comparison of ANNs to Logistic Regression

With the development of ANNs as an alternative to logistic regression for prediction studies, research has been conducted to explore the differences in the two techniques. One study specifically examined the advantages and disadvantages of using artificial neural networks compared to logistic regression techniques to predict medical outcomes (Tu 1996). Tu noted that for predicting dichotomous outcomes, logistic regression has emerged as the statistical technique of choice. However, he also concluded that neural networks are not constrained by a predefined mathematical relationship between dependent and independent variables and therefore can have more accurate prediction models.

There are advantages and disadvantages to the use of artificial neural networks as a classification tool. As noted, the most important advantage of neural networks is the ability to detect complex non-linear relationships between input and output variables. The hidden layers and neurons as well as the node connections allow ANNs to have non-linear relationships between the input values, nodes, and output value. Another advantage is that ANNs can detect all possible interactions between input variables. The previous statistical model that was developed only evaluated two term interactions as well as only second power terms. The inherent design of a neural network evaluates and considers every possible variable interaction and power. Finally, ANNs have the advantage that they can be developed and evaluated using different learning techniques and different objective functions. This allows the creator of the neural network the ability to try different techniques to determine the optimal classification model.

The use of ANNs also has some disadvantages compared to statistical techniques such as logistic regression. One disadvantage is that computation time is longer for ANNs. This may be an important factor for large problems. Also, ANNs do not give a value for the probability of an outcome; instead they only classify an outcome as a failure or non-failure. If there is utility in having a quantitative sense of the likelihood of an outcome, i.e. a probability, then the logistic regression technique is better. Another disadvantage of neural networks is that they are "black box" models, meaning that the logic and quantitative functional relationships within a neural network are not easily reproduced. This limits the ability to explain what the model is doing and why. It means that a user cannot evaluate the possible relationships between input variables for an ANN model.

In Tu's (1996) research on ANNs, he concluded, "Neural networks may be particularly useful when the primary goal is outcome prediction and important interactions or complex nonlinearities exist in a dataset…Logistic regression remains the clear choice when the primary goal of model development is to look for possible causal relationships between independent and dependent variables, and a modeler wishes to understand the effect of predictor variables on the outcome…It is possible that some form of hybrid technique that incorporates the best features of both logistic regression and neural network modeling might lead to the best possible outcome prediction model."

An objective of my study was to use simple ANNs, and to evaluate the use of hybrid ANN/Logistic Regression (ANN/LR) models for the purposes of producing a more accurate service failure prediction model.

4.3.3 ANN Classification Model

A stand-alone artificial neural network was developed for classification of specific track locations as either failure or non-failure for a service failure. As described earlier, the dataset used contained 3,675 cases of track segments in which approximately 50% experienced a service failure in the two year time period. The ANN learning method used was backpropagation, which is a type of "supervised learning". Supervised learning is when both the input and output values for each case are entered into the network and the objective of the learning function is to reduce the mismatch between the neural network output and the actual output value. This is the most common form of computer learning for ANNs.

The input parameters used for development of the ANN were the same as those used for the logistic regression model. The ANN classified 3,223 of the 3,675 cases (87.7%) correctly (Table 4.2), as compared to the logistic regression model which classified 3,212 (87.4%) of the cases correctly. The optimal number of hidden neurons used to ensure robustness of the model was 77 for this network. The computation time for development of the ANN was eight seconds, as compared to two seconds for the logistic regression model. The ANN also produced more false positives than false negatives, similar to the logistic regression model (Table 4.2). The ANN was tested for robustness against unseen data in the same way as the logistic regression model. The results were that the model predicted correct classification of either failure or non-failure in 86.8% of the cases in the testing sample.

Table 4.2 Classification Results of Artificial Neural Network Model

Model Type	Outcome of Model Classification	Cases	Percent of Total
Artificial Neural Network	Correct Prediction	3,223	87.7%
	False Positive	283	7.7%
	False Negative	169	4.6%

4.3.4 Hybrid ANN / Logistic Regression Classification Model

As discussed, ANNs have some disadvantages compared to statistical modeling techniques. However, previous work has investigated the possible benefits of using hybrid ANNs and logistic regression techniques to overcome some of the disadvantages of simple ANNs (Spackman 1992). One study compared the use of ANNs, statistical models, and hybrid models for corporate firm failure (Yim & Mitchell 2003). The authors studied two different forms of hybrid networks for combining ANNs and logistical regression techniques. They concluded that the best statistical model was the logistic regression, but found that the results from the ANN were similar. However, they also found that the performance of the ANN was improved when hybrid models were considered.

The two common types of hybrid ANN/Logistic Regression (ANN/LR) models studied in previous research were the pre-selection of input variables and the addition of a probability input value. The first hybrid ANN/LR model form is for the pre-selection of variables (Logit-ANN). Logit-ANN uses logistic regression to determine the most influential input factors and develops an ANN based only on these factors for predicting failure or non-failure. The second ANN/LR model form is developed using the logistic regression to calculate the probability of failure for each case and adding that value as an additional input parameter into the ANN (PLogit-ANN). Again, the ANN is used to predict either failure or non-failure.

The two ANN/LR models produce advantages over the simple ANN classification technique. First, the hybrid models will decrease the number of cases used for learning the ANN, meaning that more cases can be devoted to optimizing the network instead of learning the network. Secondly, the hybrid models condense information for very large problems by pre-selection of variables. Finally, hybrid models may have a decreased amount of learning time required due to the pre-selected or condensed information. A decrease in learning time can be a significant factor for very large datasets with a large number of input variables, but this was not an important factor in this study.

The first hybrid model (Logit-ANN) developed for this analysis was the method of pre-selection of input variables. The logistic regression technique was used and the following parameters were identified for inclusion in the ANN: rail age, degree of curve, annual traffic loads, rail weight, annual number of wheel passes, average dynamic wheel load, and the presence of a turnout. The results from the ANN model were that 3,218 of the 3,675 cases (87.6%) were correctly classified (Table 4.3). The optimal number of hidden neurons was 74.

Table 4.3 Classification Results of Logistic Regression Artificial Neural Network Hybrid Models

Model Type	Outcome of Model Classification	Cases	Percent of Total
Logit-ANN Hybrid	Correct Prediction	3,218	87.6%
	False Positive	283	7.7%
	False Negative	174	4.7%
PLogit-ANN Hybrid	Correct Prediction	3,220	87.6%
	False Positive	290	7.9%
	False Negative	165	4.5%

The second hybrid model (PLogit-ANN) developed included probability of failure calculated using logistic regression in the ANN. Probability of failure for each case was entered as a new input variable into the ANN and the neural network was constructed again. The PLogit-ANN model classified 3,320 of the 3,675 cases (87.6%) correctly (Table 4.3). The optimal number of hidden neurons for this model was 78.

In both cases the hybrid ANNs took the same amount of time to construct the neural network. Also, both cases were tested for robustness against unseen data in a similar fashion as the previous models. The Logit-ANN hybrid model 86.4% accurate for the testing samples, while the PLogit-ANN hybrid model was 87.1% accurate for the testing data. Overall, these accuracy values were very similar to those produced by the previous models (Table 4.4).

Table 4.4 Summary of Classification Results for all Prediction Models

Model Type	Correct Prediction (%)	False Positives (%)	False Negatives (%)	Computation Time (sec)	Testing Sample Accuracy (%)
Logistic Regression	87.4%	8.2%	4.4%	2	86.8%
Artificial Neural Network	87.7%	7.7%	4.6%	8	86.8%
Logit - ANN Hybrid	87.6%	7.7%	4.7%	8	86.4%
Plogit - ANN Hybrid	87.6%	7.9%	4.5%	8	87.1%

4.4 Conclusions

In this study four different classification models were developed and analyzed for the purpose of predicting service failure. The first two models used stand-alone logistic regression and neural network techniques. The second two were constructed using hybrid combinations of the two techniques.

Four conclusions can be determined from this analysis. The first is that the simple ANN model and the hybrid ANN/LR models performed as well as the logistic regression model for classification purposes. This means that all models had a similar predictive ability for determining track segments that had a high likelihood of experiencing a service failure. The second conclusion is that all the models were robust against unseen data and were equivalently accurate for predicting service failures for unseen or unknown track segments. Additionally, the ANNs had a longer computation time compared to simple logistic regression analysis, but because of the limited data and input variables, computation time was short, eight seconds versus two seconds, respectively. However, computation time could become a factor for very large datasets.

Finally, all models had more false positives than false negatives. This means that all the models were more conservative when predicting service failures. This outcome was not intended, however, this is probably more desirable than the reverse for service failure prediction.

4.4.1 Next Steps in Service Failure Prediction Modeling

The next step is to apply the insights gained from the logistic regression, ANN, and hybrid ANN/LR models to a new, expanded dataset. This analysis will use recent data and include more input variables such as maintenance activities, rail testing results, and additional track, infrastructure, and traffic data. Also, the dataset will be expanded to evaluate a longer time period of study.

CHAPTER 5: STATISTICAL AND NEURAL NETWORK BROKEN RAIL PREDICTION MODELS

The purpose of this study was to consider the factors that influence the occurrence of broken rails and improve our understanding of the quantitative effect of these factors. In Chapter 4 I examined BNSF service failure data from a two-year period from May 1998 to May 2000 for a limited number of factors. The factors previously evaluated included only rail and traffic characteristics. In this chapter I expand on the previous work by analyzing more recent, comprehensive service failure data for a four-year period from 2003 through 2006. I have also included additional factors believed to affect service failures. The factors considered in this analysis included rail characteristics, infrastructure features, maintenance activity, operational information, and rail testing results. Two analytical approaches were used to understand the factors that affect service failures; statistical regression and artificial neural networks (ANN). Both approaches have relative strengths and weaknesses, and for that reason, hybrid models were also developed. The ultimate objective of this research was to develop a method that enables railroads to more effectively allocate resources to prevent the occurrence of broken rails.

5.1 Introduction to Broken Rail Prediction

Understanding the factors causing service failures and broken rail derailments is an important topic for U.S. freight railroads and is becoming more so because of the increase in their occurrence in recent years. This increase is due to several factors, but the combination of increased traffic and heavier axle loads are probably the most important. Broken rails are caused by the undetected growth of either internal or surface defects in the rail. The prediction of fracture growth within a rail once a defect is detected has been examined previously (Kim & Kim 2002, da Silva et al. 2003, Skyttebol et al. 2005). However, the majority of broken rails occur where a defect has not previously been detected. This is due to both the rapid growth of defects under load and various impediments to detection of certain types of defects, allowing them to grow to criticality without being detected. Previous research as focused on both mechanistic analyses (Aglan & Gan 2001, Kim & Kim 2002, da Silva et al. 2003, Fletcher et al. 2004, Skyttebol et al. 2005, Smith 2005, Fischer et al. 2006) and statistical analyses (Shry & Ben-Akiva 1996, Dick 2001, Dick et al. 2003, Zarembski & Palese 2005, Sourget & Riollet 2006) in order to understand the factors that cause crack growth in rails and ultimately broken rails.

The primary objective of this analysis was to develop a predictive tool that will enable railroads to identify locations with a high probability of broken rail occurrence based on service failure data and other possible influence factors. All of the available parameters that might affect service failure occurrence, for which data were available, were analyzed. These included rail characteristics, infrastructure data, maintenance activity, operational information, and rail testing results.

I developed several new predictive models using various techniques to attempt to predict broken rail locations. These included logistic regression (LR), artificial neural networks (ANN) and hybrid models that combined LR & ANN techniques. Previous work has shown that hybrid ANN/LR models outperform purely statistical approaches in other fields (Yim & Mitchell 2003), but this approach has not previously been applied to the prediction of broken rails. Each of these models was evaluated and a practical model was determined. The practical model was used to create a prospective service failure prediction model. The objectives of this analysis were as follows:

- Evaluate the previous prediction model developed using current data,
- Develop a new prediction model using various methods and techniques,
- Determine a practical prospective prediction model, and
- Examine the use of the model based on a hypothetical case study.

5.2 BNSF Service Failure Data

In order to develop a predictive model, it is desirable to initially consider as many factors as possible that might affect the occurrence of broken rails. Dick (2001) conducted an in-depth analysis of possible track and traffic factors based on data available to him at the time. In my study I considered these factors, as well as additional variables. From the standpoint of rail maintenance planning it is just as important to determine which factors are correlated with broken rails, as it is to determine which are not. Therefore the analysis included a wide-range of possible variables for which data were available. This included track and rail characteristics such as rail age, rail curvature, track speed, grade, and rail weight. Also, changes in track modulus due to the presence of infrastructure features such as bridges and turnouts have a potential effect on rail defect growth and were examined as well. Additionally, maintenance activities are included that can reduce the likelihood of a broken rail occurrence, such as rail grinding and tie replacement. Finally, track geometry and ultrasonic testing for rail defects are used by railroads to assess the condition of track and therefore the results of these tests are included as they may provide predictive information about broken rail occurrence.

The BNSF Railway provided relevant information regarding the location of service failures. A "BNSF service failure" is defined as any incident where track must be taken out of service for repair or replacement. For this study I define a "service failure" as an incident where a track was taken out of service due to a broken rail. Therefore, my definition of a service failure does not include incidents where trains are halted due to a rail found to be badly worn or damaged. Broken rail events in this analysis are then categorized as either service failures or broken rail derailments. Service failures may be detected in a number of ways including signal system, track inspector, or train crews. A broken rail derailment is defined as a broken rail that causes a train to derail.

A database was developed from approximately 23,000 miles of mainline track maintained by the BNSF Railway covering the four-year period, 2003 through 2006. The data available included specific locations for service failures occurring across the network. BNSF experienced 12,685 service failures during the four-year period (Table 5.1). Additionally, rail characteristics, infrastructure data, maintenance activity, operational information, and track testing results were linked to each of these service failures, for an overall total of 28 variables (Table 5.2).

Table 5.1 Summary of BNSF Network Data, 2003-2006

Event	Frequency
Annual Number of Geometric Defects	93,684
Annual Number of Ultrasonic Defects	45,294
Annual Number of Service Failures	3,171
Annual Number of Broken Rail Derailments	19
Track-miles Operated in 2006	37,003

Table 5.2 Variables Included in Service Failure Analysis

Rail weight	Degree of curvature
Rail type (bolted or welded)	Length of curve
Age of rail	Degree of superelevation
Maximum allowable track speed	Percent rise of grade
Annual number of trains	Length of grade
Annual number of tons	Recent tie replacement or tie work
Accumulated tons on rail	Presence of a bridge
Annual number of cars	Presence of a culvert
Average tons per car	Presence of a tunnel
Average dynamic tons per car	Presence of a diamond
Annual number of wheel passes	Presence of a turnout
Occurrence of a internal defect	Presence of a grade crossing
Occurrence of a geometric defect	Curve rail grinding activity
Severity of a geometric defect	Out-of-face rail grinding activity

BNSF’s network was divided into 0.01-mile-long segments (approximately 53 feet each) and the location of each service failure recorded. The initial dataset comprised the 12,685 0.01-mile track segments that experienced a service failure during the study period. For the case of modeling rare events it is common to sample all of the rare events and compare these with a similar sized sample of instances where the event did not occur (McCullagh & Nelder 1989). Therefore an additional 12,685 0.01-mile segments that did not experience a service failure during the four-year period were randomly selected from the same BNSF network of maintained track. Additionally, the non-failure locations were assigned a random date within the four-year time period for use in evaluating certain temporal variables that might be factors, such as the recent occurrence of an internal defect. Therefore, the dataset used in the remainder of this analysis included 25,370 total segment locations, each with a particular date, from the railroad's network.

5.3 Evaluation of Previous Service Failure Classification Model

The most relevant previous work on this topic was a study conducted by Dick (2001) for the purpose of predicting service failures based on relevant track and traffic data. The outcome of this study was a multivariate statistical model that was able to quantify the probability of a service failure at any particular location based on a number of track and traffic related variables. The model’s classification equation is as follows:

$$P_{\mathrm{SF2}} = \frac{e^u}{(1+e^u)} \tag{5.1}$$

$$\begin{aligned} U = Z &+ 0.059A + 0.025AC - 0.00008A^2C^2 + 5.101\tfrac{T}{S} \\ &+ 217.9\tfrac{W}{S} - 3861.6\tfrac{W^2}{S^2} + 0.897(2N-1) - 1.108\tfrac{P}{S} \end{aligned} \tag{5.2}$$

where,

P_{SF2} = probability that a service failure occurred during a two-year period

Z = -4.569, model specific constant

A = rail age (in years)

C = curvature of track (in degrees)

T = annual traffic (in million gross tons)

S = rail weight (in pounds per yard)

W = annual number of wheel passes (in millions)

P = dynamic wheel load (in tons)

N = presence of turnout (1 if present, 0 otherwise)

L = weight of car (in tons)

V = track speed (in miles per hour)

Dick (2001) determined that an optimal probability threshold for Equation 5.1 was 0.5 to classify each location as either failure or non-failure. The data used included a total of 1,903 service failures from a two-year period from May 1998 to May 2000. This model was found to classify locations correctly with 87.4% accuracy when using a dataset that was composed of half failures and half non-failures. This model was not tested against any “unseen” cases, or validation data, at the time it was developed.

The next step was to test Dick’s model against a two-year period of current service failure data. During the time period of 2005 through 2006, the BNSF experienced 6,613 service failures. These service failures, as well as 6,613 random non-failure locations, were entered into the above model in Equation 5.2. Again, using a probability threshold of 0.5 it was determined that the previous model classified 7,247 of the 13,226 cases correctly (54.8%). However, the new optimal probability threshold was found to be

0.1 with an accuracy of 57.2% (Table 5.3). It is evident that the previous model had less predictive power for more recent service failure occurrences.

Table 5.3 Results of Testing Previous Service Failure Model with Current Data

Probability Threshold	Correct Predictions	Accuracy	False Positives	False Negatives
0.1	**7,566**	**57.21%**	**29.79%**	**13.00%**
0.2	7,500	56.70%	24.01%	19.29%
0.3	7,402	56.00%	19.69%	24.35%
0.4	7,338	55.48%	16.83%	27.69%
0.5	7,247	54.80%	13.96%	31.24%
0.6	7,070	53.50%	11.13%	35.42%
0.7	6,931	52.40%	7.96%	39.63%
0.8	6,807	51.50%	5.06%	43.47%
0.9	6,663	50.40%	2.10%	47.52%

These results raised the question as to why the model had lower predictive power than it previously had. During the two-year interval that the model was based on the rate of service failures occurrences on the BNSF network was approximately 952 per year. During the more recent four-year period, BNSF experienced a total of 12,685 service failures, or approximately 3,171 per year. This more than three-fold increase in service failures was substantially higher than the increase that BNSF and other railroads had actually experienced during this interval. It suggested some unknown difference between the earlier and more recent databases.

A closer examination of the earlier service failure dataset compare to the more recent dataset revealed that there was a difference in the acquisition criteria for the two. The difference was due to a misinterpretation in the different reporting techniques for BNSF network locations that are on single track compared to those at multiple track locations. The service failure data provided by BNSF during the two-year study period of May 1998 to May 2000 may have included only service failures occurring in locations of multiple track lines. Whereas, the data provided for the more recent four-year study period included data for both single track as well as multiple track locations. During the recent four-year study period, BNSF experienced 4,689 service failures in areas of multiple track, or approximately 1,172 per year. This corresponds to a 23% increase in service failures in areas with multiple tracks compared to the previous study. This increase is much closer to the increase reported by railroads for this time period.

The distribution of BNSF service failures across their network can be graphically depicted using geographic information system (GIS) software and rail network data. Figure 5.1 shows the frequency of service failures per track mile on the BNSF network. A few line segments on the BNSF network are not included in this figure due to missing information. Overall, this figure shows that, as expected, many service failures occur on high density lines, such as the BNSF Transcon from Los Angeles to Chicago and the line extending east from the Powder River Basin in Wyoming and Nebraska. However, the figure also shows that there are service failures occurring elsewhere across the entire network, including many areas that are single track or unsignaled (dark) territory. Due to the inherent difference between the previous and more recent dataset, a new classification model was developed based only on the most current four-year study period.

5.4 Statistical Classification Model

The first new classification model that was developed to predict service failure locations used the same logistic regression techniques as Dick's (2001) previous work. However, unlike the previous work, more factors that might influence crack growth in rails were included to develop the model, such as infrastructure data, maintenance activities, and track testing results. The logistic regression technique was selected because it is a discrete choice model that calculates the probability of failure based on the input variables. These probabilities are used to classify each case as either failure or non-failure. A statistical regression equation was developed based on the significant input parameters to determine the probability of failure. To find the optimal classification model, the input parameters were evaluated with and without multiple term interactions allowed. Multiple term interaction allows for more complex relationships and dependencies that may not have been previously known. Additionally, a number of computational techniques for logistic regression are possible, and these were also examined and evaluated in this analysis.

5.4.1 Logistic Regression Methodology and Techniques

The logistic regression method (LR) uses a transformation that creates a prediction equation that calculates a value between 0 and 1. LR predicts the natural log of the odds for a case being in one category or the other. LR is widely used in multivariate regression problems in which the dependent variable is binary, or has only two levels, such as failure or non-failure (Cody & Smith 1997). Logistic regression analysis has been widely used in fields, such as medicine, engineering, business, and physiology (Carthey et al. 2003, Lei & Jing-feng 2006, Sagberg 2006, Mojsilovic et al. 2007).

Four possible computation techniques exist, and each technique was evaluated in this analysis, for the development of a logistic regression model. The simplest method is referred to as "full-model", or variable selection type "none" in SAS. The full-model method uses every available input variable to determine the best regression model. The other three methods are models which use variable selection techniques.

The next method examined in this analysis is selection type “forward”. Forward selection evaluates each input variable and initially adds the most significant variable to the model. Next, the forward selection method adds the variable that, when evaluated in conjunction with the first variable, produces the greatest improvement. This process continues until no additional variables meet a defined significance level for inclusion in the model. The entry and removal level used in this analysis was a 0.05 significance threshold.

The next logistic regression technique that was used was the “backward” selection. This selection method starts with all input variables included in the model. In the first step, the model determines the least significant effect that does not meet the defined significance level and removes it from the model. This process continues until no other variables included in the model meet the defined level of removal.

The final logistic regression selection technique used was “step-wise” selection. The step-wise selection method is similar to the forward selection method because the model begins with the most significant terms and continues adding terms step-by-step. However, unlike the forward selection method, the step-wise process evaluates the importance of all model terms after each step. If any term is determined to be insignificant, based on the defined significance level, than that term is removed and the process continues. The step-wise selection process ends when no further variables are added or removed from the model based on the defined entry and exit thresholds. Each of these four logistic regression models was used in the following analysis for both single and multiple term interaction.

Previous work has shown that use of the above described variable selection techniques may not lead to the optimal logistic regression model (Hocking 1976). Hocking stated that none of the variable selection techniques are superior to others, but instead all methods should be used to find the best model for the dataset. Hocking concluded that the developed models from each of the techniques should be compared for similarities and that these similarities may reveal a near-optimal model. For this study, each of the four variable selection techniques was evaluated.

5.4.2 Simple Logistic Regression Model

The first logistic regression model constructed was a multivariate analysis that did not allow for variable interaction. All four logistic regression selection techniques were implemented to determine the best model. SAS software was used (SAS 2006) for model construction. All four models had a similar classification accuracy of 66.3% for 25,370 cases being classified (Table 5.4). Additionally, the three selection models used the same 23 variables and developed the exact same logistic regression equation. The developed logistic regression equation was:

$$P_{\mathrm{SF2}} = \frac{e^{u}}{(1+e^{u})} \tag{5.3}$$

$$\begin{aligned} U = Z &- 0.0486S - 1.32R - 0.00362A - 0.0447V + 0.0000520F + 0.0313T \\ &- 0.000150H - 0.0542L + 0.0487P - 0.487W + 1.60I + 0.689G + 0.0501C \\ &+ 0.0637E - 1.42\text{x}10^{-6}J - 0.107M + 1.62B + 0.157K + 3.01D + 0.980N \\ &+ 0.361X + 0.778O + 0.589Q \end{aligned} \tag{5.4}$$

where,

P_{SF2} = probability that a service failure occurred during a four-year period

Z = 6.32, model specific constant

S = rail weight (in pounds per yard)

R = rail type (1 if welded, 0 if bolted)

A = rail age (in years)

V = track speed (in miles per hour)

F = annual number of trains (total, both directions)

T = annual traffic (in million gross tons)

H = accumulated tons on rail since rail was installed (in millions)

L = weight of car (in tons)

P = dynamic wheel load (in tons)

W = annual number of wheel passes (in millions)

I = presence of an ultrasonic defect in the last three years (1 if present, 0 otherwise)

G = presence of a geometric defect in the last three years (1 if present, 0 otherwise)

C = curvature of track (in degrees)

E = superelevation of track (in inches)

J = length of grade (in feet)

M = recent tie replacement or tie work in last three years (1 if present, 0 otherwise)

B = presence of a bridge within 200 feet of segment (1 if present, 0 otherwise)

K = presence of a culvert 200 feet of segment (1 if present, 0 otherwise)

D = presence of a diamond within 200 feet of segment (1 if present, 0 otherwise)

N = presence of a turnout within 200 feet of segment (1 if present, 0 otherwise)

X = presence of a grade crossing within 200 feet of segment (1 if present, 0 otherwise)

O = out-of-face rail grinding activity performed (1 if present, 0 otherwise)

Q = curve rail grinding activity performed (1 if present, 0 otherwise)

Table 5.4 Results of Simple Logistic Regression Service Failure Classification Models using Four Regression Techniques

Regression Technique	Number of Parameters in Model	Number of Cases Correctly Classified	Accuracy of Classification	False Positives	False Negatives
Full-Model	28	16,820	66.30%	12.80%	20.90%
Forward	23	16,822	66.31%	12.83%	20.86%
Backward	23	16,822	66.31%	12.83%	20.86%
Step-wise	23	16,822	66.31%	12.83%	20.86%

This new statistical model contains more variables that contribute to the likelihood of a service failure as compared to the previous classification model developed by Dick (2001). This is because the previous model only examined 11 of the possible prediction factors; whereas the new model evaluated 28 possible factors. The optimal probability threshold for classification was determined to be 0.05. Altering the threshold of probability will change the classifications of the model (Table 5.5).

The new step-wise model increased the accuracy of classification for the most recent service failure data by 11.5% over the previous model developed by Dick (2001). Therefore, the development of a new model, with the inclusion of additional possible factors leading to service failures, increased the model's predictive ability. In particular the first five terms, or most significant factors, entered into the new model were: presence of an ultrasonic defect, rail type, annual MGTs, average tons per car, and presence of a geometric defect. Of these five terms neither ultrasonic nor geometric defects had been

Table 5.5 Results of Step-wise Logistic Regression Model for Varying Levels of Probability Threshold Classification

Probability Threshold	Correct Predictions	Accuracy	False Positives	False Negatives
0.1	12,721	50.14%	49.83%	0.02%
0.2	13,298	52.42%	47.04%	0.54%
0.3	14,711	57.99%	38.99%	3.03%
0.4	16,545	65.21%	23.33%	11.45%
0.5	**16,822**	**66.31%**	**12.83%**	**20.86%**
0.6	16,485	64.98%	7.16%	27.86%
0.7	15,825	62.38%	3.83%	33.80%
0.8	14,787	58.29%	1.54%	40.17%
0.9	13,402	52.83%	0.26%	46.92%

included in the previous model. Additionally, the presence of infrastructure features, such as bridges, grade crossings, and diamonds, were not previously evaluated, but also have influence in the new statistical model.

The next step was to examine if the model's accuracy could be improved with changes to the logistic regression method. One type of change that was investigated was transforming some of the input parameters from continuous variables to discrete choice variables. Many of the inputs, such as the rail weight and the presence of infrastructure, were evaluated in the previous model as continuous variables, and may be better represented as discrete variables. For example, rail weight has 13 different entries: 89, 90, 100, 110, 112, 115, 119, 129, 131, 132, 136, 140, and 141 pounds per yard. In the previous model, rail weight was a continuous variable, meaning that the change in rail weight was directly proportional to the change in likelihood of a service failure. However, this practice is limiting because the change from 110 to 119-lb. rail may not be proportional to a change from 132 to 141 lb rail, even though each of these cases show a 9-lb. rail increase. Additionally, different rail weights have different cross sections, and therefore a change in rail weight may not be directly correlated with a change in likelihood of a service failure. Therefore, rail weight can be transformed to a discrete variable.

To transform a variable from continuous to discrete the addition of more input parameters is needed. The SAS software that was used to construct the logistic regression equations allows for the transformation of input parameters. For example, the rail weight variable has 13 unique entries; therefore, a total of 12 new input variables were created to represent the different entries (Figure 5.2). By definition, binary choice variables that have unique entries of only 1 and 0 are already discrete variables. Therefore, no transformation of binary variables, such as the presence of a bridge, was needed. Additionally, other variables that are continuous in nature, such as superelevation and annual tonnage, should not be transformed or the model may become over-fitted, or have a limited ability for prediction of future events.

Class Level Information													
Class	Value	Design Variables											
RAIL_WGT	89	1	0	0	0	0	0	0	0	0	0	0	0
	90	0	1	0	0	0	0	0	0	0	0	0	0
	100	0	0	1	0	0	0	0	0	0	0	0	0
	110	0	0	0	1	0	0	0	0	0	0	0	0
	112	0	0	0	0	1	0	0	0	0	0	0	0
	115	0	0	0	0	0	1	0	0	0	0	0	0
	119	0	0	0	0	0	0	1	0	0	0	0	0
	129	0	0	0	0	0	0	0	1	0	0	0	0
	131	0	0	0	0	0	0	0	0	1	0	0	0
	132	0	0	0	0	0	0	0	0	0	1	0	0
	136	0	0	0	0	0	0	0	0	0	0	1	0
	140	0	0	0	0	0	0	0	0	0	0	0	1
	141	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Figure 5.2 Transformation Chart for Rail Weight Variable from Continuous to Discrete Choice by Creation of New Inputs

Of the 28 variables, 13 were binary discrete and only one, rail weight, was a discrete variable with multiple unique values. All other variables were continuous and therefore could not be transformed to a discrete variable without arbitrary divisions, or bins, being created. The four logistic regression techniques were developed in a similar manner with the transformed dataset. This model performed better than the previous model with an accuracy of 68.5%, showing an increase in accuracy of 2.2% (Table 5.6). Additionally, the three variable selection techniques selected the same variables to include in the model and developed the same regression equation. None of the models in

Tables 5.4 and 5.6 allowed for variable interaction; therefore, the next step was to consider models with interactive terms.

Table 5.6 Results of Simple Logistic Regression Models using Variable Transformation and Four Regression Techniques

Regression Technique	Number of Parameters in Model	Number of Cases Correctly Classified	Accuracy of Classification	False Positives	False Negatives
Full-Model	40	17,351	68.39%	13.10%	18.51%
Forward	34	17,372	68.47%	13.01%	18.51%
Backward	34	17,372	68.47%	13.01%	18.51%
Step-wise	34	17,372	68.47%	13.01%	18.51%

5.4.3 Multiple Term Interaction Logistic Regression Model

The next logistic regression model considered the allowance of input variable interaction. Variable interaction is important to consider for prediction models due to the fact that some of the input variables may not be independent and may have a combined effect on the location of a service failure. For example, rail age and degree of curve may each have an independent effect on the likelihood of a service failure; however, the combined effect of rail age multiplied by the degree of curvature may produce an even stronger correlation with service failure locations. The software package included with SAS allows for a calculation of two and three-term interaction possibilities. However, only two-term interaction is possible for this analysis because of the large number of input variables included in the dataset. The initial model considered 28 independent variables; the two-term interaction model therefore considers 406 possible variables. Three-term or higher interaction was limited by the computational power available to process the large number of possible variable interactions.

Again, each of the four logistic regression techniques was used to develop service failure classification models with two-term interaction. The procedures followed in this part of the analysis were the same as in the simple logistic regression model. Initially, all input variables were considered to be continuous variables. The most accurate model, using two-term interaction, is the backward selection technique. This model classified 71.1% of the cases correctly, or an increase in accuracy of 2.6% over the best previous model (Table 5.7). In this case, the full-model included only 363 of the total possible 406 variables due to the fact that the interaction of some of the variables produces the same value as another input parameter. For example, the product of the variables rail age and annual gross tons is the same value as the parameter for accumulated tons on the rail. In each instance where the SAS software encounters a situation such as this, the variable is removed to prevent redundancy.

Table 5.7 Results of Two Term Interaction Logistic Regression Service Failure Classification Models using Four Regression Techniques

Regression Technique	Number of Parameters in Model	Number of Cases Correctly Classified	Accuracy of Classification	False Positives	False Negatives
Full-Model	363	17,921	70.64%	12.22%	17.14%
Forward	90	17,840	70.32%	12.23%	17.45%
Backward	145	18,025	71.05%	12.06%	16.89%
Step-wise	62	17,779	70.08%	12.20%	17.72%

As with the simple logistic regression models, the next step was to develop two-term interaction models in which applicable input parameters were transformed from continuous to discrete variables. Thus, the input variable for rail weight was split into 12 different input parameters to differentiate between the 13 unique rail weight values (Figure 5.2). Therefore, the total number of possible input parameters increased to 820, or more than double that of the model that did not account for discrete variables. The logistic regression equations were constructed using a similar procedure of evaluating all four possible regression techniques. The model that presented the highest level classification accuracy was the forward selection technique at 72.3% accuracy (Table 5.8), which was higher than any of the previous models. Overall, the results of the two-term interaction models produced higher levels of accuracy, but also included substantially more input parameters than the single-variable techniques.

Table 5.8 Results of Two Term Interaction Logistic Regression Models using Variable Transformation and Four Regression Techniques

Regression Technique	Number of Parameters in Model	Number of Cases Correctly Classified	Accuracy of Classification	False Positives	False Negatives
Full-Model	565	18,333	72.26%	12.47%	15.27%
Forward	336	18,340	72.29%	12.47%	15.24%
Backward	262	18,316	72.20%	12.63%	15.18%
Step-wise	44	17,123	67.49%	13.75%	18.75%

Two problems arise from models similar to those evaluated in this analysis with a large number of input parameters. First, such models are prone to over-fitting the data. Over-fitting will occur when a model creates relationships that are not actually factors that lead to failure but instead happen to fit the current set of data more accurately. To evaluate the robustness of the models produced in these analyses, the models must be tested against validation data, or "unseen" cases and this is examined in the following sections.

The second problem that arises from large models is the fact that they may be unreasonable to explain, define, and therefore implement in practice. For example, the most accurate model in this analysis was two-term interaction with forward selection; this technique had an accuracy of 72.3% with 336 input parameters. However, the step-wise simple logistic regression model shown in Equation 5.4 had an accuracy of 66.3% but only included 23 input parameters. In many cases it may be more practical to use a simpler classification model despite it being less powerful. The optimal solution will be a model that combines sufficiently high accuracy with a limited number of variables.

5.4.4 Development of a Practical Statistical Classification Model

The purpose of this analysis was to create a prediction model for service failures that was both understandable and useable with the intention of implementing it as a maintenance planning tool. As described in the previous sections, many of the logistic regression models included a large number of parameters and therefore are not conducive for understanding and use by a railroad. A practical model is thus needed that limits the number of input parameters but still has a sufficiently high level of accuracy. Such a model was developed by examining which input parameters of the dataset were most significant in predicting service failures.

To determine a simplified model, the logistic regression method was used with the "score" variable technique. This technique calculated the most important variables for accurate prediction of the logistic regression model. Due to limitations in computational power, the score program does not allow consideration of multiple-term interaction or the use of discrete variable transformation. For example, the score technique was used to determine the most powerful model if the input terms were limited to only the best five. In this case the five input parameters for the best model were: type of rail, annual gross tons, average tons per car, presence of an ultrasonic defect, and presence of a geometric defect. This five term model had an accuracy level of 64.0%. As compared to the previous simple step-wise regression model which had an accuracy level of 66.3%, but used 23 different parameters. A similar analysis was completed to calculate the best model for a varying number of parameters. The results of this analysis for inclusion of one to 23 parameters are shown in Table 5.9. This table shows what variables were removed and added at each iteration from the previous model. The base model of 23 parameters is the same as shown in Equation 5.4.

Table 5.9 shows that, in most cases, as the number of parameters decrease the classification accuracy of the model also decreases. Figure 5.3 is a graphical representation of the change of accuracy versus the number of model parameters. In some cases, the addition of another variable did not improve the model, most notably the six-variable model. This is because the model is being forced to create a model based on the best six variables, which has a lower level of accuracy then the five-variable model in this case. For example, if the step-wise regression technique were used, instead of forcing the model to be created with six terms, the sixth term would have been removed and the five-term model would be selected. In situations like this, the larger model is undesirable and would not be selected as an optimal model.

As shown in Table 5.9, generally the best model at each step is the same as the previous model with the least significant term removed. However, this is not true for the variable rail age. The rail age term was not present in models of sizes 20 through 10, but then added back in for model size 9. This means that rail age may not be significant when a number of other factors are included, but as the factors become limited, rail age is relatively more important. Another generality that can be drawn from Table 5.9 is that the presence of infrastructure features, except for bridges and grade crossings, are not significant factors. From this analysis, I determined that a “practical” model, that balanced both the number of input variables and the accuracy of classification, was the eight-parameter model. The calculated logistic regression equation presented a reasonably simple model that can be understood and used, but also has an accuracy level of 64.7%.

Table 5.9 Practical Service Failure Classification Models of Simple Step-wise Logistic Regression Technique by Number of Allowed Parameters

Parameters in Model	Parameters Removed	Parameters Added (if any)	Number of Cases Correctly Classification	Accuracy of Classification	False Positives	False Negatives
23	--	--	16,822	66.31%	12.83%	20.86%
22	Turnout	--	16,811	66.26%	12.86%	20.88%
21	Age of Rail	--	16,830	66.34%	12.95%	20.71%
20	Length of Grade	--	16,825	66.32%	12.92%	20.76%
19	Degree of Curve	--	16,856	66.44%	12.80%	20.76%
18	Average Tons per Car	--	16,854	66.43%	12.83%	20.74%
17	Culvert	--	16,816	66.28%	12.95%	20.77%
16	Tie Work Completed	--	16,835	66.36%	12.83%	20.81%
15	Superelevation	--	16,825	66.32%	12.95%	20.73%
14	Diamond	--	16,826	66.32%	12.98%	20.70%
13	Grade Crossing	--	16,727	65.93%	13.18%	20.88%
12	Speed, Annual Trains, Average Dynamic Tons, & Annual Wheel Passes	Average Tons per Car, Degree of Curve, & Grade Crossing	16,564	65.29%	13.52%	21.19%
11	Grade Crossing	--	16,571	65.32%	14.05%	20.63%
10	Out-of-Face Rail Grinding	--	16,560	65.27%	13.12%	21.60%
9	Accumulated MGTs & Degree of Curve	Age of Rail	16,429	64.76%	13.05%	22.20%
8	Curve Rail Grinding	--	16,407	64.67%	12.84%	22.49%
7	Age of Rail	--	16,303	64.26%	12.10%	23.64%
6	Bridge	--	16,064	63.32%	12.93%	23.75%
5	Rail Weight	--	16,235	63.99%	11.15%	24.86%
4	Geometric Defect	--	15,841	62.44%	10.89%	26.67%
3	Average Tons per Car	--	15,935	62.81%	10.74%	26.45%
2	Annual MGTs	--	15,213	59.96%	4.44%	35.60%
1	Rail Type	--	14,265	56.23%	1.65%	42.12%
0	Ultrasonic Defect	--	--	--	--	--

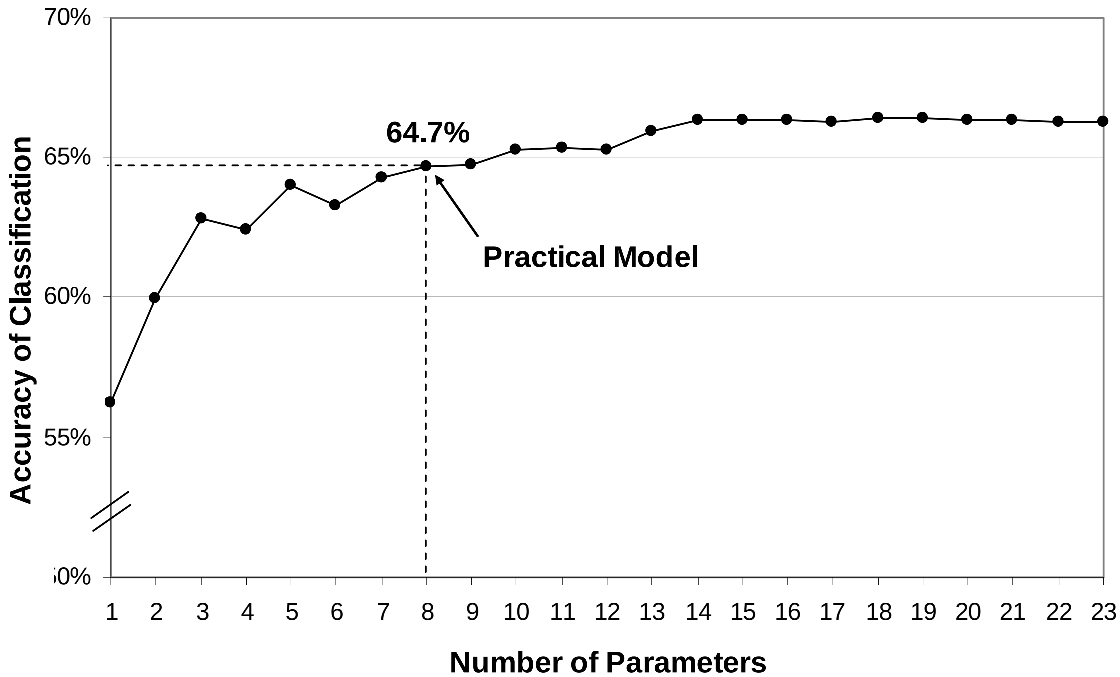

Figure 5.3 Accuracy of Classifications based on Number of Allowable Parameters for Simple Step-wise Logistic Regression Models

This is a decrease of only 1.6% from the basic step-wise regression model with 23 parameters, but is obviously much simpler to use and evaluate. The "practical" eight-term regression model was:

$$P_{\mathrm{SF2}} = \frac{e^u}{(1+e^u)} \tag{5.5}$$

$$\begin{aligned} U &= Z - 0.0454S - 1.35R - 0.0106A + 0.00899T \\ &\quad + 0.0232L + 1.61I + 0.823G + 1.63B \end{aligned} \tag{5.6}$$

where,

P_{SF2} = probability that a service failure occurred during a four-year period

Z = 4.94, model specific constant

S = rail weight (in pounds per yard)

R = rail type (1 if welded, 0 if bolted)

A = rail age (in years)

T = annual traffic (in million gross tons)

L = weight of car (in tons)

I = presence of an ultrasonic defect in the last three years (1 if present, 0 otherwise)

G = presence of a geometric defect in the last three years (1 if present, 0 otherwise)

B = presence of a bridge within 200 feet of segment (1 if present, 0 otherwise)

I next examined the applicability of this model as a tool for railway engineering professionals. The eight terms in this model can be ranked by how significant they are to the prediction power of the model (Table 5.10). The most important prediction term is the presence of an ultrasonic defect (I). The coefficient of this term in Equation 5.6 is positive, indicating that the presence of an ultrasonic defect in a 0.01 mile track segment increases the likelihood of a service failure occurring at a later point on the same track segment. This correlation between detected defects and the likelihood of a service failure is similar to the conclusions presented by Zarembski & Palese (2005). The model Zarembski & Palese developed showed that the risk of broken rail derailments is directly related to the rate of rail defect development.

Table 5.10 Ranking of Top Eight Input Parameters for Service Failure Regression Model

Ranking	Input Parameter
1	Ultrasonic Defect Present
2	Rail Type
3	Annual MGTs
4	Average Tons per Car
5	Geometric Defect Present
6	Rail Weight
7	Bridge Present
8	Age of Rail

The next most important parameter is the type of rail; if the rail segment is in bolted-rail territory the likelihood of a service failure is increased. The next two most significant parameters correspond to loading on the rail. As annual tonnage and average car weight increase, so does the probability of a service failure. The fifth most important factor is the occurrence of a geometric defect. Similar to ultrasonic defects, the presence of a geometric defect increases the likelihood that a service failure will occur in the same track segment. The next term in the model is rail weight (as a continuous variable). Rail weight is inversely related to service failure probability. The seventh most important term in the model is the presence of a bridge within 200 feet of the track segment location. The presence of this term in the model is consistent with conventional thinking that the change in track modulus often associated with the transition between track and bridges increases the dynamic load on the track and thereby increasing the likelihood of crack growth. The final term in the practical model is rail age. The regression equation indicated that rail age is inversely related to service failure occurrence. This relationship is counterintuitive but may be explained by the fact that rail life is relatively short in high density track but low density track may have very old rail. This simple eight-term model produced a classification accuracy of 64.7%.

5.4.5 Evaluation of Statistical Classification Models

The final step to determine the best statistical prediction model was to test the robustness of each of the regression equations developed. This was completed by testing each model against validation data, or "unseen" cases. Models that include a significant number of parameters are prone to over-fitting and are therefore poor prediction models for events that have not yet occurred. All of the service failures were separated into two groups. Of the 25,370 total cases, 15,222 cases (60%) were included in a training sample. The remaining 10,148 cases (40%) were retained in a testing sample, or validation group. The cases that were included in each dataset were selected at random over the four-year study period. The best model from each particular logistic regression technique was selected to test against the testing sample. Each of the five specific model techniques were used to create logistic regression equations based only on the training sample in the same procedure as previously described. The regression equations were then used to calculate the probability of failure for each case in the testing sample. Again, a probability threshold level of 0.5 was used to classify cases as either failure or non-failure. The predicted classification was then compared to the actual event that occurred in each case and overall model accuracy was determined.

The results from this analysis showed that the three logistic regression models that did not allow for variable interaction performed well against the testing sample and therefore are robust for service failure prediction (Table 5.11). Each of the three simple regression models had less than a 1% difference in accuracy between the training sample and the testing sample. The two models that allowed variable interaction over-fit the data and therefore had a large decrease in accuracy for the testing sample (Table 5.11). The most accurately constructed model, the two term interaction model with variable transformation, over-fit the training sample by almost 20%. Therefore, the variable interaction models are not useful prediction tools for calculating the probability of service failures.

Table 5.11 Evaluation of Statistical Service Failure Prediction Models by Testing Against Validation Data

Logistic Regression Model	Regression Technique	Accuracy of Initial Classification Model	Accuracy for Training Sample	Accuracy for Testing Sample	Change
Simple Logit	Step-wise	66.31%	66.73%	66.34%	-0.39%
Simple Logit w/ Transformation	Step-wise	68.47%	67.80%	67.69%	-0.11%
Two Term Interaction Logit	Backward	71.05%	71.17%	57.07%	-14.10%
Two Term Interaction Logit w/ Transformation	Forward	72.29%	86.74%	67.33%	-19.41%
Eight-term Logit Model	Step-wise	64.67%	65.05%	64.12%	-0.93%

5.5 Development of Artificial Neural Network Models

The next classification models developed used artificial neural networks (ANN). ANNs have been used in various studies of event prediction, in particular classifying future events into either failure or non-failure. The use of ANNs has been shown to be a more powerful alternative to logistic regression models in certain applications. In one previous study, a neural network model was developed for predicting bankruptcy failure of firms using limited financial data (Odom & Sharda 1990). The authors concluded that the neural network developed showed a higher level of prediction accuracy and robustness compared to previous statistical modeling techniques.

As described in the previous chapter, an ANN is a computational tool that has the ability to "learn" mathematical relationships between a series of input variables and their respective output value. The internal structure of an ANN is an interconnected group of neurons that have the ability to change its structure and connection weights based on information that flows through the network. With the development of ANNs as an alternative to logistic regression for prediction studies, research has been conducted to explore the differences in the two techniques (Tu 1996). As described by Tu, ANNs have two distinct advantages over traditional neural network models. One advantage is that ANNs have the ability to detect complex non-linear relationships between input and output variables that statistical analysis does not. The second advantage is that ANNs inherently detect all possible interactions between input variables as part of the learning process, unlike the statistical analysis which only evaluated two-term interactions.

5.5.1 Simple ANN Classification Model

An artificial neural network model was developed for classifying track segment locations as either failure or non-failure. The same service failure data, as well as non-failure locations, that were used to develop the logistic regression models were used again for construction of the neural networks. However, only 15,999 randomly selected cases could be analyzed due to limitations of the neural network software. The software used for construction of the artificial neural network was "NeuroShell Classifier" developed by Ward Systems Group, Inc (Ward Systems Group, Inc. 2006). Using the data from the four-year period, a neural network was developed using back-propagation. The ANN classified 67.7% of the cases correctly, an improvement of 1.4% over the previous simple step-wise logistic regression model (Table 5.12). The ANN model classified 12.7% false positives and 19.6% false negatives. The number of hidden neurons constructed for this neural network was 76. An ANN is constructed by algorithms that add hidden neurons one by one until the optimal network is determined. An optimal network and the optimal number of neurons represent a balance of model accuracy and robustness. A network that generalizes well is one that is not over-fit and therefore is able to provide good results for validation data. In other words, the software attempts to learn and produce a neural network that is both accurate and robust for unseen cases.

Table 5.12 Results of Developed Artificial Neural Network and Hybrid Service Failure Classification Models

ANN Model Type	Accuracy of Classification	False Positives	False Negatives	Number of Hidden Neurons
ANN	67.72%	12.64%	19.64%	76
Logit-ANN	67.52%	13.03%	19.45%	71
PLogit-ANN	67.93%	12.83%	19.24%	77

5.5.2 Hybrid Logit-ANN Classification Models

The final two classification models developed were ANN/Logistic Regression (ANN/LR) hybrid models. One of the disadvantages of ANNs when compared to logistic regression models is that ANNs frequently have difficulty analyzing systems that have a large number of parameters due to the amount of time required to learn the system, as well as possibly over-fitting the model during the initial learning phase. Hybrid ANN/LR models have been shown to improve classification performance when compared to traditional logistic regression techniques (Spackman 1992, Yim & Mitchell 2003).

Two types of ANN/LR models that have been developed in previous work (Yim & Mitchell 2003) were examined for this study. The first type of hybrid model is constructed using logistic regression to pre-select variables based on their significance in the prediction model (Logit-ANN). Only the factors included in the initial logistic regression model are then considered in the development of the ANN. The second type of hybrid model type is constructed using the logistic regression model to calculate the probability of failure and then adding that value as an additional input variable into the ANN (PLogit-ANN). The two hybrid models offer advantages over the logistic regression techniques. The hybrid models decrease initial learning cases for the ANN, meaning that more cases can be devoted to optimizing the network instead of learning the network. Additionally, the hybrid models condense information for very large problems thereby reducing learning time, which can be a significant factor for very large datasets.

In this study, both hybrid models were developed using the previously determined simple step-wise logistic regression model (Equation 5.4). This selection model was chosen because it produced both a high level of classification accuracy and was robust for unseen cases. The first hybrid, Logit-ANN, is constructed by first pre-selecting the input variables. The logistic regression model determined that 23 of the 28 input factors were significant for service failure prediction. Only these 23 factors were then used to construct the new ANN. The Logit-ANN hybrid model was 67.5% accurate (Table 5.12).

The second hybrid model, PLogit-ANN, was constructed using the logistic regression model to calculate the probability of failure for each case using Equations 5.3 and 5.4. This value was then added as an additional input variable for construction of the ANN. The PLogit-ANN hybrid model was found to be the most accurate model with a 67.9% correct classification rate (Table 5.12).

The PLogit-ANN hybrid model performed only modestly better than any of the other models, including the simple step-wise logistic regression technique. Additionally, Table 5.12 shows that the accuracy of the Logit-ANN hybrid model was slightly less than the stand-alone ANN; meaning the simple ANN model considered additional variables significant that the previous step-wise statistical model did not. Overall, the three artificial neural network models performed only about 1% to 2% more accurately than the simple step-wise logistic regression model. As with the statistical models, a better evaluation of the ANN models performance is to test them with unseen data.

5.5.3 Evaluation of Artificial Neural Network Classification Models

The various ANN models were evaluated using a similar procedure as the statistical models. The 15,999 cases used to construct the ANN were randomly divided into two groups, a training sample and a testing sample. 9,600 cases (~60%) were selected for the training set and 6,399 cases (~40%) were selected for the testing set. The results showed that all three models performed well against the testing sample as compared to the training sample (Table 5.13). Each model’s accuracy decreased by only 1% to 2% and therefore could be considered robust.

5.6 Final Prospective Service Failure Prediction Model

As stated previously, the objective of this analysis was to develop an accurate, understandable tool that railroads could easily implement to assist with maintenance

Table 5.13 Evaluation of ANN Service Failure Prediction Models by Testing Against Validation Data

ANN Model Type	Accuracy of Initial Classification Model	Accuracy for Training Sample	Accuracy for Testing Sample	Change
ANN	67.72%	67.75%	66.56%	-1.19%
Logit-ANN	67.52%	67.79%	66.28%	-1.51%
Plogit-ANN	67.93%	67.65%	66.43%	-1.22%

planning. Multiple models using both statistical and artificial neural network methods were developed to predict service failure locations. Each technique produced models with varying degrees of accuracy, robustness, and simplicity. Comparing ANNs to statistical methods revealed some shortcomings of neural networks. One disadvantage is that ANNs do not give a value for the probability of an outcome; the ANN only produces an output of either failure or non-failure. This is a significant disadvantage compared to the logistic regression method that can estimate the probability of failure for each case using the regression equation. Another disadvantage of neural networks is that they are "black box" models, meaning that the relationships between variables in a neural network cannot be easily understood. The ability to explain what the model is doing and why is thus limited. Overall the neural network models evaluated did not greatly increase the classification accuracy of service failure prediction. Therefore, the final prospective service failure model is based on the logistic regression methods.

The different logistic regression techniques used in this study produced 16 different classification models. Each of these models was evaluated based on unseen data, and it was determined that the single variable models were more robust for predicting service failures than the two-variable models. Of the eight single-variable logistic regression models evaluated, all contained more than 20 input parameters. However, the simplified eight-term model combined high accuracy and a low number of input parameters. This "practical" model was accurate and robust as well as easy to understand and implement. Therefore, the eight-term model was selected as the version to be adapted for predictive purposes. The SAS software's output of this model, detailing the process used to develop the logistic regression equation, is shown in Appendix A. However, to use this practical model in the field, it must first be transformed to a prospective prediction model.

Each of the classification models described in study were retrospective models created using a dataset in which half the records had a service failure and half did not. A transformation is needed to develop the statistical model into a prospective model that can be used to predict the location of service failures. Previous work has shown how the transformation can be done using a logistic regression model (McCullagh & Nelder 1989, Dick et al. 2003). The transformation was completed with adjustment of the model specific constant, Z, to reflect the average service failure probability across the entire system. During the four-year period, there were 12,685 service failures on the railroad that were classified according to which of the 0.01-mile segments they occurred on. In 2006, BNSF maintained 23,358 miles of mainline track (STB 2006). This corresponds to a total of approximately 2.34 million 0.01-mile-long segments. The average probability that a service failure will occur on any particular segment during a similar four-year period is thus 0.00543. This probability was converted into a new model-specific constant using the log-odds operator:

$$Z^* = Z + \ln\left(\frac{P_{SF_{AVG}}}{1-P_{SF_{AVG}}}\right) = 4.94 + \ln\left(\frac{.00543}{1-.00543}\right) = -0.270 \tag{5.7}$$

$$P_{SF2} = \frac{e^u}{(1+e^u)} \tag{5.8}$$

$$\begin{aligned} U &= Z^* - 0.0454S - 1.35R - 0.0106A + 0.00899T \\ &\quad + 0.0232L + 1.61I + 0.823G + 1.63B \end{aligned} \tag{5.9}$$

where,

Z^* = -0.270, adjusted model constant

Z = 4.94, model specific constant

P_{SF2} = probability that a service failure occurred during a four-year period

S = rail weight (in pounds per yard)

R = rail type (1 if welded, 0 if bolted)

A = rail age (in years)

T = annual traffic (in million gross tons)

L = weight of car (in tons)

I = presence of an ultrasonic defect in the last three years (1 if present, 0 otherwise)

G = presence of a geometric defect in the last three years (1 if present, 0 otherwise)

B = presence of a bridge within 200 feet of segment (1 if present, 0 otherwise)

The previous model specific constant, Z, is replaced in the logistic regression equation by the adjusted constant, Z^*, as shown in Equation 5.9. Therefore, Equation 5.7 represents the prospective service failure model, with updated value for U, for the prediction of service failures during a four-year period. This equation can be used to determine specific locations with a high likelihood of a service failure, and the overall service failure rate for a specific line.

5.7 Service Failure Prediction Case Study

The prospective prediction model can be used to calculate annual service failure rates for specific track segments and estimate the number of service failures on a particular line. This information may enable more efficient railroad maintenance planning to reduce the likelihood of broken rails. A summation of the probabilities for 100 consecutive 0.01 mile segments will yield the expected number of service failures per mile. However, the equation also calculates the probability of a service failure occurring over a four-year time frame. Therefore, assuming that service failures are distributed linearly over time, the number of expected service failures is divided by four to determine the annual rate of service failures. The transformed rate equation to calculate service failures per mile per year is:

$$E_{\mathrm{SF}} = \frac{100e^{u}}{4(1+e^{u})} \tag{5.10}$$

$$\begin{aligned} U = {} & Z^{*} - 0.0454S - 1.35R - 0.0106A + 0.00899T \\ & + 0.0232L + 1.61I + 0.823G + 1.63B \end{aligned} \tag{5.11}$$

where,

E_{SF} = expected number of service failures per mile per year on a specific segment

Z^* = -0.270, adjusted model constant

S = rail weight (in pounds per yard)

R = rail type (1 if welded, 0 if bolted)

A = rail age (in years)

T = annual traffic (in million gross tons)

L = weight of car (in tons)

I = presence of an ultrasonic defect in the last three years (1 if present, 0 otherwise)

G = presence of a geometric defect in the last three years (1 if present, 0 otherwise)

B = presence of a bridge within 200 feet of segment (1 if present, 0 otherwise)

Use of the prospective model to calculate expected service failure rate is illustrated using the following hypothetical case study (Figure 5.4, Table 5.13). The constants for this case study are that it is 3.5 miles in length, has continuously welded rail, and carries 55 million gross tons per year with an average weight of 75 tons per car. The rail weight and rail age vary along the segment. Ultrasonic defects were previously detected at mile posts 0.875 and 1.245. Geometric defects were detected at mile posts 0.485, 2.125, 2.635, and 3.335. Additionally, a 400 foot bridge is present from mile post 1.44 to 1.52.

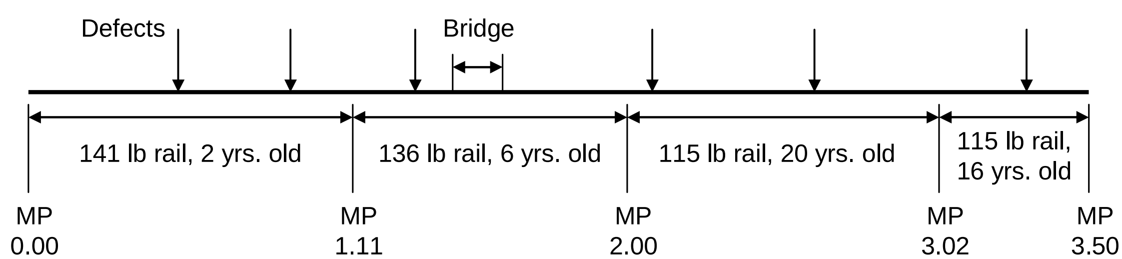

Figure 5.4 Graphical Representation of Hypothetical Case Study

The probability of a service failure for each particular 0.01 mile segment was calculated using Equation 5.8. Segments of similar characteristics were grouped together and the calculated U values are shown in Table 5.14. The number of service failures per mile per year for each segment was determined using Equation 5.10 based on each segment's U value. The expected number of service failures for each segment was then calculated by multiplying the service failure rate by the length of that particular segment. Finally, the total expected number of service failures for this 3.5-mile line was calculated

Table 5.14 Input Parameters by Mile Post for Hypothetical Cased Study

Mile Post Start	Mile Post End	Z*	S	R	A	T	L	I	G	B	U
0.00	0.48	-0.27	141	1	2	55	75	0	0	0	-5.81
0.48	0.49	-0.27	141	1	2	55	75	0	1	0	-4.99
0.49	0.87	-0.27	141	1	2	55	75	0	0	0	-5.81
0.87	0.88	-0.27	141	1	2	55	75	1	0	0	-4.20
0.88	1.11	-0.27	141	1	2	55	75	0	0	0	-5.81
1.11	1.24	-0.27	136	1	6	55	75	0	0	0	-5.62
1.24	1.25	-0.27	136	1	6	55	75	1	0	0	-4.01
1.25	1.40	-0.27	136	1	6	55	75	0	0	0	-5.62
1.40	1.56	-0.27	136	1	6	55	75	0	0	1	-3.99
1.56	2.00	-0.27	136	1	6	55	75	0	0	0	-5.62
2.00	2.12	-0.27	115	1	20	55	75	0	0	0	-4.82
2.12	2.13	-0.27	115	1	20	55	75	0	1	0	-4.00
2.13	2.63	-0.27	115	1	20	55	75	0	0	0	-4.82
2.63	2.64	-0.27	115	1	20	55	75	0	1	0	-4.00
2.64	3.02	-0.27	115	1	20	55	75	0	0	0	-4.82
3.02	3.33	-0.27	115	1	16	55	75	0	0	0	-4.78
3.33	3.34	-0.27	115	1	16	55	75	0	1	0	-3.95
3.34	3.50	-0.27	115	1	16	55	75	0	0	0	-4.78

by summing over all of the segments. This segment of track is expected to have 0.54 service failures in the next year (Table 5.15). The changes in service failure rates due to various factors across the line segment are shown in Figure 5.5.

5.8 Conclusions

20 different prediction models were developed to predict service failures; including several different logistic regression and ANN models. Service failure data from BNSF's network were used for a four-year time period. A previous service failure classification model using logistic regression, that incorporated only track and traffic characteristics, was evaluated and determined to have limited predictive ability for current service failure data. New logistic regression models were developed that included additional factors such as infrastructure data, maintenance activities, and rail testing results. The logistic regression models were constructed using various techniques and each model was tested against validation data. A practical logistic regression model was also developed that reduced the complexity of the model and maintained a high level of accuracy. This practical model was determined to be 64.7% accurate for classifying track segments. An ANN model was also developed to classify cases as either failure or

Table 5.15 Calculation of Service Failures per Year and Total Expected Service Failures for Hypothetical Case Study

Mile Post Start	Mile Post End	Length	U	Service Failures per mile per year	Expected Service Failures
0.00	0.48	0.48	-5.81	0.075	0.036
0.48	0.49	0.01	-4.99	0.170	0.002
0.49	0.87	0.38	-5.81	0.075	0.028
0.87	0.88	0.01	-4.20	0.370	0.004
0.88	1.11	0.23	-5.81	0.075	0.017
1.11	1.24	0.13	-5.62	0.090	0.012
1.24	1.25	0.01	-4.01	0.444	0.004
1.25	1.40	0.15	-5.62	0.090	0.013
1.40	1.56	0.16	-3.99	0.453	0.072
1.56	2.00	0.44	-5.62	0.090	0.040
2.00	2.12	0.12	-4.82	0.200	0.024
2.12	2.13	0.01	-4.00	0.452	0.005
2.13	2.63	0.5	-4.82	0.200	0.100
2.63	2.64	0.01	-4.00	0.452	0.005
2.64	3.02	0.38	-4.82	0.200	0.076
3.02	3.33	0.31	-4.78	0.209	0.065
3.33	3.34	0.01	-3.95	0.471	0.005
3.34	3.50	0.16	-4.78	0.209	0.033
				TOTAL EXPECTED	**0.541**

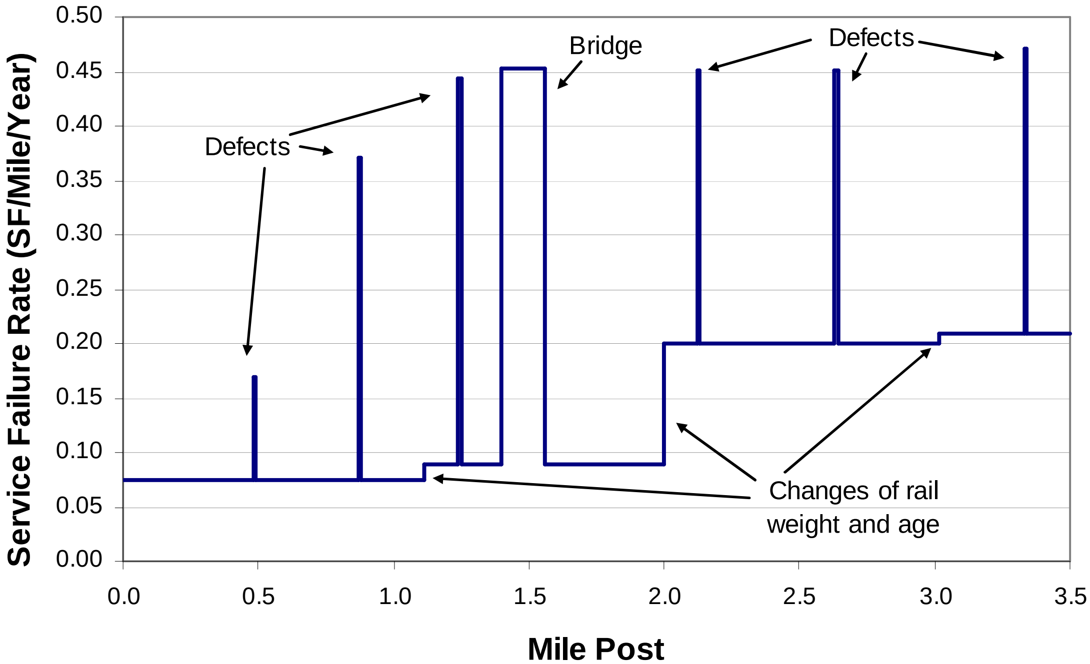

Figure 5.5 Service Failure Rate vs. Mile Post for Hypothetical Case Study

non-failure. Additionally, two ANN/LR hybrid classification models were developed. Each of the three advanced models performed only slightly better than the traditional logistic regression techniques. Finally, the practical logistic regression model was transformed into a prospective prediction model based on the overall probability of service failures on the BNSF network. This prospective prediction model was then used in a hypothetical case study to examine how this prediction tool can be used to evaluate specific track segments.

This analysis revealed that service failures can only be predicted with accuracy levels ranging from about 65% to 70% using the data available for this study and the methods presented. The different methods and techniques used to find the best prediction model had only slightly different classification accuracies, indicating that 30% or more of the cases presented here were not able to be correctly classified from the available data. This indicates that the models developed were not accounting for about 30% to 35% of the variance. Examples of these sources of variance may include the occurrence of thermally-induced stress in the rail, the frequency and magnitude of dynamic loading events from out-of-round wheels, and other characteristics of rail steel and fatigue-crack growth.

The models developed in this study are intended to assist railroads to more effectively allocate resources to prevent the occurrence of broken rails. The models can be implemented in two different ways involving maintenance planning. They can be used tactically for short-term maintenance assistance, such as determining specific track segments to monitor closely or repair. They can also be used strategically for long-term maintenance planning and renewal activities. The two most common prevention techniques for broken rails are rail grinding and rail replacement. Both of these activities require long lead times for planning and have high associated costs.

5.8.1 Future Work on Service Failure Prediction

This analysis included many of the available variables that potentially affect the growth of defects and the occurrence of broken rails. However, as discussed above, some additional factors that could be considered include climatic data for track locations, track inspection frequency, and density of service failures and defects. Climate effects, especially in areas of continuously welded rail experience high tensile stress that may affect the growth of rail defects and the occurrence of broken rails. Also, more frequent track inspection probably reduces service failure occurrence because cracks are more likely to be detected and repaired (Zarembski & Palese 2005). Additionally, it may be possible to determine the overall density of service failures and detected defects for each particular piece of rail, thereby increasing the accuracy of the prediction model.

Future work might also include determination of an optimized statistical model that considers the trade-off between the number of parameters and the accuracy of prediction. This is based on the idea that as additional parameters are added to the model there may be an incremental cost. The incremental cost to add each parameter may or may not be linear. However, as the level of accuracy is increased there should be a savings due to the reduced occurrence of broken rails. With this information a "utopia point" solution balancing cost versus benefit could be developed to find the optimal model (Marler & Arora 2004).

Another area of possible future work would be to evaluate different artificial neural network models. This includes different learning techniques and objective functions to determine the optimal neural network. Experimenting with different numbers of neurons in the hidden layer may also lead to a more accurate ANN, but care must be taken to not over-fit the data with additional factors. Finally, the use of neuro-fuzzy networks might be possible to apply to this topic. Fuzzy Neural Networks (FNN) may be able to classify each case as well as produce an output value regarding how strongly each case is in its respective classification. This output value may be a close approximation to determine the probability of failure; a value a simple ANN cannot produce.

CHAPTER 6: ECONOMIC IMPACT OF BROKEN RAILS

The purpose of this study was to understand the economic impact of service failures, broken rail derailments, and their respective prevention techniques. In Chapter 5 I considered factors and various analytical techniques to predict locations that have a high probability of broken rail occurrence. The most important factors were rail weight, rail type, rail age, annual traffic, average weight of cars, presence of an ultrasonic defect, presence of a geometric defect, and the presence of a bridge. However, understanding where broken rails are most likely to occur is necessary, but not sufficient for cost-effective management of the problem. Additional information on the economic impact of broken rails as well as the cost and effectiveness of various preventive strategies is also needed. In this chapter I quantify the cost of broken rails. In particular, costs associated with broken rail derailments, service failures, train delay, and typical prevention measures are examined. The results of this study are intended to assist railroads to make better informed decisions regarding maintenance and prevention of broken rails.

6.1 Introduction to Economic Study of Broken Rails

Broken rails are generally caused by either internal or surface defects in the rail (Sperry Rail Service 1999). Internal defects are generally present due to the growth of minute flaws introduced during rail manufacture. These flaws grow due to cyclic vertical and horizontal loading of the rail and consequent fatigue crack growth. Surface defects are generally caused by wheel-rail contact stresses at the running surface of the rail from passing trains. In either case, if a flaw is allowed to grow large enough, at some point the rail may be subject to a load that it does not have sufficient strength to withstand and it fractures. Broken rails are separated into two categories, those that result in a derailment and those that are detected by some other means and are typically called “service failures”. The consequences of these two events are quite different, but understanding their economic impact is important to making better informed decisions regarding their prevention. Overall, broken rails were responsible for 335 mainline derailments on Class I freight railroads from 2003 through 2006 (FRA 2007a). These derailments resulted in over $176 million of equipment and track damage. However, the economic impact of broken rails includes many other costs besides track and equipment damage.

Economic analysis of railroad engineering and operations is a topic that has been the subject of extensive study for over a century and a half. Among the most well-known early treatises on railway economics was by Wellington (1887). More recent research has focused on specific topics of railway economics; some that is applicable to this analysis, includes the expected life of rail and rail renewal (Zhao et al. 2006, Ling 2006). However, the economic costs specifically associated with broken rails have not previously been quantified. The potential costs associated with broken rails have been explored as part of an overall analysis of rail defects (Cannon et al. 2003). Cannon et al. stated that the cost of broken rails includes inspection of track, train delay, remedial treatments, pre-emptive treatments, derailments, and loss of business.

The cost of a specific broken rail event will vary based on many factors, but the intention of this analysis was to calculate typical expected costs based on past averages of similar events. However, some costs, such as loss of business, were difficult to quantify, while others could not be obtained because of the sensitivity of the information. The objectives of this analysis were to:

- Quantify the costs associated with broken rail derailments and service failures,
- Determine costs associated with train delay,
- Develop a train delay cost calculator based on the density of a line, and
- Quantify the costs associated with preventive measures.

6.2 Costs Associated with Broken Rail Derailments

The economic impact of a broken rail derailment can be severe. Such accidents are also a major disruption to railroad operations. Railroads spend a great deal of time and money in their efforts to prevent broken rails. The costs associated with broken rail derailments include track damage, equipment damage, accident clean-up, labor and materials for repair, train delay, lading damage, and loss of future business. Railroads are generally apprehensive about sharing complete information on derailment related costs; but some information is publicly available (FRA 2007a). This information was supplemented by interviews with railroad industry experts and other research to further understand the associated costs.

6.2.1 Track and Equipment Damage of Broken Rail Derailments

The Federal Railroad Administration (FRA) requires that railroads report equipment and track damage for railway accidents that exceed a specified monetary threshold ($7,700 in 2006). The cost of equipment damage includes any repair or replacement of on-track equipment such as cars, locomotives, and maintenance equipment. This also includes any necessary labor or materials needed for equipment repair or replacement. Track damage reported to the FRA is a broad category, as it includes the costs associated with repair or replacement of any track, signals, or track structures, such as bridges and grade crossings, including any labor and materials needed for repair or replacement. Track damage also includes the costs of accident clean-up, such as clearing the right of way of damaged cars, spilled lading, and the cost of third-parties contracted to assist with accident clean-up. The financial impact associated with environmental and hazardous material clean-up due to an accident is not included in the FRA reportable costs.

FRA accident data are publicly available at the FRA Office of Safety website. Train accident data were downloaded and analyzed from 2003 through 2006 (FRA 2007a). The FRA defines 14 unique accident cause codes related to broken rails (FRA 2007b). During the four-year study period, U.S. Class I freight railroads experienced 335 mainline broken rail derailments (Table 6.1). The total costs of these derailments exceeded $176 million in equipment and track damage, for an average cost of $525,400 per incident. By contrast, an examination of siding derailments revealed that there had been 40 derailments during the same interval with an average cost of $76,490 per incident (Table 6.2).

6.2.2 Other Related Broken Rail Derailment Costs

Besides reportable equipment and track damage, another important cost associated with broken rail derailments is train delay. Train delay cost is based on the time of track-outage as well as the number of trains delayed. A train delay cost calculator

Table 6.1 Equipment and Track Damage for Class I U.S. Freight Railroad Mainline Broken Rail Derailments, 2003-2006

FRA Code	Cause Description	Frequency	Total Cost	Cost Per Incident
T201	Bolt hole crack or break	14	$12,854,596	$918,185
T202	Broken Base	26	5,804,332	223,244
T203	Broken Weld (plant)	3	1,026,794	342,265
T204	Broken Weld (field)	25	17,338,957	693,558
T207	Detail fracture from shelling or head check	82	52,792,131	643,806
T210	Head and web separation (outside joint bar)	23	5,380,144	233,919
T211	Head and web separation (within joint bar)	7	2,042,042	291,720
T212	Horizontal split head	5	1,967,657	393,531
T213	Joint bar broken (compromise)	6	3,111,204	518,534
T214	Joint bar broken (insulated)	9	8,152,304	905,812
T215	Joint bar broken (noninsulated)	15	14,225,856	948,390
T219	Rail defect with joint bar repair	1	664,622	664,622
T220	Transverse/compound fissure	90	42,315,036	470,167
T221	Vertical split head	29	8,333,190	287,351
		335	**$176,008,865**	**$525,400**

Table 6.2 Equipment and Track Damage for Class I U.S. Freight Railroad Siding Broken Rail Derailments, 2003-2006

FRA Code	Cause Description	Frequency	Total Cost	Cost Per Incident
T201	Bolt hole crack or break	1	$14,688	$14,688
T202	Broken Base	4	211,636	52,909
T203	Broken Weld (plant)	0		
T204	Broken Weld (field)	0		
T207	Detail fracture from shelling or head check	11	1,303,978	118,543
T210	Head and web separation (outside joint bar)	0		
T211	Head and web separation (within joint bar)	0		
T212	Horizontal split head	0		
T213	Joint bar broken (compromise)	1	15,237	15,237
T214	Joint bar broken (insulated)	0		
T215	Joint bar broken (noninsulated)	1	128,879	128,879
T219	Rail defect with joint bar repair	0		
T220	Transverse/compound fissure	16	975,877	60,992
T221	Vertical split head	6	409,321	68,220
		40	**$3,059,616**	**$76,490**

was developed as part of this study and is described in a later section. Discussions with industry experts indicated that the time of track-outage from a broken rail derailment depends on the situation. Some of the factors that affect track-outage time are the severity of the accident, access to the site, if hazardous materials were involved, or if the accident is near a metropolitan area. Track damage due to broken rail derailments is typically restricted to about 500 to 1,500 feet as a result of the cars piling up. A moderate to large scale broken rail derailment will take approximately 24 hours to return the track to service. One expert stated that if the expected outage is several days or more, then arrangements to reroute trains will be made if possible.

Other costs associated with broken rail derailments include lading damage and loss of business or customers. The cost of lading damage is not required to be reported to the FRA and therefore is not publicly available. The cost of lading damage was unavailable and is highly variable. Depending on what is lost or damaged, it can vary from a few thousands to millions of dollars per incident. The loss of future business due to broken rail derailments is difficult to quantify.

6.3 Costs Associated with Service Failures

The next step was to examine the costs associated with service failures. Service failures have a much lower economic impact than broken rail derailments, but occur much more frequently. One major Class I railroad experienced 3,171 service failures and 19 broken rail derailments per year from 2003 through 2006 (167:1 ratio). Generally, service failures are detected by the signal system, a track inspector, or a train crew. Once detected, trains typically do not proceed over that section of track until the rail has been repaired. Although, FRA regulations allow trains to be “walked” over a broken rail while the break is monitored by a qualified railroad employee, this practice is not generally used by the major U.S. railroads. Instead, trains are halted and a repair crew is dispatched to remove and replace the broken rail.

The costs associated with service failures include material, labor, and train delay costs. One railroad industry expert stated that the average material and labor cost for a service failure is $1,500, which includes mobilization of the crew and materials. Another railroad provided further details based on their estimates for average labor and materials for rail repair. The estimated material cost of a 15 foot section of 136 pound rail and two welds is $373.60 (Table 6.3). The estimated labor cost for removing the old rail, unloading and placing new rail, installation of the requisite other track material (OTM), and installing field welds totals $370.00 (Table 6.3). The combined total cost, not including mobilization, is $743.60. The difference between this value and the $1,500 estimate is the cost of mobilizing the labor and materials needed for the repair. A number of factors must be considered to evaluate the cost of mobilization, such as the time of day, time of year, location of service failure, and availability of materials.

Table 6.3 Estimated Labor and Material Repair Cost for a Service Failure

Item	Cost per Unit	Cost
136 lb. CWR rail	$17.84 per foot	$267.60
Welding kit	53.00 each	106.00
	Total Material Cost	373.60
Remove & load old rail	3.12 per foot	46.80
Unload new rail and OTM	0.49 per foot	7.35
Place new rail and OTM	2.19 per foot	32.85
Install field welds	141.50 each	283.00
	Total Labor Cost	370.00
	TOTAL REPAIR COST	**$743.60**

The final cost associated with service failures is the cost of train delay. The length of the delay will be affected by a number of factors. Railroad industry experts indicated that a typical service failure will result in approximately a four-hour track-outage, from initial notification of the failure until the line is reopened for normal operation. Again, the delay cost will also depend on the number of trains delayed and can be estimated using the train delay calculator described in the next section.

6.4 Train Delay Cost Calculator

Train delay cost is affected by both broken rail derailments and service failures. The total cost due to train delay is based on the cost of delay per train-hour, the number of trains delayed, and the total length of the delay. Industry experts estimated that the cost due to delay of a single train is in the range of $200 to $300 per train-hour. The purpose of this analysis was to calculate an updated value based on operating averages for all U.S. Class I railroads. Additionally, a formula was developed to determine the cost of delay based on the number of trains delayed and their length of delay based on any given line density.

6.4.1 Calculation of Train Delay Cost per Train-hour

Interviews with industry experts led to the conclusion that single-train delay cost per train-hour includes four components: car cost, locomotive cost, fuel cost, and crew labor cost. Car delay cost refers to the cost of railroad-owned cars that are delayed and therefore cannot be used elsewhere. Privately owned cars are excluded from this analysis because, in many cases, they are charged by the mile and do not have a direct cost to railroads if delayed. The average number of cars per train in 2006 was 69.2 cars (AAR 2006) and 39.8% were railroad owned (AAR 2007). To determine the cost per car, an average car-hire rate was used. Industry experts indicated that a reasonable estimate for this was $0.75 per car-hour in 2006. The total average car delay cost per train was computed to be about $20.67 per hour in 2006 (Table 6.4).

The second component of train delay cost is that associated with delay of locomotives. Similar to car delay, locomotive delay was determined by estimating the opportunity cost due to unavailability of locomotives for other applications. This value can be estimated based on the locomotive depreciation that occurs during the time of delay. The average numbers of locomotives per train in 2006 was 2.7 (AAR 2006). A 2008 survey of Class I railroad data revealed that the average cost of a new road locomotive, which varies greatly based on type, was approximately $1,877,500 and the salvage value after 25 years of road life was approximately $250,000 (Murray 2008). Assuming a discount rate of 10%, the annual locomotive depreciation per locomotive for 2006 was $209,383. Therefore, the locomotive cost per locomotive-hour was $23.90 or $64.54 per train-hour (Table 6.4).

The third component of train delay is the cost of fuel consumed during the delay. The cost of diesel fuel purchased by Class I railroads in 2006 was $1.93 per gallon (AAR 2006). The average fuel consumed per locomotive-hour in idle was approximately 3.5 gallons based on information provided for different locomotive types from a major U.S.

Table 6.4 Breakdown of Train Delay Cost per Train-hour

Car Cost	
Average number of Cars per train	69.2
Car Hire per hour	$0.75
Percent of cars owned by railroad	39.8%
Total per Train-hour	**$20.67**
Locomotive Cost	
Average number of Locomotives per train	2.7
Average cost of new locomotive	$1,877,500
Average Life of locomotive	25
Average salvage value	$250,000
Discount Rate	10%
Average cost per locomotive year	$209,382.57
Total per Train-hour	**$64.54**
Fuel Cost	
Average number of Locomotives per train	2.7
Gallons per Hour In idle	3.5
Cost per Gallon	$1.93
Total per Train-hour	**$18.24**
Crew Cost	
Number of employees per train	2
Average Hourly Pay	$21.40
Average Overtime Pay	$31.45
Percent of wage for fringe benefits	75%
Total per Train-hour	**$110.08**
TOTAL COST PER TRAIN HOUR	**$213.52**

freight railroad. Based on an average of 2.7 locomotives per train, the total fuel cost was $18.24 per train-hour in 2006 (Table 6.4).

The fourth component of train delay is labor cost. Average hourly wages for train and engine crews for Class I U.S. freight railroads in 2006 was $21.40 for straight time and $31.45 for overtime pay (STB 2008). For the calculation of labor cost, only the overtime rate was used based on the assumption that a train delay of more than a few hours will generally result in overtime pay for the train crew. Additionally, the labor cost includes fringe benefits, such as vacation pay, holiday pay, railroad retirement, unemployment, health welfare, and group life insurance. Fringe benefits are estimated to be approximately 75% of wages in 2006. Therefore, based on a two-person train crew, the labor cost of delay was $110.08 per train-hour (Table 6.4).

A summation of the four components of train delay yields a total cost of $213.52 per train-hour in 2006. However, this estimate is still only a partial estimate because additional costs due to train delay are not considered. For example, in some delay situations crews must be replaced due to federal hours-of-service regulations that limit crews to a maximum period on duty of 12 hours. There may also be some extra stopping and starting of the train resulting in extra fuel consumption and wear and tear on brakes and other components.

6.4.2 Cost of Multiple Train Delay

The number of trains delayed and the duration of their delay during a track-outage must be considered in the calculation of train delay cost. These values can be approximated based on the density of the line and the number of mainline tracks. To determine the number of trains delayed, I assumed that trains will arrive in constant time intervals from both directions. The average train operated for Class I U.S. railroads was 6,312 gross tons, including cars and locomotives, in 2006 (AAR 2006). The number of trains per year for a particular line is the annual gross tonnage (in millions) (ANMGT) of that line divided by 0.006312 million-tons per train. The interval between trains, *t*, was determined by dividing the number of hours per year, 8,766, by the number of trains per year:

$$n = \text{Number of trains per year} = \frac{\text{Annual MGTs}}{\text{tons per train (millions)}} = \frac{ANMGT}{0.006312} \tag{6.1}$$

$$t = \text{hours per train arrival} = \frac{\text{hours per year}}{\text{trains per year}} = \frac{8{,}766}{n} = \frac{55.33}{ANMGT} \tag{6.2}$$

The total cost of train delay can then be calculated by the cost of delay per train-hour and the hours per train arrival of the particular line. The total number of trains delayed is determined by dividing the total delay time by the hours per train arrival. The length of delay for each train is based on the time of their respective arrival. The total cost due to train delay from a service interruption can be calculated using the following formula:

$$C = Tx + \sum_{n=1}^{m}(T - nt)x \tag{6.3}$$

where,

C = total train delay cost for multiple trains
T = total delay time for service interruption
x = cost of delay per train-hour ($213.52)
m = number of following trains delayed = T / t (rounded to the nearest integer)
t = hours per train arrival = *55.33 / ANMGT*

The total train delay cost presented in Equation 6.3 is valid for broken rail scenarios in which no trains can proceed due to the event. This would include service failures on single track territory or broken rail derailments on multiple track or single track territory because no trains would be able to proceed. However, Equation 6.3 must be adjusted for situations in which a service failure occurs on one track in multiple track territory. In these situations trains will be able to proceed on the other mainline track and there will not be complete (100%) delay. It can be assumed that a service interruption on a single track may cause up to half the trains (50%) to be delayed (i.e. traffic in one direction stops). However, in most cases, less than 50% of trains would be delayed, and the actual amount of delay is dependent on the density of the line.

A sensitivity analysis was conducted on a hypothetical case study to evaluate how Equation 6.3 can be used to evaluate train delay cost of broken rail events on various density lines (Table 6.5). The case study assumed a single track mainline with either a service failure with a track outage of 4 hours or a broken rail derailment with a track outage of 24 hours. Using Equation 6.3, the train delay cost on a 60 MGT line was calculated to be $2,235 for a service failure or $69,244 for a broken rail derailment.

Table 6.5 Sensitivity Analysis of Train Delay Costs

Annual Gross	Broken Rail Event	
Tons (Million's)	Service Failure	Derailment
15	$854	$19,332
30	1,314	35,902
45	1,775	52,472
60	2,235	69,244
75	2,695	85,853

6.5 Costs Associated with Broken Rail Preventive Measures

To determine the economic impact of broken rails, the costs associated with preventive measures must also be examined. Typical, preventive measures include rail inspection for defects, rail grinding, and rail replacement and renewal. These preventive measures decrease the likelihood of a broken rail; however, they are also used to extend the life of the overall track structure. For example, rail grinding improves the overall wheel-rail interface and the allocation of the cost related to broken rail prevention cannot be determined. Therefore, the cost of broken rail preventive measures is an indirect cost related to broken rails (Table 6.6).

Table 6.6 Estimated Indirect Costs of Broken Rails

Broken Rail Prevention Technique	Annual Cost per Track Mile ($)
Ultrasonic and Geometric Track Inspection	900
Rail Grinding	1,900
Rail Surfacing	700
Rail Renewal and Replacement	2,500
TOTAL INDIRECT COST	**6,000**

One of the most effective broken rail prevention measures is the use of ultrasonic and geometric inspection of track (Zarembski & Palese 2005). Its cost is dependent on the frequency of inspections. The BNSF Railway uses a risk-based approach for inspection frequency (Palese & Zarembski 2001). As defined by BNSF, the calculated risk factor of any particular line depends on many variables, including the number of previously detected defects, if the line carries passengers and/or hazardous materials, and the railroad-determined “importance” of the line. Ultrasonic and geometric inspections are estimated to cost approximately $900 per track mile in 2007 (Table 6.6).

Class I U.S. freight railroads also use rail grinding to prolong the life of rail and to eliminate surface defects. Grinding locations are typically based on the life of the rail, the density of the line, and the number of previously detected surface defects. Rail grinding is estimated to cost approximately $1,900 per track-mile (Table 6.6). Additionally, railroads complete rail surfacing projects to maintain stable and properly aligned track structure for many reasons, including slowing potential crack growth. These costs are estimated to be approximately $700 per track-mile for capital surfacing projects (Table 6.6). Finally, rail renewal and replacement projects are crucial for large railroads to prevent broken rails and other track-related accident causes. However, rail replacement occurs due to both wear and fatigue crack growth, so allocation of this expense is not possible without knowing the percentage of rail replaced due to these different causes. The average amount of rail replaced by Class I U.S. railroads in 2006 was 3.25 tons per track-mile (AAR 2006). One railroad reported that the cost of 141-lb. rail in 2006 was approximately $760 per ton. Therefore, the cost of rail renewal is approximately $2,500 per track-mile (Table 6.6). Additional information on preventive measures, such as the cost of local rail surfacing projects, was difficult to determine and are not included in this analysis.

6.6 Conclusions

The costs associated with broken rails are a substantial concern to all major U.S. railroads. Based on this analysis, an average broken rail derailment will have a direct economic impact to the railroad of $550,000 or higher depending on the density of the line, plus indirect costs associated with preventive measures. A service failure on a medium density line will incur a direct cost of at least $2,500, plus the indirect costs of prevention. Broken rail derailments from 2003 through 2006 resulted in a total FRA reportable damage to equipment and track of $176 million, or approximately $44 million per year; however, the economic impact of broken rails is greater when the additional costs considered in this analysis are included. Assuming that service failure rates are similar for all Class I U.S. railroads and that the average line traffic density is 30 MGTs, the annual direct cost of broken rails for all U.S. Class 1 railroads was approximately $83 million (Table 6.7). The annual indirect cost from broken rail preventive activities was approximately $855 million, based on a reported 142,428 track-miles operated by U.S. Class I freight railroads (AAR 2006) (Table 6.7).

Table 6.7 Estimated Annual Broken Rail Costs for U.S. Class I Railroads

Broken Rail Associated Cost	Cost ($)
Derailment Damage	44,002,216
Derailment Delay Cost	3,006,793
Service Failure Repair Cost	19,141,747
Service Failure Delay Cost	16,768,170
TOTAL DIRECT COST	**82,918,926**
Track Inspection	128,185,200
Rail Grinding	270,613,200
Rail Surfacing	99,699,600
Rail Renewal and Replacement	356,070,000
TOTAL INDIRECT COST	**854,568,000**

The values presented in this analysis are overall averages for Class I U.S. freight railroads and will vary based on the circumstances of each broken rail event. Pertinent factors include the severity, location, density of line, and availability of labor and materials. The methodologies developed here could be used by railroads to estimate the costs associated with broken rails based on their own line and operating characteristics.

6.6.1 Future Work on Economic Impact of Broken Rails

Additional work on this topic could include research on hard-to-quantify values associated with broken rails. Some of these factors include the average cost of lading loss in a broken rail derailment, the financial impact of loss of business from accidents and train delays, and the costs associated with rail surfacing at the local level. Additional future research might also be useful to further understand and refine the costs associated with train delay. A better estimate for the cost of locomotive delay may be based on leasing rates for locomotives or number of additional locomotives purchased to compensate for delays. It also may be possible to refine the car delay cost to include privately owned cars.

CHAPTER 7: CONCLUSIONS & FUTURE RESEARCH

Understanding the factors affecting the likelihood of train accidents is essential for identifying and implementing the most effective and efficient accident reduction measures. In Chapter 3 I presented an analysis of the effect of train length on accident rates. Train accident likelihood is dependent on both car-mile and train-mile-related causes and FRA train accident causes were classified into these two categories using a new quantitative metric. Updated mainline car-mile and train-mile-related accident rates were calculated for U.S. Class I freight railroads using the reclassified accident causes. In 2005 the accident rate for car-mile-related causes was 1.05×10^{-8} or about 0.011 accidents per million car-miles, and the train-mile-related accident rate was 8.62×10^{-7} or about 0.86 accidents per million train-miles. The model developed here enables quantification of the effect of operational changes in train length on accident rate at both the individual train, and system-wide levels.

In Chapter 4 I presented an analysis of recent train accident data and found that broken rails are the leading cause of major derailments on U.S. railroads. A previous study by Dick (2001) used logistic regression and data on 11 possible track and traffic characteristics to develop a statistical model to predict service failures. I conducted further evaluation of Dick's model using unseen data and found it to be robust for the time period and dataset used in his study. Additionally, artificial neural networks (ANNs) and hybrid ANN/Logistic Regression (ANN/LR) models were used in an attempt to improve the accuracy of the logistic regression model. Although neither ANNs nor ANN/LR models improved the predictive performance, all the models had a similar level of accuracy and were robust for unseen cases.

In Chapter 5 I extended Dick's (2001) study, by conducting an analysis using a new, more recent dataset that included an expanded number of factors with the potential to affect rail crack growth rate and service failure occurrence. The new dataset included 28 factors on track and rail characteristics, infrastructure features, maintenance activities, and on-track testing results. The predictive accuracy of Dick's (2001) model was reduced when applied to the new data. Consequently several types and variations of modeling techniques were tried, including logistic regression, ANNs, and hybrid

ANN/LRs. Each was evaluated for accuracy and robustness to determine the best prediction model. A "practical" logistic regression model was ultimately selected because it was accurate, understandable, and useable. This model included only eight parameters and had a predictive accuracy of 64.7%. The most important factors related to service failures were rail weight, rail type, rail age, annual traffic, weight of car, presence of an ultrasonic defect, presence of a geometric defect, and the presence of a bridge. A prospective prediction model was developed based on service failure rates for the BNSF Railway. The objective of the models developed in this analysis was to provide railroads with tools to help them identify locations with a high likelihood of service failure occurrence and more effectively allocate resources to prevent them.

In Chapter 6 I presented an analysis of the economic impact of broken rails. From 2003 through 2006 broken rails were responsible for 335 FRA-reportable mainline derailments on Class I freight railroads, or about 84 per year. The average cost of damage to track and equipment from these accidents was $525,000 per incident and the average annual cost was $44 million. In addition to the FRA-reportable costs, expenses due to broken-rail-accident-caused train delay are estimated to be about $3 million per year. Railroads also incur about $19 million per year for repair of service failures, and an additional expense of $17 million due to the train delay that results from their occurrence. The costs of preventive measures, such as rail inspection, rail grinding, track surfacing, and rail replacement were estimated to be approximately $855 million per year, but because these activities provide multiple benefits to railroads it was not possible to determine what share of this amount should be allocated to broken rail prevention.

7.1 Future Research

Future research specific to the topics presented in this thesis is described at the end of each chapter. However, some of the most important topics are summarized here.

The analysis presented in Chapter 3 was based on a binary classification of accident causes that assumed that they could be classified as either car-mile or train-mile-related. Future work regarding train accident causes may reveal that some accidents are a function of both car and train-miles. Therefore, it may be possible to develop a function for each cause group that accounts for the effect of both. This work might further refine the understanding of train accidents and train accident rates.

It may also be possible to improve the accuracy of the service failure prediction models. The models presented in this analysis had an accuracy ranging from about 65% to 70% for prediction of service failures leaving about 30% to 35% of the variance unexplained. Examination of additional factors that affect crack growth may help to understand this variance. Some of the factors that could be considered include location-specific climatic data, flat wheel incidence, and track inspection data.

Incorporation of climatic data may enable quantification of cyclic longitudinal loading due to fluctuations in thermal stress. It may also be possible to determine the extent of cyclic vertical loading that rail at specific locations experiences under various degrees of tensile stress.

Out-of-round (flat) wheels are known to affect crack growth in rails. Therefore, incorporation of wheel impact load detector (WILD) data might enable the occurrence and magnitude of these loads to be included as a parameter in a prediction model.

More sophisticated use of ultrasonic and geometric inspection data might also improve the predictive ability of the model in two ways. First, the incidence of broken rails is commonly believed to be related to the rate of both rail defect and service failure occurrence, but neither of these was accounted for in the models that were developed. Second, inspection frequency was also not accounted for. This parameter might be inversely related to service failures because problems are more likely to be detected and corrected before a broken rail occurs.

Finally, it may also be possible to apply the modeling techniques explored in this thesis to other accident causes. For example, the likelihood of track buckling, also known as “sun kinks”, is affected by a variety of factors. The same multivariate modeling techniques that were applied to service failure prediction in this thesis might be able to be adapted to determine the factors related to the occurrence of track buckles.

REFERENCES

Aglan, H. & Gan, Y.X. (2001). Fatigue crack growth analysis of a premium rail steel. *Journal of Materials Science*, vol. 36, no. 2, pp. 389-397.

Anderson, R.T. (2005). Quantitative analysis of factors affecting railroad accident probability and severity. *Masters Thesis*. University of Illinois at Urbana-Champaign.

Anderson, R.T. & Barkan, C.P.L. (2004). Railroad accident rates for use in transportation risk analysis. *Transportation Research Record*, no. 1863, pp. 88-98.

Anderson, R.T. & Barkan, C.P.L. (2005a). Derailment probability analysis and modeling of mainline freight trains. *Proceedings of the 8th International Heavy Haul Conference*, Rio de Janeiro, June 2005, pp. 491-497.

Anderson, R.T. & Barkan, C.P.L. (2005b). Train accidents as functions of train-miles and car-miles. Working paper, University of Illinois at Urbana-Champaign.

Arthur D. Little, Inc. (ADL) (1996). *Risk Assessment for the Transportation of Hazardous Materials by Rail, Supplementary Report: Railroad Accident Rate and Risk Reduction Option Effectiveness Analysis and Data (Second Revision).* ADL, Cambridge, Mass.

Association of American Railroads (AAR) (2006). *Analysis of Class I railroads*. Association of American Railroads, Washington, D.C.

Association of American Railroads (AAR) (2007). *Railroad Facts, 2000-2006 editions.* Association of American Railroads, Washington, D.C.

Barkan, C.P.L., Dick, C.T., & Anderson, R.T. (2003). Railroad derailment factors affecting hazardous materials transportation risk. *Transportation Research Record*, no. 1825, pp. 64-74.

Ben-Akiva, M. & Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Mass.

Bouteiller, F., Grisso, B.L., Peairs D.M., & Inman, D.J. (2006). Broken rail track detection using smart materials. *Proceedings of SPIE Smart Structures and Materials/NDE*, San Diego, CA, Feb. 26 - Mar. 2.

Cannon, D.F., Edel, K-O., Grassie, S.L., & Sawley, K. (2003). Rail defects: an overview. *Fatigue & Fracture of Engineering Materials & Structures*, vol. 26, no. 10, pp. 865-886.

Carthey, J., de Leval, M.R., Wright, D.J., Farewell, V.T., Reason, J.T. & all UK Paediatric Cardiac Centers (2003). Behavioural markers of surgical excellence. *Safety Science*, vol. 41, no. 5, pp. 409-425.

Center for Chemical Process Safety (CCPS) (1995). *Guidelines for Chemical Transportation Risk Analysis*. American Institute of Chemical Engineers, New York.

Cody, R. & Smith, J.K. (1997). *Applied Statistics and the SAS Programming Language.* Prentice-Hall, Upper Saddle River, New Jersey.

da Silva, L.F.M., Stewardson, D.J., de Oliveira, F.M.F., & de Castro, P.M.S.T. (2003). Fatigue crack growth of rails for railways. *Proceedings of the Institution of Mechanical Engineers, Part F - Journal of Rail & Rapid Transit,* vol. 217, no. 2, pp. 89-97.

Dennis, S.M. (2002). Changes in railroad track accident rates. *Transportation Quarterly,* vol. 56, no. 4, pp. 161-174.

Dick, C.T. (2001). Factors affecting the frequency and location of broken railway rails and broken rail derailments. *Masters Thesis*. University of Illinois at Urbana-Champaign.

Dick, C.T., Barkan, C.P.L., Chapman, E., & Stehly, M.P. (2003). Multivariate statistical model for predicting occurrence and location of broken rails. *Transportation Research Record,* no. 1825, pp. 48-55.

Dougherty, M. (1995). A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies,* vol. 3, no. 4, pp. 247-260.

Fanning, K.M. & Cogger, K.O. (1994). A comparative analysis of artificial neural networks using financial distress prediction. *International Journal of Intelligent Systems in Accounting, Finance, and Management*, vol. 3, pp. 241-252.

Federal Railroad Administration (FRA) (2006). *Railroad Safety Statistics Annual Report, 2005*. Federal Railroad Administration, Washington, D.C., safetydata.fra.dot.gov/OfficeofSafety/Forms/Default.asp. [Dec. 15, 2006]

Fischer, F.D., Daves, W., Pippan, R., & Pointner, P. (2006). Some comments on surface cracks in rails. *Fatigue & Fracture of Engineering Materials & Structures,* vol. 29, no. 11, pp. 938-948.

Fletcher, D.I., Hyde, P., & Kapoor, A. (2004). Growth of multiple rolling contact fatigue cracks driven by rail bending modeled using a boundary element technique. *Proceedings of the Institution of Mechanical Engineers, Part F - Journal of Rail & Rapid Transit*, vol. 218, no. 3, pp. 243-253.

FRA Office of Safety (2007a). *Download Data on Demand*. safetydata.fra.dot.gov/ officeofsafety/Downloads/Default.asp. [May 10, 2007]

FRA Office of Safety (2007b). *Train Accident Cause Codes*. safetydata.fra.dot.gov/ OfficeofSafety/Downloads/Default.asp. [May 10, 2007]

FRA Office of Safety (2007c). *Train Accident Trends*. safetydata.fra.dot.gov/ OfficeofSafety/Query/Default.asp?page=inctally1.asp. [July 31, 2007]

Hay, W.W. (1982). *Railroad Engineering, Second Edition*. Wiley & Sons, New York.

Hocking, R.R. (1976). A biometrics invited paper: the analysis and selection of variables in linear regression. *Biometrics*, vol. 32, no. 1, pp. 1-49.

Kawprasert, A. & Barkan, C.P.L. (2008). Reducing the risk of rail transport of hazardous materials by route rationalization. *Transportation Research Record.*

Kim, J.K. & Kim, C.S. (2002). Fatigue crack growth behavior of rail steel under mode I and mixed mode loadings. *Materials Science and Engineering A,* vol. 338, no. 1-2, pp. 191-201.

Lei, Z. & Jing-feng, H. (2006). GIS-based logistic regression method for landslide susceptibility mapping in regional scale. *Journal of Zhejiang University - Science A*, vol. 7, no. 12, pp. 2007-2012.

Ling, D.J., Roy, R., Shehab, E., Jaiswal, J., & Stretch, J. (2006). Modelling the cost of railway asset renewal projects using pairwise comparisons. *Proceedings of the Institution of Mechanical Engineers -- Part F -- Journal of Rail & Rapid Transit,* vol. 220, no. 4, pp. 331-346.

Marler, R.T. & Arora, J.S. (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp. 369-395.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall, New York.

Mojsilovic, A., Ray, B., Lawrence, R., & Takriti, S. (2007). A logistic regression framework for information technology outsourcing lifecycle management. *Computers & Operations Research,* vol. 34, no. 12, pp. 3609-3627.

Murray, T. (2008). How much does it cost? *Trains*, vol. 67, no. 1, pp. 34-43.

Odom, M. & Sharda, R. (1990). A neural network model for bankruptcy prediction. *IJCNN International Joint Conference on Neural Networks*, vol. 2, pp. 163-168.

Palese, J.W. & Zarembski, A.M. (2001). BNSF tests risk-based ultrasonic detection. *Railway Track & Structures Magazine*, vol. 97, no. 2, pp. 17–21.

Poole, E.C. (1962). *Costs – A Tool for Railroad Management*. Simmons-Boardman Publishing Corporation, New York.

Saccomanno, F.F. & El-Hage, S. (1989). Minimizing derailments of railcars carrying dangerous commodities through effective marshaling strategies. *Transportation Research Record,* no 1245, pp. 34-51.

Saccomanno, F.F. & El-Hage, S. (1991). Establishing derailment profiles by position for corridor shipments of dangerous goods. *Canadian Journal of Civil Engineering,* vol. 18, no.1, pp. 67-75.

Sagberg, F. (2006). Driver health and crash involvement: a case-control study. *Accident Analysis & Prevention,* vol. 38, no. 1, pp. 28-34.

SAS Institute, Inc. (2006). *SAS 9.1*. sas.com. [October 1, 2006]

Shry, F.Y. & Ben-Akiva, M. (1996). Modeling rail fatigue behavior with multiple hazards. *Journal of Infrastructure Systems,* vol. 2, no. 2, pp. 73-82.

Skyttebol, A., Josefson, B.L., & Ringsberg, J.W. (2005). Fatigue crack growth in a welded rail under the influence of residual stresses. *Engineering Fracture Mechanics,* vol. 72, no. 2, pp. 271-285.

Smith, R.A. (2005). Railway fatigue failures: an overview of a long standing problem. *Materialwissenschaft und Werkstofftechnik*, vol. 36, no. 11, pp. 697 – 705.

Sourget F. & Riollet, A-M. (2006). PROBARAIL: a statistical tool to improve preventive maintenance on rails. *Proceedings of the 7th World Congress on Railway Research*, Montreal, June 2006.

Spackman, K.A. (1992). Combining logistic regression and neural networks to create predictive models. *Proceedings of the Sixteenth Annual Symposium on Computer Applications in Medical Care,* Baltimore, MD, pp. 456-459.

Sperry Rail Service (1999). *Rail Defect Manual*. Sperry Rail Service, Danbury, Conn.

Surface Transportation Board (STB) (2002). *Draft Environmental Impact Statement: Finance Docket no. 34079. San Jacinto Rail Limited and the Burlington Northern and Santa Fe Railway Company Construction and Operation of a Rail Line from the Bayport Loop in Harris County, Texas*. Decision ID No. 33200.

Surface Transportation Board (STB) (2006). *BNSF Railroad 2006 Annual Report.* stb.dot.gov. [February 27, 2008].

Surface Transportation Board (STB) (2008). *Economic Data*. stb.dot.gov/econdata.nsf/. [April 7, 2008].

Towell, G.G. & Shavlik, J.W. (1994). Knowledge-based artificial neural networks. *Artificial Intelligence,* vol. 70, no. 1-2, pp. 119-165.

Transport Canada (2006). *Evaluation of Risk Associated with Stationary Dangerous Goods Railway Cars*. tc.gc.ca/tdc/summary/14600/14690e.htm [May 10, 2007]

Treichel, T.T. & Barkan, C.P.L. (1993). Mainline freight train accident rates. Working paper, unpublished report to the Association of American Railroads.

Tu, J.V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal Clinical Epidemiology,* vol. 49, no. 11, pp. 1225-1231.

Ward Systems Group, Inc. (2006). *NeuroShell Classifier*. wardsystems.com. [December 1, 2006].

Wellington, A. (1887). *The Economic Theory of Railway Location, Second Edition*. John Wiley & Sons, New York.

Yim, J. & Mitchell, H. (2003). A comparison of corporate failure models in Australia: hybrid neural networks, logit models, and discriminant analysis. *Proceedings of the 16th International Conference on Developments in Applied Artificial Intelligence,* Laughborough, UK, pp. 348-358.

Zarembski, A.M. (2005). *The Art and Science of Rail Grinding*. Simmons-Boardman Books, Omaha, Nebraska.

Zarembski, A.M. & Palese, J.P. (2005). Characterization of broken rail risk for freight and passenger railway operations. *Proceedings of the AREMA 2005 Annual Conference*, Chicago, IL.

Zarembski, A.M., Palese, J.W., & Euston, T. (2005). Monitoring grinding effectiveness. *Railway Track & Structures Magazine*, June 2005.

Zhao J., Chan, A.H.C., Roberts, C., & Stirling, A.B. (2006). Assessing the economic life of rail using a stochastic analysis of failures. *Proceedings of the Institution of Mechanical Engineers Part F - Journal of Rail and Rapid Transit,* vol. 220, no. 2, pp. 103-111.

Zhao J., Chan, A.H.C., & Burrow, M.P.N. (2007). Probabilistic model for predicting rail breaks and controlling risk of derailment. *Transportation Research Record,* no 1995, pp. 76-83.

Zumpano, G. & Meo, M. (2006). A new damage detection technique based on wave propagation for rails. *International Journal of Solids and Structures,* vol. 43, no. 5, pp. 1023-1046.

APPENDIX A: SAS OUTPUT FOR PRACTICAL SERVICE FAILURE PREDICTION MODEL

Model Information	
Data Set	WORK.TRIAL_1
Response Variable	BROKEN_RAIL
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	25370
Number of Observations Used	25370

Response Profile		
Ordered Value	**BROKEN_RAIL**	**Total Frequency**
1	1	12685
2	0	12685

Probability modeled is BROKEN_RAIL=1.

Step-wise Selection Procedure

Step 0. Intercept entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

-2 Log L	=	35170.288

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
3595.0046	8	<.0001

Step 1. Effect ULTRASONIC_DEFECT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35172.288	33941.932
SC	35180.429	33958.215
-2 Log L	35170.288	33937.932

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1232.3557	1	<.0001
Score	1142.0348	1	<.0001
Wald	944.4974	1	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
2565.4889	7	<.0001

Note: No effects for the model in Step 1 are removed.

Step 2. Effect RAIL_TYPE entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	35172.288	33323.061
SC	35180.429	33347.485
-2 Log L	35170.288	33317.061

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	1853.2271	2	<.0001
Score	1715.0878	2	<.0001
Wald	1471.8881	2	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
2023.9459	6	<.0001

Note: No effects for the model in Step 2 are removed.

Step 3. Effect ANNUAL_MGT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35172.288	32728.646
SC	35180.429	32761.211
-2 Log L	35170.288	32720.646

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2449.6421	3	<.0001
Score	2273.7011	3	<.0001
Wald	1980.8580	3	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1471.8589	5	<.0001

Note: No effects for the model in Step 3 are removed.

Step 4. Effect AVE_TONS entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35172.288	32241.567
SC	35180.429	32282.273
-2 Log L	35170.288	32231.567

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2938.7214	4	<.0001
Score	2701.5118	4	<.0001
Wald	2313.8615	4	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
1005.8357	4	<.0001

Note: No effects for the model in Step 4 are removed.

Step 5. Effect GEOMETRIC_DEFECT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35172.288	31931.386
SC	35180.429	31980.234
-2 Log L	35170.288	31919.386

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3250.9021	5	<.0001
Score	2973.5499	5	<.0001
Wald	2540.4068	5	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
705.7051	3	<.0001

Note: No effects for the model in Step 5 are removed.

Step 6. Effect RAIL_WGT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	35172.288	31595.359
SC	35180.429	31652.348
-2 Log L	35170.288	31581.359

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	3588.9288	6	<.0001
Score	3268.5913	6	<.0001
Wald	2777.7570	6	<.0001

Residual Chi-Square Test		
Chi-Square	DF	Pr > ChiSq
370.3985	2	<.0001

Note: No effects for the model in Step 6 are removed.

Step 7. Effect BRD_PRESENT entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	35172.288	31328.628
SC	35180.429	31393.759
-2 Log L	35170.288	31312.628

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	3857.6599	7	<.0001
Score	3492.4723	7	<.0001
Wald	2945.4859	7	<.0001

Residual Chi-Square Test		
Chi-Square	**DF**	**Pr > ChiSq**
118.3244	1	<.0001

Note: No effects for the model in Step 7 are removed.

Step 8. Effect AGE_OF_RAIL entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	35172.288	31211.125
SC	35180.429	31284.397
-2 Log L	35170.288	31193.125

Testing Global Null Hypothesis: BETA=0			
Test	**Chi-Square**	**DF**	**Pr > ChiSq**
Likelihood Ratio	3977.1626	8	<.0001
Score	3595.0046	8	<.0001
Wald	3025.2001	8	<.0001

Note: No effects for the model in Step 8 are removed.

Note: All effects have been entered into the model.

Summary of Step-wise Selection							
	Effect						
Step	**Entered**	**Removed**	**DF**	**Number In**	**Score Chi-Square**	**Wald Chi-Square**	**Pr > ChiSq**
1	**ULTRASONIC_DEFECT**		1	1	1142.0348		<.0001
2	**RAIL_TYPE**		1	2	605.0523		<.0001
3	**ANNUAL_MGT**		1	3	586.6958		<.0001
4	**AVE_TONS**		1	4	474.1906		<.0001
5	**GEOMETRIC_DEFECT**		1	5	308.3612		<.0001
6	**RAIL_WGT**		1	6	337.4356		<.0001
7	**BRD_PRESENT**		1	7	253.0159		<.0001
8	**AGE_OF_RAIL**		1	8	118.3244		<.0001

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.9373	0.3219	235.3116	<.0001
RAIL_WGT	1	-0.0454	0.00218	435.8151	<.0001
RAIL_TYPE	1	-1.3464	0.0551	598.0631	<.0001
AGE_OF_RAIL	1	-0.0106	0.000978	117.4865	<.0001
ANNUAL_MGT	1	0.00899	0.000356	638.1390	<.0001
AVE_TONS	1	0.0232	0.00133	306.7840	<.0001
ULTRASONIC_DEFECT	1	1.6130	0.0571	796.7727	<.0001
GEOMETRIC_DEFECT	1	0.8227	0.0455	327.0920	<.0001
BRD_PRESENT	1	1.6294	0.1119	212.0442	<.0001

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
RAIL_WGT	0.956	0.952	0.960
RAIL_TYPE	0.260	0.234	0.290
AGE_OF_RAIL	0.989	0.988	0.991
ANNUAL_MGT	1.009	1.008	1.010
AVE_TONS	1.023	1.021	1.026
ULTRASONIC_DEFECT	5.018	4.486	5.613
GEOMETRIC_DEFECT	2.277	2.083	2.489
BRD_PRESENT	5.101	4.096	6.351

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	71.7	**Somers' D**	0.437
Percent Discordant	28.0	**Gamma**	0.439
Percent Tied	0.3	**Tau-a**	0.219
Pairs	160909225	**c**	0.719

Wald Confidence Interval for Adjusted Odds Ratios				
Effect	**Unit**	**Estimate**	**95% Confidence Limits**	
RAIL_WGT	1.0000	0.956	0.952	0.960
RAIL_TYPE	1.0000	0.260	0.234	0.290
AGE_OF_RAIL	1.0000	0.989	0.988	0.991
ANNUAL_MGT	1.0000	1.009	1.008	1.010
AVE_TONS	1.0000	1.023	1.021	1.026
ULTRASONIC_DEFECT	1.0000	5.018	4.486	5.613
GEOMETRIC_DEFECT	1.0000	2.277	2.083	2.489
BRD_PRESENT	1.0000	5.101	4.096	6.351

Partition for the Hosmer and Lemeshow Test					
		BROKEN_RAIL = 1		**BROKEN_RAIL = 0**	
Group	**Total**	**Observed**	**Expected**	**Observed**	**Expected**
1	2537	502	611.85	2035	1925.15
2	2537	750	814.76	1787	1722.24
3	2537	832	910.77	1705	1626.23
4	2542	1065	995.19	1477	1546.81
5	2540	1364	1086.92	1176	1453.08
6	2537	1249	1202.03	1288	1334.97
7	2537	1273	1353.28	1264	1183.72
8	2543	1591	1624.56	952	918.44

Partition for the Hosmer and Lemeshow Test					
		BROKEN_RAIL = 1		BROKEN_RAIL = 0	
Group	Total	Observed	Expected	Observed	Expected
9	2537	1900	1888.49	637	648.51
10	2523	2159	2196.91	364	326.09

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
196.6711	8	<.0001

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.000	12685	0	12685	0	50.0	100.0	0.0	50.0	.
0.100	12685	31	12654	0	50.1	100.0	0.2	49.9	0.0
0.200	12601	448	12237	84	51.4	99.3	3.5	49.3	15.8
0.300	12105	2179	10506	580	56.3	95.4	17.2	46.5	21.0
0.400	9828	6628	6057	2857	64.9	77.5	52.3	38.1	30.1
0.500	6979	9428	3257	5706	64.7	55.0	74.3	31.8	37.7
0.600	5258	11087	1598	7427	64.4	41.5	87.4	23.3	40.1
0.700	4227	11540	1145	8458	62.1	33.3	91.0	21.3	42.3
0.800	2076	12334	351	10609	56.8	16.4	97.2	14.5	46.2
0.900	674	12604	81	12011	52.3	5.3	99.4	10.7	48.8
1.000	0	12685	0	12685	50.0	0.0	100.0	.	50.0