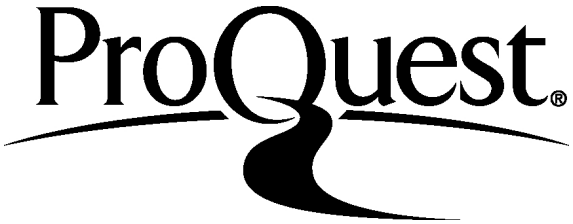
THE DEVELOPMENT OF RAIL DEFECTS DUE TO THE PRESENCE OF GEOMETRY DEFECTS IN CLASS 1 RAILROADS

by

Daniel Einbinder

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Civil Engineering

Summer 2015



THE DEVELOPMENT OF RAIL DEFECTS

DUE TO THE PRESENCE

OF GEOMETRY DEFECTS

IN CLASS 1 RAILROADS

by

Daniel Einbinder

Approved: _____

Allan Zarembski, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____

Harry W. Shenton III, Ph.D.
Chair of the Department of Civil and Environmental Engineering

Approved: _____

Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____

James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Zarembski, for all the support he has given me over the past couple of years. If not for his extensive knowledge in this subject, his constant push to get work done, his patience and overall guidance I would not be where I am today. I am extremely glad and lucky that he introduced me to the railroad world. I could not have asked for a better person to do this.

Second, I would like to thank Prof. Attoh-Okine for teaching many new and interesting ways on how to use statistics. I am grateful for the encouragement he provided and that he was always able to pick up my spirits.

I would also like the thank Dr. Riddell. Dr. Riddell is who first truly introduced me to doing research work. Without him, I would have never pursued an advanced degree.

Last but not least, I would like to thank my friends and family for always being there for me when I needed a break and for nagging me to get work done when I did not want to.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

The purpose of this thesis is to develop a relationship between the development of a rail defect and the presence of a geometry defect prior to the occurrence of the rail defect. This was done with the use of a Class 1 Railroad's geometry and rail defect data for their whole 22,000 mile system over a five year period. Regression analyses were performed using multilinear regression and multivariate adaptive regression splines (MARS) that showed that the occurrence of a geometry exception or defect resulted in approximately 30% loss in rail life. A Bayesian Network was then developed with the use of NETICA. This network showed that the presence of a single geometry defect increased the likelihood of a rail defect by approximately seven, with multiple geometry defects increasing this likelihood even more. These findings show that geometry defects do have an impact on the development of a rail defect.

Chapter 1

INTRODUCTION

The goal of this project was to determine and then develop engineering and/or statistical relationships between different types of geometry exceptions and the development of rail defects and failures. A detailed look at the relationships used can be seen in the BACKGROUND AND LITERATURE REVIEW chapter. The data used to determine these relationships was provided by a major US Class 1 railroad (CSX) to include rail and geometry defect data for a five year period. This data was correlated between the geometry exceptions and rail defects. The project focused primarily on developing statistical models with engineering reasoning behind their development. The project was broken into four different tasks.

The first task was the collection, consolidation, and processing of the data. The data collected included the following:

- 2.5 years of CSX rail defect data
- 5 years of CSX rail defect data
- 5 years of annual CSX tonnage data
- 5 years of Vertical Track Interaction data
- CSX track curvature data

This data was consolidated and processed using a combination of Excel and MATLab. More information of data collection and processing can be found in the DISCUSSION OF DATA chapter.

The second task was to determine the key parameters to be considered in the analysis. The parameters were determined by performing different data correlations and with the use of different regression models. This can be seen in more detail in the MULTILINEAR AND MARS ANALYSES section.

The third task was to develop statistical relationships and models using the constructed database and the determined key parameters. These models can be seen in the PROBABILITY ANALYSIS section.

The forth task was using two separate data sets to validate the relationships and models developed in the previous task. The validation can be seen in the VALIDATION OF MODELS section.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

Rail Stresses

The relationship between rail stresses and the track support structure has been known since Winkler's [1] Beam on Elastic Foundation Theory in 1867. This theory over the years has been further refined by Hetenyi[2], Hay[3], and Kerr[4]. The relationship between rail stress, contact stresses, and thermal stresses has also been explored by Zarembski [5] and Steele [6].

The stress within the rail is complex and can involve several different load cases, such as thermal stress, contact stress, and bending stress. An illustration of these different stresses within the rail can be seen in Figure 1. Contact stress is the most important stress in terms of maintenance, since it can be most directly affected by maintenance decisions.

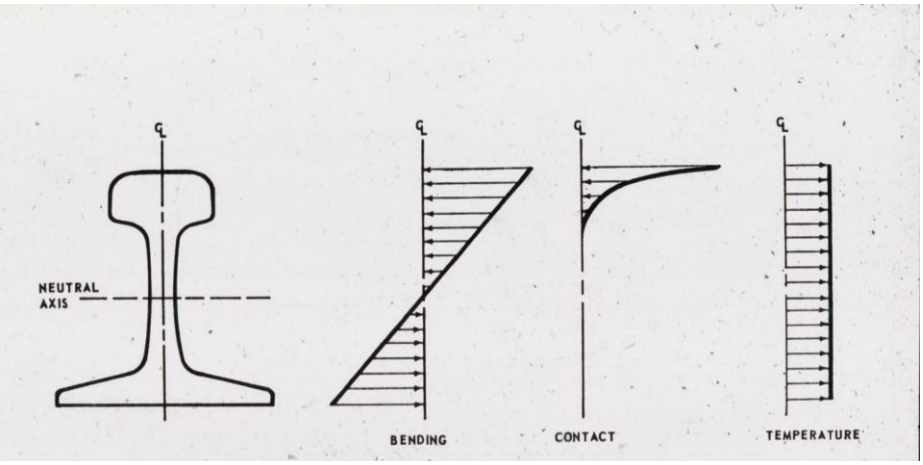


Figure 1 Different Stresses in a Rail

Geometry Defect Effects on Dynamic Loading

There are several different analyses that show the relationship between the presence of a surface defect, either geometry or rail, and the increase of dynamic loading on the rail. A particular analysis of interest performed by Jenkins et al. [7] focuses on the dynamic impact force created by a moving train on a track with a vertical geometry deviation. These dynamic impact forces can be split into two peaks which are called P1 and P2. Figure 2 shows the location of these values and how impactful they are. The P1 force, shown in the figure, has a larger amplitude with a shorter duration, where the P2 force has a lower amplitude with a longer duration. Since P2 duration is longer with the heightened dynamic impact load, it tends to cause more issues with track degradation. (P1 generates higher rail stresses)

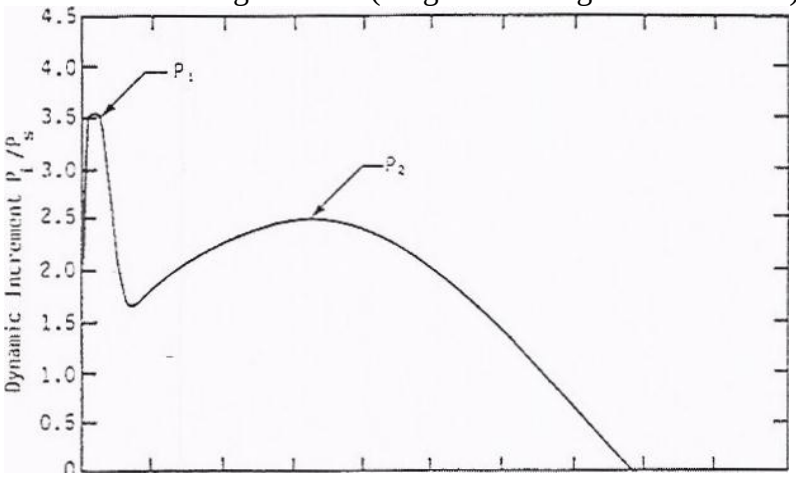


Figure 2 P1 and P2 Force Diagram

P1 and P2 forces are defined in Jekins et al. [7] and Ahlbeck [8] a as follows in Equations 1 and 2.

$$P_1 = P_0 + 2\alpha V \sqrt{\frac{K_h M'}{1 + \frac{M'}{M_u}}} \qquad \text{Equation 1}$$

$$P_2 = P_0 + \left[1 - \pi\xi^2 \frac{M_t}{M_u + M_t}\right]\left[2\alpha V\right]\sqrt{K_t M_u}\left[\frac{M_u}{M_u + M_t}\right]^{\frac{1}{2}} \qquad \text{Equation 2}$$

Where the terms are defined as:

P_0 = Static Wheel Load, (lbs)
a= Rail Joint Dip Angle, (rad)
V= Vehicle Speed, (in/sec)
K_h = Hertzian Contact Stiffness, (lb/in)
M' = Effective Mass of Rail and Tie, (lb-sec^2/in)
M_u = Unsprung Vehicle Mass, (lb-sec^2/in)
ξ = Track Effective Damping Ratio
K_t = Effective Track Stiffness, (1b/in)

To determine these input parameters and for a more detailed description of the parameters refer to Jekins et al. [7] and Ahlbeck [8].

Wheel/Rail Interaction

As discussed earlier the contact stresses (wheel/rail contact) have an effect on the overall stress in the rail. These stresses have a larger impact on the development of detail fracture defects (TDD) in the surface and gage corner of the rail [9]. The values of these stresses can be determined by using Hertzian Contact Stress Theory [10]. Figure 3 below shows a more detailed look at the contact stress developed by the wheel/rail interaction.

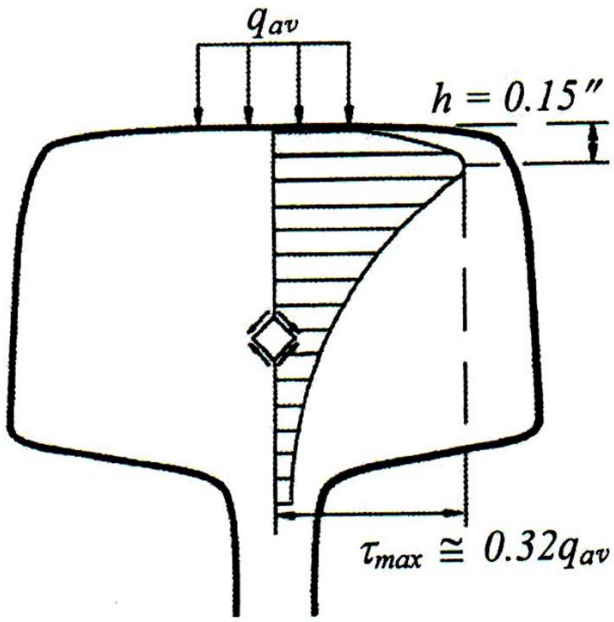


Figure 3 Stresses induced by wheel/rail contact

A photo-elastic look at the wheel/rail interaction, Figure 4, shows that the contact points of the wheel and rail is important when determining the magnitude of these stresses. This means the rail cant, and to some extent gage, directly affects the magnitude of the contact stresses being developed. This means changes in rail cant, due to the presence of a gage or other geometry defect, can impact the development of rail defects in the gage corners. This can be further seen in how rail grinding can reduce these defects by correcting or redefining the wheel/rail contact [9]

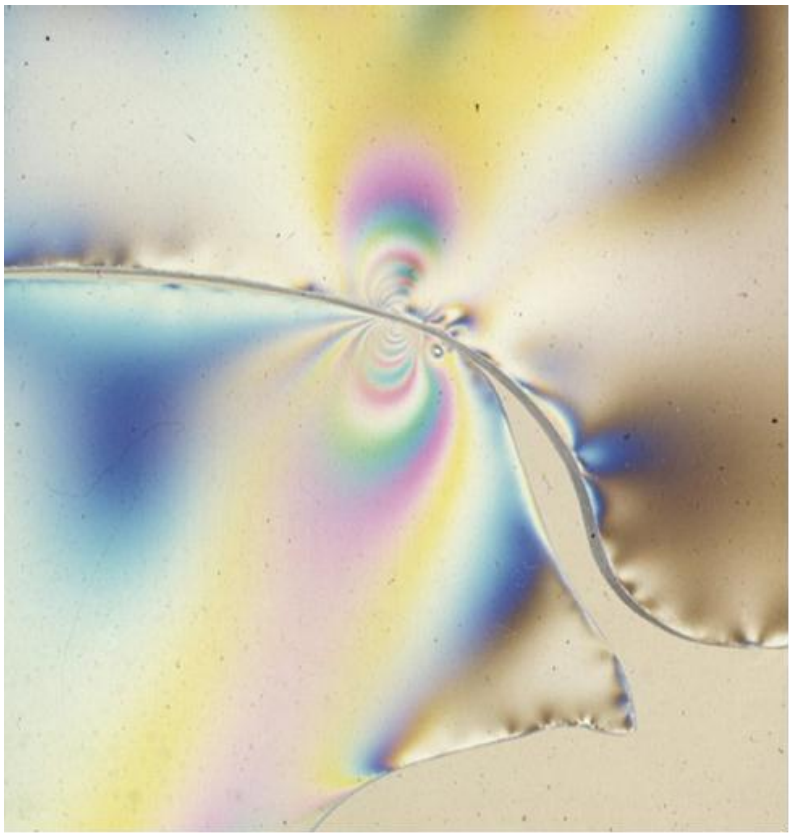


Figure 4 Photo-Elastic Wheel/Rail Contact Stress at Gauge Corner of Rail

Multivariate Adaptive Regressions Splines

Multivariate adaptive regressions splines (MARS) is a data driven analysis. The MARS analysis is based on developing a relationship between predictor variables, vector {X}, and a dependent variable, variable Y. This can be seen below in Equation 3, where ε is the deviation of the dependent variable.

$$Y = f\{X\} + \varepsilon \qquad \text{Equation 3}$$

MARS uses a set of smooth continuous splines throughout the vector {X}. This shows the shift in the relationship of the variables. The locations of these shifts are known as “knots”. These knot locations are determined through a forward pass that

reduces the sum of squared residuals. Once these knots are located and formed, a series of functions are created. These functions are called basis functions. The basis functions are shown with one of two forms shown below:

max(0,x-t) or max(0,t-x)

With these functions formed, a backwards pass is performed. This backward pass deletes unimportant knots and basis functions, giving the final form of the MARS analysis. The knot selection and deletion can be seen in more detail in Friedman [11], Sephton [12], and Attoh-Okine [13]

Bayes' Theorem

Bayes' Theorem is a theorem used to determine the conditional probability of an event to occur based off a different event occurring. It was first published by Thomas Bayes in 1763. Conditional probability is typically determined by using joint probability and the random probability of an event. This can be seen in Equation 2.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{Equation 4}$$

By using the multiplication rule and the definition of marginal probability, this conditional probability can be rewritten. This new form, shown below in Equation 3, is called Bayes' Theorem [14].

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} \quad \text{Equation 5}$$

Bayes' Theorem can also be shown as:

$$Posterior = \frac{Likelihood\ X\ Prior}{Evidence} \quad \text{Theorem 1}$$

Where:

$Posterior = P(B|A)$

$Likelihood = P(A|B)$

$Prior = P(B)$

$Evidence = P(A)$

Bayes' theorem is the base of all Bayesian statistics and is used in both Naïve Bayes and Bayesian Network methods.

Naïve Bayes

Naïve Bayes, a Bayesian Classifier, is a very commonly used method to determine the likelihood of an event to occur based off the occurrence of different types of event occurring. This method is typically seen in spam filters for emails. Using Naïve Bayes, the Bayes' Theorem is rewritten, show in Equation 4.

$$P(B|A) = \frac{P(B)*\sum_i^n P(A_i|B)}{\{P(B)*\prod_i^n P(A_i|B)\}+\{P(\neg B)*\prod_i^n P(A_i|\neg B)\}} \quad \text{Equation 6}$$

Using this method introduces an independence assumption into the interaction between the different events, as in the independent variable events (A_i) do not have an effect on the occurrence on other independent variable events. In the example of determining if email is spam or not, the occurrence of the word "Sir" does not increase the chance that the word "buy" would occur. If both words were to occur though, the probability, or likelihood, of the email being spam may increase. More can be seen on the uses of Naïve Bayes in Lewis [15] and Androutsopoulos et al. [16].

Bayesian Network

Bayesian Networks are a probabilistic model where variables are shown with nodes and their relationships are shown with directed edges. Each nodes' value is determined by a conditional probability table that is attached to that variable and expresses the probability of an event (variable) occurring. The nodes are broken into two sets of variables. Parent variables are where the link (direct edge) starts from.

Child variables are where the link ends. The child variables probability is thus directly conditioned by the parents' probability. Due to these links, the independence assumptions made while using Naïve Bayes are minimized and controlled, giving a much better understanding of the relationships between all the different variables. Below in Figure 5 is an example of what a Bayesian Network looks like. These networks are typical created with the use of software such as NETICA. The use of NETICA is explained later with the data in the RESULTS AND DISCUSION chapter. A more detailed look of how Bayesian networks are formed can be seen in Catselletti [17] and Bielza [18].

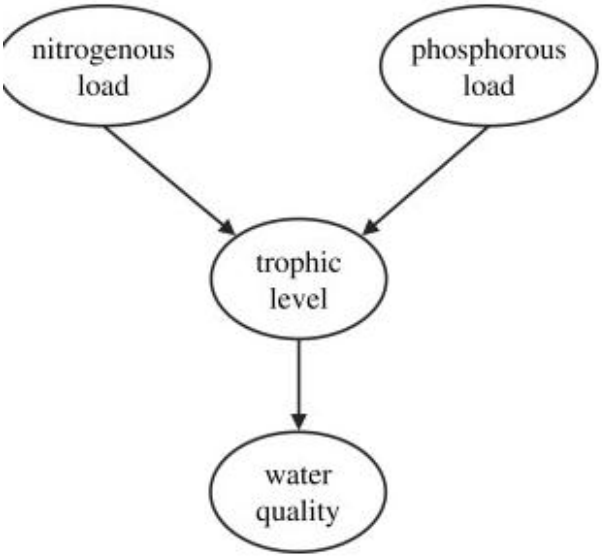


Figure 5 Example Bayesian Network

Chapter 3

DISCUSSION OF DATA

As stated in the INTRODUCTION chapter, a large amount of data was collected and provided for this project. The majority of this data came from a major Class 1 Railroad (CSX) as well as from the FRA. The data that was received is listed below:

- 2.5 Years of CSX Rail Defect Data (Shown in Appendix A1)
- 5 Years of CSX Geometry Exception Data (Shown in Appendix A2)
- 5 Years of CSX Annual Tonnage Data (Shown in Appendix A3)
- 5 Years of CSX Vertical Track Interaction (Shown in Appendix A4)
- 5 Years of FRA Track Geometry Data (Shown in Appendix A5)

The CSX data contained approximately 22,000 track miles. CSX has 11 Divisions that were being examined. These divisions are as follows; Albany, Appalachian, Atlanta, Baltimore, C&O, Chicago, Florence, Great Lakes, Jacksonville, Louisville, and Nashville. Each division had a subset of subdivisions. The average annual traffic on the rail system, as defined in terms of Millions of Gross Tons of traffic or MGT was determined as 21.3 MGT.

Preprocessing of Data

The rail data provided was that of the whole system, approximately 22,000 miles. A sample of the raw format of this data can be seen in Appendix A1. Below is a smaller sample in Figure 6. There was approximately 2.5 years' worth of rail data with a total of 50,000 rail defects. Each rail defect entry had a defect code which

represented which type of rail defect it is. Below in Figure 7 is a table showing what each code relates to in terms of rail defects.

DIVISION	SUBDIVISION	PREFIX	MILEPOST	TRACK TYPE	TRACK CODE	SIDE	DEFECT TYPE	SIZE	DATE FOUND	CURVE - TANG	ROLLED YEAR	MILL	WEIGHT
HU	**JAMES RIVER**	**CAB**	**185.7**	**SG**	**M**	R	TDD	10	09/03/2010	T	1988	TH	136

Figure 6 Example of rail defect data

DESCRIPTION	RAIL TEST CAR CODES	CSX DEFECT CODES
Transverse Fissure	TDT	TDT
Compound Fissure	TDC	TDC
Detail Fracture	TDD	TDD
Engine Burn Fracture	EBF	EBF
Defective Plant Weld	DWE/DWG	EFBW/OAW
In-Track Electric Flash Butt Weld	DWE	EFBW
Defective Field Weld	DWF	TW
Horizontal Split Head	HSH	HSH
Vertical Split Head	VSH/VSJ	VSH/VSH
Split Web	SWO/SWJ	SW/SW
Piped Rail	PRO/PRJ	PIPE/PIPE
Head Web Separation	HWO/HWJ	HW/HW
Bolt Hole Crack	BHO/BHJ	BHB/BHB
Broken Base	BBO/BBJ	BB/BB
Flattened Head (Flattened out across the width of the rail head, occurrences have no repetitive regularity, does not include corrugations and has no apparent localized cause such as a weld or engine burn.)	FH	FH

LER (Loss of Expected Response is an area with contamination that causes a loss of ultrasonic signal)	LER	LER
SSC Length of Track (due to Shells, Spalling, or Corrugation conditions that causes a loss of ultrasonic signal.)	SSC	SSC
Crushed Head (Flattening and widening of the head for several inches, with entire head sagging, creep or bleeding under head and small cracks in the depression on the running surface.)	CH	CH
Engine Burn – (Not fractured) Shelly Spots	EB SD	EB SD
Shelly Spots in Dead Zone of Switch	SDZ	SDZ

Figure 7 Description of rail defect codes

Using the figure above, certain rail defect codes and entries were determined to be not of interest. These defects include; engine burn (EB or EBF), loss of expected response (LER or SSC), and shelly spots in dead zone of switch (SDZ). Once these rail defects were removed, the number of rail defects of interest came to approximately 26,000 rail defects. These remaining rail defects were separated into one of the 11 divisions for easier processing later on.

A similar process of splitting the data into divisions was done for the geometry data, the tonnage data, and the VTI data. Samples of all these can be seen in Appendix A. It should be noted that the VTI data of interest was the measured dynamic impact forces, labeled as AXV1 and AXV2. This data represents the measured impact force, in pounds, peaked over a 1sec period of time, with a threshold of 100 Kips. Since

these values were the only values of interest, less than 20% of the original 1,000,000 records were used.

Development of Database

Once the data was preprocessed, the database was ready to be formed. In order to provide for initial data base construction and early assessment of data, the initial data entered into the database focused on CSX track geometry and rail defect data, together with tonnage data and location reference information. The initial set of data provided by CSX was for the Baltimore Division, which has approximately 1,040 miles of track with an average annual tonnage of 24 MGT. The Baltimore Division consisted of 18 subdivisions ranging in size from less than 10 miles to 315 miles and with annual tonnage of less than 1 MGT to 42 MGT. The Baltimore Division had approximately 18,351 reported track geometry defects (over the five year period) and 1,543 rail defects (over the two and a half year period noted above). Of the 18,351 geometry defects, 11,462 were in unique locations (based on a 50 foot location resolution length). The remainder was repeat geometry defects at the same location, but with either a different date or multiple types of defects.

For the Baltimore division, the database construction was done manually in Excel. The first step was to define the location of each defect, with the initial focus on the rail and geometry defects provided by CSX from its own database (for the Baltimore Division). The location for a rail defect was reported in two formats; mile post (which includes Division, Subdivision, milepost, footage and track number) and longitude/latitude coordinates. Geometry defects were reported with the same two formats, however since many of the geometry defects were of measurable length; these were also reported with the addition of start and ending locations of the reported

defect. Though longitude and latitude provide an accurate location, a large portion of rail defects did not have these coordinates, therefore milepost location was used. It was decided that for this analysis, the precision of location was set to be +/- 25 foot for a 50 foot defect length (approximately 0.01 miles). This was used for the matching of defects.

Once each defect was located, the database was structured to determine what rail and geometry defects "matched" each other. In order to be considered matched, several conditions must be met.

The first condition is location of both rail and geometry defects. The location matching of the defects was based on the matching of several parameters. The division was the first parameter to be looked at. Since this first set of data was all from the same division this was easily handled. The second parameter was the subdivision. The subdivision location was defined for each of the provided defect data points. For the Baltimore Division, all of the defect data was separated by subdivision. A total of 18 subdivisions were accounted for during this process. Once the defect data was separated by subdivision, it was then sorted by mile post location. The location precision of 0.01 miles (approximately 50 feet) was used to determine if a geometry defect influenced a rail defect. If these conditions were met then this constituted that the geometry defect is "near" the rail defect.

Noting that geometry defects can have more than one defect at one specific location, this was noted with all matches recorded and filtered later on in the process.

After the location condition is met, the time condition is analyzed, which in the case of both rail and geometry defects are the dates of inspection. In order for a

geometry defect to influence a rail defect and be considered a match, it must occur prior to the rail defect.

Once these conditions are met, the raw geometry defect and raw rail defect data are combined into one file containing all the matches for the division. After all matches are found, rail defects with more than one geometry defect must be filtered down into one unique geometry defect per unique rail defect. The order of importance for selecting a geometry defect to be used is as follows; priority critical, most recent defect, cant geometry defect, crosslevel geometry defect, and lastly profile geometry defect. A new sheet is then created for unique locations of matches. In addition, a record is kept of all "multiple" matches, i.e. rail defects that match with 2 or more track geometry defects at the same location, with the geometry defect dates always occurring prior the rail defect date.

The final step before analyzing the data is determining the MGT that each track experiences. Though several years of data for MGT was given it was determined to use only 2012's data due to inconsistency between the formats of each year and 2012 having the best and most consistent overall format. The MGT for each track was determined using a weighted average, with the weighting based on segment length and MGT on each segment. The equation used for weighted MGT can be seen below in Equation 7.

$$MGT\ (weighted)\ =$$
$$\Sigma\ Segment\ length * MGT\ (segment) / \Sigma\ Segment\ lengths \qquad \text{Equation 7}$$

Using the hand developed data base for the Baltimore Division, a set of preliminary analyses were performed to allow for the development of data extraction techniques and the generation of Data Tables. This preliminary data analysis began

with a detailed breakdown of each division into a table format using a Microsoft Excel structure. Each subdivision was placed in this breakdown, even if it did not contain any matches. After the subdivision name is the total track miles that subdivision contains as well as its average MGT for all of its tracks. The track length and average MGT is then further broken down into each track number for that subdivision. Placed next into this table is the total number of geometry and rail defects that occur in a subdivision, followed by a ratio of geometry defects to rail defects. To determine the number of defects a subdivision had, the original defect data was sorted by subdivision with the use of a filter tool. The original data was sorted by mile post, track number, and subdivision. Sorting in this fashion made it easy to count the number of defects that occurred within a subdivision. The number of defects in a subdivision was counted by selecting all the defects within the subdivision and then counting the number of defects that occurred. This value was recorded into the table. The unique matched defects were then sorted and counted in the same fashion as the raw data and placed in the table.

One rail defect, detail fraction (TDD), was of particular interest, since it represented over 45% of all the rail defects in the Baltimore and was the single largest defect category by far. To determine the number of TDD defects that occurred in a subdivision, another layer to the filter was applied. The unique matched and raw rail defect data was sorted by defect type. After being sorted, the number of defects was counted in the same manner as previously stated. The number of TDDs was also added onto this table. The number of matches of all rail defects and TDD defects that were located on a curve was also recorded in the table. For this analysis, the raw rail defect data and the unique match data was sorted by whether it was located on a curve or on

a tangent[1]. After being sorted this way, it was re-sorted by milepost, track number, and subdivision. This separated the curve and tangent locations apart from each other. The defects that were on a curve were then selected and counted. The last piece of data place on this table was the percent of the rail defects and TDD defects that were matched. A compressed example of this chart can be seen below in Figure 8.

Sub Division	Length in Miles	Subdivision MGT	Geometry Defects	Rail Defects	Ratio	Matches	# of TDD Match	Repeat	% of Rail Defects Matches	% of Match with Curves	% of TDD on Curve	% of Matched TDD
BA	314.89	32.96	7271	520	14.0	86	60	21	17%	74%	50%	20%

Figure 8 Example of Baltimore Division Summary Master Chart

After this master chart for a division is created, it was noted that there were a large number of subdivisions and segments with low tonnage levels that would not generally generate a large number of rail defects. Using a threshold of 20 MGT (the actual threshold was set at 19.5 MGT), a MGT filter was applied to the Baltimore Division data. This filter was applied based on MGT of each track segment within a subdivision. Any track segment with a MGT of less than 19.5 MGT was removed from the chart along with all of its geometry and rail defects. The filtered summary chart, for Baltimore Division is presented in Table XX. Note, while the number of subdivisions is reduced from 18 to 7 and the total track miles is reduced from 1,043 to 798, the number of rail defects is only reduced from 1543 to 1330, and geometry

---

[1] Tangent refers to a section of straight track with no curvature.

[2] Also the majority of all defects

[3] Rail roll date is the date of manufacture of the rail which is usually close to the

defects from 18,351 to 16,207. Furthermore the number of matched defects is only reduced from 204 to 182.

An additional data sort and comparison is made for comparison of defect type; specifically rail defect type vs. track geometry defect type. This is illustrated in Appendix B which shows rail defect type across the top vs geometry defect type along the vertical axis. A compressed example of this can be seen below in Figure 9, where TDD is detail fracture from shell and BRO is a broken rail outside joint area.

DEFECTS	TDD	BRO
CROSSLEVEL	14	1
WARP	25	0
PROFILE	5	1
RAIL CANT	42	0
GAGE	18	0
TOTAL	104	2

Figure 9 Example of Baltimore Division Defect Matrix

After matching and initially analyzing the Baltimore Division by hand, a MATLab code was developed to perform the matching process automatically. The code starts by pulling all of the raw data out of the original Excel data files and placing them into two cell arrays, one for rail defects and one for geometry defects. Once the data is in the cell arrays, the code moves to nested for loop in order to determine which two specific defects to compare. The first for loop is rail defects and the second for loop is geometry defects. Once in the nested for loop, the code determines if the defects are a match following the same procedure as the hand method in nested if statements. If a rail defect is determined to not have a match, it is moved to a separate

cell array and the code moves to the next rail defect. If a rail defect is determined to have a match, or multiple repeat matches, both the rail and geometry defects are moved into a new cell array. Once all rail defects were looked at, the cell arrays were transported back into the data format. Divisions were no longer broken into subdivisions prior to matching. This code allowed for quicker construction of the initial data base and easy manipulation of the data. A code was also developed for filtering out the multiple matches of geometry defects to rail defects. The code would start by converting the matched defect data file into a cell array. A nested for loop is used again to compare different matches. This comparison is done with if statements using the same method and order of importance as before to determine repeats and unique locations. The code also checks to make sure the rail defect being compared is the same. Once a defect is done being compared, the unique location is pulled out. A similar code to the filtering unique matches code was developed for filter just geometry defects. The code setup is near identical with the exception that it does not check rail defect type and the total number of geometry defects that are repeats are counted. This was done to find the total unique locations that had a geometry defect. To further quicken the data processing, the MGT of each track was added to the initial rail defect data. This allows for easier MGT filtering during the master chart filtering process. Skeleton versions of these codes can be found in Appendix C.

Once every division was matched and a database was formed, new summary master charts were created for each subdivision. These charts were formatted in the same fashion as the Baltimore division. These master charts were then combined into a final master chart, which replaced subdivisions with divisions. This made comparing the different divisions easier. This chart followed the same format as the Baltimore

division chart with one exception; the track segments (SG track, Track 1, Track 2, and etc.) were not included. This master chart had the divisions in both a MGT filtered and non-filtered format. This format was used for when comparing individual division. A second master chart format was made that only included the combined values of the data. This master chart had the same information as the previous chart but with three changes. The first is that the data was broken into tangent, curve, and all track categories. The second change was the orientation in which the chart was presented. The last change was that this format also included more detailed information about the key geometry defects that are discussed later. Both of these charts can be seen in Appendix D.

The VTI impact data was also matched with geometry and rail defect data to form another data base. This data was matched in the same manner as the geometry defect data, i.e. the VTI impact came prior to the rail defect. When the whole system was matched using the developed MATLab code, only three matches were found. It was determined from this low match amount that the VTI data was not important in these relationships.

Chapter 4

RESULTS AND DISCUSSION

This chapter discusses the relationships that were developed using the constructed databases. These relationships are based primarily off the statistical methods discussed earlier.

Defect Correlation

The initial analysis preformed on the data was a system wide correlation analysis between geometry defects and rail defects. After the initial formation of the database, it was observed that there was a large amount of rail defects with multiple geometry defects prior to it. These were called “repeats” and were looked at separately from the other matches. Figure 10 below has a breakdown of repeat matches for each division as well as each division with a 20 MGT filter. In the figure, it is shown that a large portion (38%) of the matches are made up of repeat geometry defect matches. This effect does not change much on the filtered track. As a result of these high amounts of repeat matches, 4.2% of all rail defects have multiple geometry defects prior to it occurring. These values vary depending on which division is being observed, though they tend to stay close to these values. This shows that multiple geometry defects have some impact on the development of rail defects. This effect is explored more in the next two chapters.

Figure 11 is a summary of the total values for both the filtered and non-filtered data. This figure has the match values, repeat match values, as well as the percent

make up of these values of the rail defect population. Figure 12 presents the same data for curve segments only. Looking at the curve specific data only, it was observed that that one fifth of all rail defects on curves had multiple geometry defects that occurred prior to the rail defect. This suggests that there is a relationship of some sort between multiple geometry defects and rail defects on curves.

Divisions	Length in Miles	# of Repeats	% of Repeats	% Repeat Matches All
Baltimore Full	1042.79	31	15.20%	2.01%
Baltimore > 19.5	798.4	30	16.48%	2.26%
Atlanta Full	2042.1	40	26.67%	1.75%
Atlanta > 19.5	1405.9	33	27.50%	1.89%
Albany	1871.05	50	31.85%	2.82%
Albany >19.5	840.46	21	26.92%	2.25%
Appalachian Full	2221.24	343	44.95%	7.17%
Appalachian >19.5	952.44	191	41.08%	8.13%
C&O Full	2051.42	234	47.18%	9.86%
C&O >19.5	770.04	168	54.90%	13.50%
Chicago Full	1630.73	53	41.41%	3.62%
Chicago > 19.5	400.8	19	37.25%	3.11%
Florence Full	3056.51	106	37.19%	2.64%
Florence > 19.5	1125.85	61	38.61%	3.32%
Great Lakes Full	2326.16	61	31.28%	2.42%
Great Lakes > 19.5	1840.21	58	32.77%	2.52%
Jacksonville Full	2920.07	64	35.96%	2.82%
Jacksonville > 19.5	784.37	32	42.67%	4.33%
Louisville Full	1406.79	65	41.67%	3.92%
Louisville > 19.5	559.94	45	45.00%	4.58%
Nashville Full	1658.86	72	34.95%	4.12%
Nashville > 19.5	1202.42	37	34.26%	2.74%
Total Full	**22227.7**	**1119**	**38.35%**	**4.23%**
Total Filter	**10680.8**	**695**	**38.19%**	**4.50%**

Figure 10 Summary of Repeat Matches

	Length in Miles	Annual MGT	Reported Geo Defects	Unique Geo Defects	Rail Defects	Matches	**% of All Matched**	# of Repeats	% of Repeats	**% Repeat Matches All**
Total Full	22227.7	21.33	334937	202341	26440	2918	**11.04%**	1119	38.35%	**4.23%**
Total Filter	10680.8	35.94	173314	104952	15428	1820	**11.80%**	695	38.19%	**4.50%**

Figure 11 Summary of Matched and Repeat Matched Rail Defects

	Rail Defects	Rail Defects on Curves	Matches	Matched on Curves	**% of All Matched**	% of Repeat Matches on Curves	**% of Repeat Matches With Curves**
Total Full	26440	8870	2918	1871	**21.09%**	45.70%	**9.60%**
Total Filtered	15428	5225	1820	1113	**21.30%**	43.10%	**9.20%**

Figure 12 Summary of Matched and Repeat Matched Rail Defects on Curves

As part of the correlation analysis, the matched rail defects and geometry defects for the whole CSX system were correlated by the specific defect type, shown in full in Figure 13. As can be seen in this table, detail fracture defects (TDD) make up the majority of matched rail defects[2] at 44% of the matched. Rail Cant and Warp defects make up the majority of the geometry defects. As shown in Figure 13, many of the geometry defect consist of geometry defects of the same type with slight

[2] Also the majority of all defects

differences, such as rail cant left and rail cant right. There are also several rail defects that made up a small amount of the matched rail defects (1% of the matched defects). Taking these two observations in mind, Figure 13 was consolidated into Figure 14. This consolidation combined similar geometry defect types (such as rail cant left and rail cant right) into one geometry defect type (rail cant). This consolidation also removed the rail defects that made up a small portion of the matched rail defects. In this consolidated form, detail fracture defects (TDD), rail cant geometry defects, and warp geometry defects still make up a majority of the matched defects.

Figure 15 presents an exploratory statistical cross-correlation between the rail defects, track geometry defects (consolidated defect classes), and the cumulative MGT of matched rail defects, which will later be presented in the form of multilinear regression equations, MARS equations, and Bayesian statistical models. In this cross-correlation it is observed that there are strong initial indications of a negative relationship between several geometry defects and cumulative MGT of a rail defect, Warp being the highest. The negative correlation between the geometry defects and cumulative MGT indicates that the presence of a geometry defect will reduce the cumulative MGT of a rail defect. This correlation was done using 25 MGT frequency brackets.

DEFECTS	TDD	BB	BHB	BRO	TDT	TDC	OAW	EFBW	TW	HSH	VSH	SW	HW	FH	CH	SD	TOTAL	% of Total	
ALINGMENT	7	1	3	0	0	0	0	0	2	0	1	0	0	0	1	0	15	0.51%	ALINGMENT
ALINGMENT LEFT	22	0	8	2	1	0	0	1	8	2	3	0	3	1	0	5	56	1.92%	ALINGMENT LEFT
ALINGMENT RIGHT	18	0	6	0	0	0	0	2	7	0	0	0	4	0	0	4	41	1.41%	ALINGMENT RIGHT
CLIM	18	1	41	3	3	0	1	2	2	1	13	4	24	1	1	0	115	3.94%	CLIM
CROSSLEVEL	159	2	24	10	8	1	10	35	66	22	16	7	30	16	19	20	445	15.25%	CROSSLEVEL
Curve Speed 3IN	5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	7	0.24%	Curve Speed 3IN
EXECESS ELEVATION	17	0	0	1	0	0	1	3	1	0	4	0	0	0	1	0	28	0.96%	EXECESS ELEVATION
GWR 2ND LEVEL	13	0	0	1	0	0	0	0	2	0	1	0	1	0	0	0	18	0.62%	GWR 2ND LEVEL
LEFT RAIL CANT	301	2	15	9	3	4	4	31	33	10	20	3	20	4	7	55	521	17.85%	LEFT RAIL CANT
Left Vert ACC	4	0	3	0	0	0	0	0	2	0	0	0	1	1	0	1	12	0.41%	Left Vert ACC
LOADED GAGE	6	0	0	1	1	0	0	2	2	0	0	0	1	3	1	8	25	0.86%	LOADED GAGE
PLG 24 2ND LEVEL	24	0	1	1	0	0	0	0	3	3	0	0	0	0	0	6	38	1.30%	PLG 24 2ND LEVEL

PROFILE LEFT	31	0	11	6	0	1	1	13	15	6	4	1	12	5	3	10	119	4.08%	PROFILE LEFT
PROFILE RIGHT	42	0	14	6	1	1	3	11	25	4	2	7	13	5	12	4	150	5.14%	PROFILE RIGHT
RIGHT RAIL CANT	205	1	12	9	3	1	2	21	28	5	12	2	4	2	4	34	346	11.86 %	RIGHT RAIL CANT
TIGHT GAGE	11	0	4	0	2	0	0	1	4	1	0	2	1	0	0	0	26	0.89%	TIGHT GAGE
WARP 31	207	1	45	7	2	2	7	27	61	15	21	7	35	7	15	56	517	17.72 %	WARP 31
WARP 62	65	0	18	3	6	0	3	10	30	8	7	4	38	4	5	7	208	7.13%	WARP 62
WIDE GAGE	131	1	5	5	2	2	1	13	16	4	17	2	8	2	5	10	224	7.68%	WIDE GAGE
TOTAL	1289	9	210	64	32	12	33	174	307	83	123	39	195	51	74	220	2918		
% of Total	44.17 %	0.31 %	7.20 %	2.19 %	1.10 %	0.41 %	1.13 %	5.96%	10.52%	2.84 %	4.22 %	1.34 %	6.68 %	1.75 %	2.54 %	7.54%			
	TDD	BB	BHB	BRO	TDT	TDC	OAW	EFBW	TW	HSH	VSH	SW	HW	FH	CH	SD			

Figure 13 Rail defect vs. Geometry Defect Data (CSX full system)

DEFECTS	TDD	BHB	EFBW	TW	VSH	HW	SD	TOTAL	% of Total	% of TDD
ALIGNMENT	47	17	3	17	4	7	9	104	4%	4%
CROSSLEVEL	177	65	37	68	29	54	20	450	18%	14%
ELEVATION	22	0	3	1	5	0	0	31	1%	2%
LOADED GAGE	46	1	4	7	1	2	14	75	3%	4%
RAIL CANT	506	27	52	61	32	24	89	791	32%	39%
PROFILE	73	25	24	40	6	25	14	207	8%	6%
GAGE	142	9	14	20	17	9	10	221	9%	11%
WARP	272	63	37	91	28	73	63	627	25%	21%
TOTAL	1285	207	174	305	122	194	219	2506		
	51.3%	8.26%	6.94%	12.17%	4.87%	7.74%	8.74%			

Figure 14 Rail defect vs. Geometry Defect Data (CSX full system) Consolidated

ALL DEFECTS 25 MGT FREQ								
	CUM.MGT	BHB	EFBW	HW	SD	TDD	TW	VSH
ALIGNMENT	-0.7147892	0.7199338	0.5594119	0.6002987	0.5834144	0.7139453	0.5499209	0.3811537
CROSSLEVEL	-0.7694986	0.7622581	0.6513747	0.7448449	0.5043068	0.7409107	0.6660712	0.6418097
ELEVATION	-0.3552777	0.4122566	0.3306969	0.3912977	0.2968832	0.3380639	0.2385682	0.2751114
GAGE	-0.7707247	0.8124716	0.6223919	0.7033230	0.6213678	0.7968176	0.6293880	0.6661253
LOADED.GAGE	-0.4577462	0.3755918	0.3528237	0.3538483	0.2689030	0.4545979	0.2232836	0.2040405
PROFILE	-0.6193273	0.5896705	0.6133353	0.5984654	0.5174390	0.6251029	0.6553790	0.4071633
RAIL.CANT	-0.7860837	0.6775659	0.6350637	0.4191745	0.8812645	0.9130556	0.6665193	0.5509936
WARP	-0.8847520	0.8037898	0.6833681	0.7229649	0.7494663	0.8916384	0.7541711	0.5358121

Figure 15 Preliminary Cross-Correlation of Rail Defects vs. Geometry defects.

The Effect of Geometry Defects on Rail Life

The CSX annual tonnage data was used in combination with the rail defect data to determine the cumulative MGT (life of rail) of a given rail defect. This was based on the rail roll date[3] and was calculated using the equation below in Equation 8. The annual MGT term was determined from the provided tonnage data discussed in the DEVELOPMENT OF DATABASE section. The 2012 data was used since it had the most consistency. The annual MGT was calculated by Equation 7.Equation 8 assumes a 2% growth in tonnage per year.

$$MGT\ (weighted)\ = \Sigma\ Segment\ length * \ MGT\ (segment) / \Sigma\ Segment\ lengths$$

Equation 7

$$Cum\ MGT = AGE\left(1 - 2 * \frac{Half\ 2012\ Annual\ MGT}{100}\right) * (2012\ Annual\ MGT)$$

Equation 8

[3] Rail roll date is the date of manufacture of the rail which is usually close to the installation date of the rail in track.

Where:

$$Age = (Year\ found - (Year\ Rolled + 1)$$

It should be noted that this equation does not capture the true life of the rail for relayed rail[4], since the original location of that rail is unknown. This gives an underestimation of the life of these rails. This was acceptable since only a small portion of the rail was relayed. By knowing the life of the rail, in MGT, rail defects with and without geometry defect matches can be compared to see if geometry defects have an overall effect on rail life. It should also be noted that there were some cases where the life of the rail could not be calculated due to a missing term in the original data. Figure 16 below is a table that shows a comparison between matched, matched with repeats, and none matched rail defects. The table also shows a comparison between tangent, curve, and all tracks.

[4] Relay rail is rail that has been moved from one location to another, after an extended period of time in the original or first position.

	Avg. Cum. MGT For All Division By Rail Defects														
	Tangent					Curve					ALL				
Defect	**Match Repeats Tangent**	**Match Unique Tangent**	**Not Matched Tangent**	**Reduction Repeat Tangent**	**Reduction Unique Tangent**	**Match Repeats Curve**	**Match Unique Curve**	**Not Matched Curves**	**Reduction Repeat Curve**	**Reduction Unique Curve**	**Match Repeats All**	**Match Unique All**	**Not Matched All**	**Reduction Repeat All**	**Reduction Unique All**
BB	135.65	171.26	572.36	-76%	-70%	267.41	419.7	439.71	-39%	-5%	188.35	295.5	541.96	-65%	-45%
BHB	330.07	292.35	481.67	-31%	-39%	308.61	298.3	380.77	-19%	-22%	319.07	295.2	468.46	-32%	-37%
BRO	325.67	395.12	713.61	-54%	-45%	385.27	491.5	368.91	4%	33%	359.19	442.0	610.47	-41%	-28%
CH	869.83	812.66	807.81	8%	1%	1019.5	525.9	766.35	33%	-31%	922.65	683.4	800.42	15%	-15%
EFBW	666.83	599.16	799.34	-17%	-25%	356.38	407.9	576.29	-38%	-29%	428.45	482.8	759.09	-44%	-36%
FH	652.95	705.29	854.78	-24%	-17%	395.67	773.6	735.24	-46%	5%	573.79	726.7	838.42	-32%	-13%
HSH	511.96	520.42	699.92	-27%	-26%	570.06	430	464.51	23%	-7%	536.86	477.2	671.21	-20%	-29%
HW	302.16	313.31	598.31	-49%	-48%	235.92	273.1	483.96	-51%	-44%	268.62	295.0	576.73	-53%	-49%
OAW	613.61	490.25	733.86	-16%	-33%	363.86	428.9	493.14	-26%	-13%	502.61	467.2	717.29	-30%	-35%
SD	660.68	564.5	624.7	6%	-10%	394.86	390.5	414.78	-5%	-6%	428.8	430.6	519.88	-18%	-17%
SW	503.62	401.23	578.44	-13%	-31%	451.71	327.7	404.55	12%	-19%	477.66	373.4	547.69	-13%	-32%
TD	NA	NA	NA	NA	NA	691.27	691.2	526.92	31%	31%	691.27	691.2	526.92	31%	31%
TDC	NA	377.2	657.42	NA	-43%	377.2	368.3	371.44	2%	-1%	377.2	504.8	569.19	-34%	-11%
TDD	508	519.67	690.01	-26%	-25%	404.23	423.5	434.35	-7%	-2%	477.68	448.7	506.31	-6%	-11%
TDT	240.4	155.64	750.5	-68%	-79%	322.78	315.6	662.92	-51%	-52%	263.93	207.2	729.77	-64%	-72%
TW	584.95	635.09	775.46	-25%	-18%	469.24	475.4	542.5	-14%	-12%	515.3	553.8	741.97	-31%	-25%
VSH	427.64	528.93	658.06	-35%	-20%	336.39	407.6	561.74	-40%	-27%	356.92	445.4	621.33	-43%	-28%
TOTAL	**488.11**	**503.64**	**703.70**	**-31%**	**-28%**	**396.04**	**413.2**	**477.15**	**-17%**	**-13%**	**423.82**	**446.7**	**646.19**	**-34%**	**-31%**

Figure 16 Comparison of rail life by rail defect type

As shown above, there is a significant loss of rail life associated with the presence of a geometry defect, 31% for all matches and 34% for all repeat matches. From this it is noted that multiple geometry defects do have an impact on reducing rail life, but the decrease in rail life from multiple geometry is not much more than the decrease in rail life from one geometry defect. To better understand the life of the rail, distribution plots were made. These plots were based off a 25 MGT bracket shown in Figure 17. Figure 18 shows the percentage of defect in each bracket and Figure 19 shows the cumulative percentage of defects.

CUM MGT	ALIGNMENT	CROSSLEVEL	ELEVATION	LOADED GAGE	RAIL CANT	LEFT VERT ACC	PROFILE	RUN OFF	GAGE	WARP	Total	%
50	4	34	3	6	34	2	9	1	21	39	153	5.4%
100	18	39	3	10	56	0	15	0	22	41	204	7.2%
150	5	40	1	9	77	0	10	0	23	47	212	7.5%
200	11	47	2	19	38	0	15	0	20	57	209	7.4%
250	9	35	0	3	55	0	16	0	12	49	179	6.3%
300	9	11	1	3	73	1	8	0	7	40	153	5.4%
350	9	17	1	2	62	0	22	0	14	46	173	6.1%
400	4	22	1	3	59	1	7	0	13	34	144	5.1%
450	8	21	2	5	105	1	10	0	22	31	205	7.3%
500	3	18	2	2	47	1	10	0	8	39	130	4.6%
550	3	35	2	1	47	0	13	0	13	40	154	5.4%
600	2	21	0	2	33	2	9	0	11	32	112	4.0%
650	4	27	2	0	16	1	17	0	7	33	107	3.8%
700	2	18	4	1	24	0	13	0	7	31	100	3.5%
750	2	27	0	3	28	1	10	0	10	23	104	3.7%
800	1	26	1	2	18	0	10	0	9	25	92	3.3%
850	3	19	0	2	18	0	18	0	8	17	85	3.0%
900	1	9	0	5	9	0	4	0	1	10	39	1.4%
950	1	8	0	3	4	0	3	0	1	8	28	1.0%
1000	2	5	0	1	9	0	4	0	4	11	36	1.3%
1050	1	11	3	0	7	0	2	0	4	10	38	1.3%
1100	0	9	4	3	7	0	6	0	1	16	46	1.6%
1150	1	10	0	0	4	0	2	0	1	8	26	0.9%
1200	0	6	0	0	4	0	2	0	0	6	18	0.6%
1250	1	3	0	1	1	0	4	0	0	4	14	0.5%
1300	2	2	0	0	0	0	1	0	1	1	7	0.2%
1350	1	5	0	0	1	1	1	0	1	2	12	0.4%
1400	0	3	0	0	0	0	5	0	0	0	8	0.3%
1450	0	3	0	0	1	0	1	0	0	2	7	0.2%
1500	0	1	0	0	1	0	2	0	0	2	6	0.2%
1550	0	5	0	0	3	0	3	0	0	2	13	0.5%
1600	0	1	0	0	1	0	3	0	0	1	6	0.2%
1650	0	2	0	0	0	0	2	0	0	0	4	0.1%
1700	0	0	0	0	0	0	0	0	0	2	2	0.1%
1750	0	0	0	0	0	0	1	0	0	0	1	0.0%
Total	107	540	32	86	842	11	258	1	241	709	2827	100.0%
	3.8%	**19.1%**	**1.1%**	**3.0%**	**29.8%**	**0.4%**	**9.1%**	**0.0%**	**8.5%**	**25.1%**	**1**	

Figure 17 Distribution of Rail Life by 25MGT Brackets

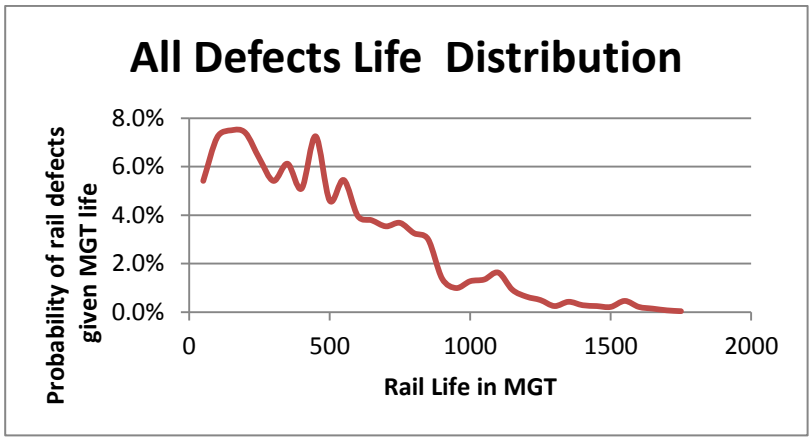


Figure 18 Percentage of defects in each MGT bracket

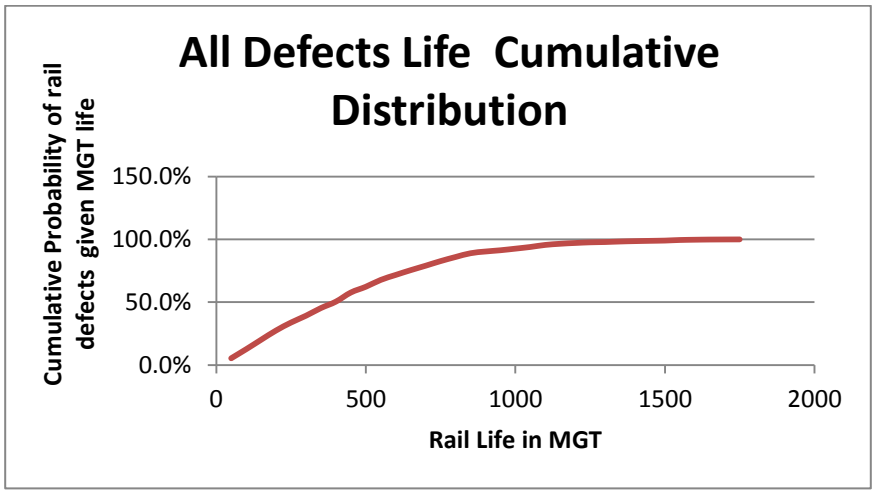


Figure 19 Cumulative Percentage of Defects by MGT

As can be observed in the above figures, approximately 50% of matched rail defects had a MGT life of less than 400 MGT. This further suggests that geometry defects have a large impact on the overall life of a rail defect, since matches occur at a much lower rail life than the average of both matched and non-matched rail defects.

Multilinear and Mars Analyses

Once it was determined that geometry defects had an impact on rail life, a series of multilinear regression analyses were performed. The first of these analyses was done with basic multilinear analysis function in R looking at all the geometry defect types and their impact on rail life. Equation 6 below shows the results of this analysis. The value of the rail life for any given geometry defect is shown in Figure 20. The a_i value is a constant value that is determined by which geometry defect (GD_i) is being observed. Though multiple defects can be used in this equation, as in more than one defect type can be entered, it is best suited for one geometry defect at a time.

$$MGT = 1750(0.8435 + \Sigma a_i GD_i)$$ Equation 9

GD_i	a_i
Alignment	-0.1585
Cant	-0.3565
Cross-level	-0.1585
Elevation	-0.0252
Loaded Gage	-0.0718
Warp	-0.3719
Run-off	-0.2552
Profile	-0.0569
Gage	+0.2341

R2 = 92.86%

alignment =	0	1	0	0	0	0	0	0	1	0	0	0	1
Cross-level=	0	0	1	0	0	0	0	0	1	1	0	1	1
Elevation	0	0	0	1	0	0	0	0	0	0	0	0	1
loaded gage	0	0	0	0	1	0	0	0	0	0	0	0	0
rail cant	0	0	0	0	0	1	0	0	0	0	1	0	0
Profile	0	0	0	0	0	0	1	0	0	0	0	1	0
warp	0	0	0	0	0	0	0	1	0	1	0	0	0
Life (MGT)	1476	1199	916	1432	1350	852	1377	825	638	265	1262	816	594
reduction in life		18.8%	38.0%	3.0%	8.5%	42.3%	6.7%	44.1%	56.8%	82.1%	14.5%	44.7%	59.7%

Figure 20 Reduction in Rail Life due to Track Geometry Defects

As shown in Figure 20, the type of geometry defect can greatly impact the life of a rail defect. To better understand the effects of certain geometry defects on rail life, and their interactions with each other, the MARS approach is used. As discussed earlier, MARS shows which predictive variable (geometry defect) are the most important and impactful. The MARS function can be shown as follows:

$$Cum\ MGT = f\{GD\} + C \qquad \text{Equation 10}$$

Using this approach the following MARS analysis was performed:

$$MGT = 447.37+86.60(BF_2)+28.65(BF_4)-47.90(BF_5) \qquad \text{Equation 11}$$

Where:

BF_2 =max(0,6-WARP)

BF_4 =max(0,18-Rail Cant)

BF_5 =max(0,Alignment-0)

The results of this analysis show that the geometry defects that have an important impact on the MGT life of the rail are Warp, Rail Cant, and Alignment defects. Below in Figure 21 is the sensitivity rail life has to different amounts of the specified geometry defects within the MARS analysis. It should be noted that with zero geometry defects, the MGT life of a rail is approximately 1500 MGT, which is a typically accepted value for the life of a rail.

Warp =	0	1	1	1	2	6	6
Rail cant =	0	0	1	1	1	18	18
Alignmen	0	0	0	1	0	0	1
BF2=	6	5	5	5	4	0	0
BF4=	18	18	17	17	17	0	0
BF5=	0	0	0	1	0	0	1
MGT	1513	1426	1398	1350	1311	477	429
reduction		5.7%	7.6%	10.8%	13.3%	68.4%	71.6%

Figure 21 MARS Sensitivity to Geometry Defects

After the initial analysis of the data using MARS, it was determined that a more in depth look of the impact of geometry defects on tangent and curve defects should be performed. To get a better understanding of the impact of geometry defects for these two cases, Warp defects were split into Warp 31 defects and Warp 62

defects. After splitting the Warp defects, the MARS analysis was rerun for both tangent and curve track. The results for both can be seen below in Equations12 and 13.

Equation 12 MARS Analysis for Tangent Track

CUM.MGT = 355.9443 + 69.37212 * max(0, 5 - CROSSLEVEL)
- 156.74 * max(0, GAGE - 1)
+ 213.68* max(0, 2 - RAIL.CANT)
+ 188.87 * max(0, 2 - Warp.62)

Equation 13 MARS Analysis for Curve Track

CUM.MGT = 258.4052 + 172.4302 * max(0, 2 - ALIGNMENT)
+ 115.71 * max(0, 2 - CROSSLEVEL)
+ 66.55 * max(0, 5 - RAIL.CANT)
+ 94.52 * max(0, 4 - WARP.31)

From these analyses, it can be observed that rail cant and warp are still very important for determining the MGT life of a rail. Tangent track is more impacted by Warp 62 where curve track is more impacted by Warp 31. A Warp 31 defect is a warp defect measured with a 31ft cord where a Warp 62 is measured with a 62ft cord. Warp 31 are more likely to occur on curve track compared to Warp 62. This is illustrated in Figures 22 through 24 in the form of distributions of these specific geometry defects. Figure 22 shows how Warp 31 is much more dominant on curves as compared to tangent track. This is reversed for Warp 62 in Figure 23. This agrees with the MARS analyses assessments of these defects, since no basis functions for Warp 31 were kept on tangent track and no Warp 62 were kept on curve track. Figure 24 shows that Rail Cant is more prominent on curve track, but more impactful per occurrence on tangent track. This also agrees with the MARS analyses. The tangent analysis shows that Rail Cant occurring has a large impact on life (200 MGT per occurrence) were the curve analysis shows that Rail Cant needs to occur often to have the same impact on life. It

should also be noted that Crosslevel, Gage, and Alignment are also impactful in determining the cumulative MGT of a rail.

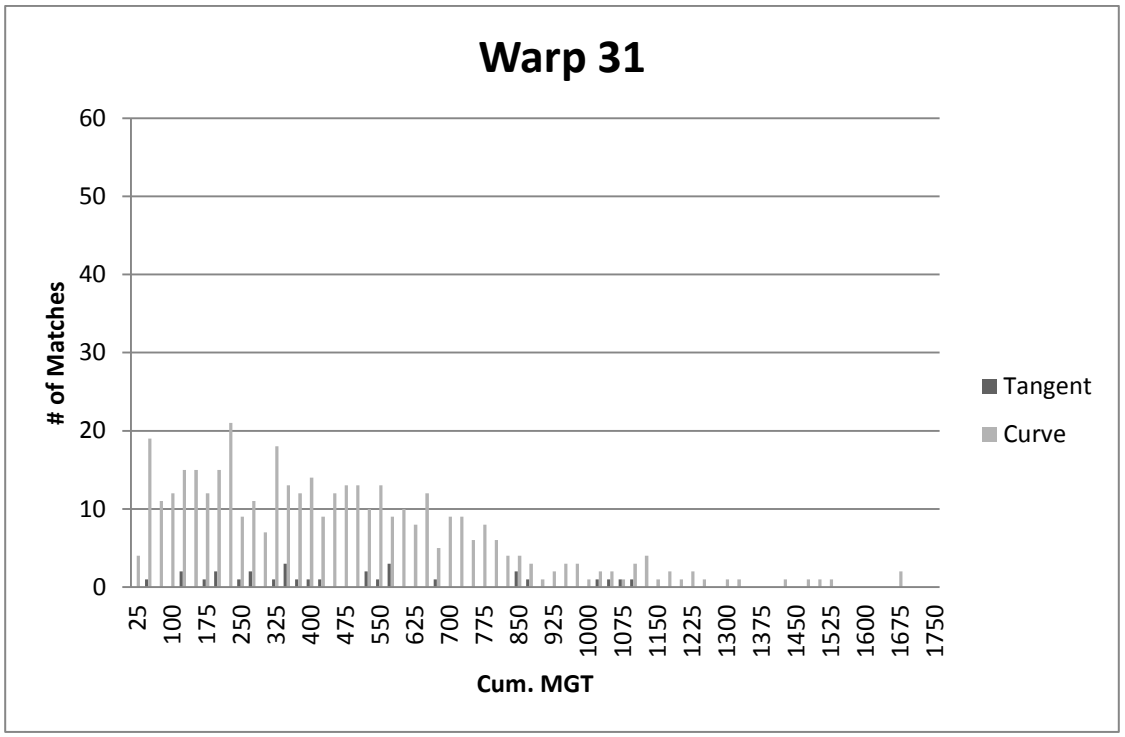


Figure 22 Distribution of Warp 31

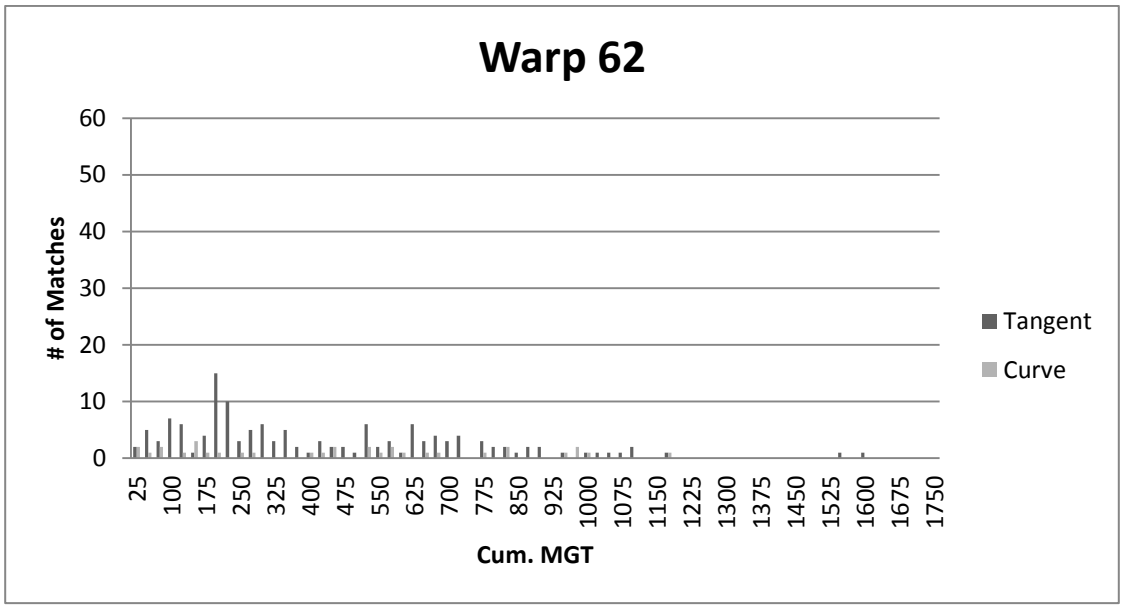


Figure 23 Distribution of Warp 62

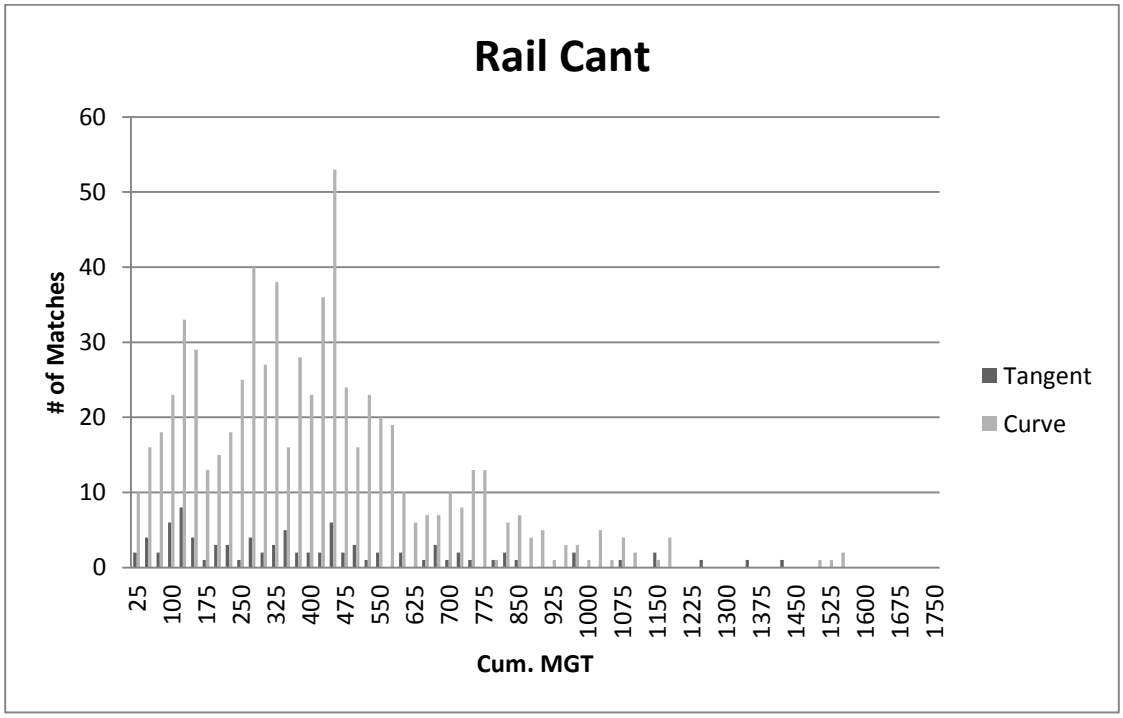


Figure 24 Distribution of Rail Cant

These six geometry defects were determined, based off the MARS analyses, to be the most important and impactful geometry defects. With this the multilinear regression analysis was revisited. Five new equations based off these six parameters were created. Their equations and sensitivity to geometry defects can be seen below.

Rail life for Tangent Track using six key geometry variables as identified by MARS Equation 14

$MGT=1750(0.739+\Sigma a_i GD_i)$

GD_i	a_i
Alignment	+0.0198
Cant	-0.4576
Cross-level	-0.2583
Warp 31	-0.1579
Warp 62	-0.3583
Gage	-0.2583

Rail life for Tangent Track using four key geometry variables as identified by MARS Equation 15

$MGT=1750(0.7156+\Sigma a_i GD_i)$

GD_i	a_i
Cant	-0.4902
Cross-level	-0.2909
Warp 62	-0.4902
Gage	-0.2249

Rail life for Curve Track using six key geometry variables as identified by MARS Equation 16

$MGT=1750(0.7777+\Sigma a_i GD_i)$

GD_i	a_i
Alignment	-0.1985
Cant	-0.1342
Cross-level	-0.1949
Warp 31	-0.4987
Warp 62	-0.0162
Gage	-0.0813

Rail life for Curve Track using four key geometry variables as identified by MARS Equation 17

MGT=1750(0.7704+Σ$a_i$$GD_i$)

GD_i	a_i
Alignment	-0.1661
Cant	-0.2211
Cross-level	-0.1661
Warp 31	-0.5193

Rail life for All Track using six key geometry variables as identified by MARS Equation 18

MGT=1750(0.784+Σ$a_i$$GD_i$)

GD_i	a_i
Alignment	-0.2305
Cant	-0.2566
Cross-level	-0.291
Warp 31	-0.3808
Warp 62	-0.1255
Gage	-0.0012

Tangent All	1750							
Intercept	0.739							
Alignment	0.0198	0	1	0	0	0	0	0
Crosslevel	-0.2582	0	0	1	0	0	0	0
Gage	-0.2583	0	0	0	1	0	0	0
Rail Cant	-0.4576	0	0	0	0	1	0	0
Warp 31	-0.1579	0	0	0	0	0	1	0
Warp 62	-0.3583	0	0	0	0	0	0	1
	MGT	1293.25	1327.9	841.4	841.225	492.45	1016.925	666.225
Reduction in Life			2.68%	-34.94%	-34.95%	-61.92%	-21.37%	-48.48%

Figure 25 Rail life for Tangent Track using six key geometry variables as identified by MARS

Tangent Specific	1750					
Intercept	0.7156					
Gage	-0.2249	0	1	0	0	0
Crosslevel	-0.2909	0	0	1	0	0
Rail.Cant	-0.4902	0	0	0	1	0
Warp 62	-0.4239	0	0	0	0	1
	MGT	1252.3	858.725	743.225	394.45	510.475
Reduction in Life			-40.65%	-68.50%	-59.24%	-39.96%

Figure 26 Rail life for Tangent Track using four key geometry variables as identified by MARS

Curve All	1750							
Intercept	0.7777							
Alignment	-0.1985	0	1	0	0	0	0	0
Crosslevel	-0.1949	0	0	1	0	0	0	0
Gage	-0.0813	0	0	0	1	0	0	0
Rail Cant	-0.1342	0	0	0	0	1	0	0
Warp 31	-0.4987	0	0	0	0	0	1	0
Warp 62	-0.01618	0	0	0	0	0	0	1
	MGT	1360.975	1013.6	1019.9	1218.7	1126.125	488.25	1332.66
Reduction in Life			-25.52%	-25.06%	-10.45%	-17.26%	-64.12%	-2.08%

Figure 27 Rail life for Curve Track using six key geometry variables as identified by MARS

Curve Specific	1750					
Intercept	0.7704					
Alignment	-0.16614	0	1	0	0	0
Crosslevel	-0.2747	0	0	1	0	0
Rail Cant	-0.2211	0	0	0	1	0
Warp 31	-0.5193	0	0	0	0	1
	MGT	1348.2	1057.455	867.475	961.275	439.425
Reduction in Life			-21.57%	-35.66%	-28.70%	-67.41%

Figure 28 Rail life for Curve Track using four key geometry variables as identified by MARS

All Track	1750							
Intercept	0.78399							
Alignment	-0.2305	0	1	0	0	0	0	0
Crosslevel	-0.291	0	0	1	0	0	0	0
Gage	-0.0012	0	0	0	1	0	0	0
Rail Cant	-0.2566	0	0	0	0	1	0	0
Warp 31	-0.3808	0	0	0	0	0	1	0
Warp 62	-0.1255	0	0	0	0	0	0	1
	MGT	1371.983	968.6075	862.7325	1369.883	922.9325	705.5825	1152.358
Reduction in Life			-29.40%	-37.12%	-0.15%	-32.73%	-48.57%	-16.01%

Figure 29 Rail life for All Track using six key geometry variables as identified by MARS

Probability Analysis

Probability Analysis examines the probability of a rail defect occurring given a geometry defect preceding it. The probability analysis approaches used in this study include:

- Random Analysis
- Conditional Probability Analysis
- Bayes' Theorem probability analysis
- Naïve Bayes probability analysis
- Bayesian network analysis

This section will look at multiple statistical (probability) approaches for predicting the probability of that a rail defect will occur given that there is a geometry defect at that same location that is present before the rail defect, i.e. the geometry defect precedes the rail defect. While there are many contributing factors into the development of a rail defect, this section addresses the effect of a geometry exception (geometry defect) in the likelihood (probability) of the development of a rail defect. Conditional Probability Analyses (Bayesian methods) are used to show this relationship.

The first step in defining the relationships between the two types of defects is to determine the random probability of a defect, either rail or geometry, occurring at any given location on the track. This was done using Equation 19 and Equation 20 below.

$$P(RD) = \frac{39ft * Annual\ Rail\ Defects}{Length\ of\ track\ in\ ft * 2} \qquad \text{Equation 19}$$

$$P(GD) = \frac{39ft * Annual\ Geometry\ Defects}{Length\ of\ track\ in\ ft} \qquad \text{Equation 20}$$

These equations give the random probability that a rail defect P(RD) or geometry defect P(GD) would occur in any one location on track independent of any other factors. A length of 39 feet was used for determining the length of track with one defect. The rail defect equation is divided by two to account for two rails per track. The complements were also determined for these defects. These are the probability that a defect type would not occur at a given location. Table 1 below has the results of these equations for the entire railroad (CSX). The geometry defects shown in Figure 30 were determined in a previous analysis to be the most significant.

Entire Railroad	
P(RD)	0.180%
P(GD)	1.19%
P(Alignment)	0.054%
P(Crosslevel)	0.30%
P(Gage)	0.15%
P(Rail Cant)	0.37%
P(Warp 31)	0.22%
P(Warp 62)	0.095%

Figure 30 Random Probability of Rail and Geometry Defects.

As shown in the table above, the random probability of a defect of any type occurring is very low. For a rail defect, the probability of the rail defect occurring in any one location randomly is 0.18%. For a geometry defect, this probability that the geometry defect will occur in any one location randomly is 1.19%. Note, this higher probability is due to the larger number of geometry defect that occur vs. rail defects, per CSX data presented previously. Also shown in Figure 30 are the probabilities for any specific geometry defect occurring on a random basis.

However, as seen from the previous analyses, rail defects do not occur randomly, but have an increased probability of occurring if preceded by a geometry defect, i.e. there is some relationship between geometry and rail defects. In order to evaluate this increased probability, a condition probability analysis with a conditional probability, P(R|G), calculated using Bayes' Theorem (Bayes' Rule).

The most common and basic method of determining conditional probability is given below in Equation 13 for the probability of a geometry defect given a rail defect follows it [P(GD/RD)]. The joint probability [P(GD∩RD)] used in Equation 21 was calculated using Equation 21, but replacing annual rail defect occurrences with annual matched rail occurrences.

$$P(GD|RD) = \frac{P(GD \cap RD)}{P(RD)} \qquad \text{Equation 21}$$

The results of the conditional probability for different geometry defects are shown below in Figure 31.

Entire Railroad	
P(Alignment|RD)	0.418%
P(Crosslevel|RD)	2.097%
P(Gage|RD)	0.933%
P(Rail Cant|RD)	3.227%
P(Warp 31|RD)	1.925%
P(Warp 62|RD)	0.776%
P(GD|RD)	9.38%

Figure 31 Conditional Probabilities for specific geometry defects given a rail defect followed.

The P(GD|RD) was first found so it could be used in Bayes' Theorem (Theorem 1) to calculate the conditional probability of a rail defect when a geometry defect preceded it.

$$Posterior = \frac{Likelihood\ X\ Prior}{Evidence}$$ Theorem 1

Thus, the equation for P(RD|GD), the conditional probabilities for a rail defect to occur given a geometry defect preceded it, is calculated using Bayes' Theorem and is rewritten as shown in Equation 22.

$$P(RD|GD) = \frac{P(GD|RD)*P(RD)}{P(GD|RD)*P(RD)+P(GD|\neg RD)*P(\neg RD)}$$ Equation 22

This conditional probability of a rail defect, P(RD/GD), was calculated in terms of each geometry defect as well as any geometry defect. The results from the Bayes' analysis are shown below in Figure 32. From this table we can determine that a rail defect is about eight times more likely to occur given there is a geometry defect that precedes it, as compared to it occurring randomly (0.18%).

Entire Railroad		
	P(RD|GD)	Likelihood more to occur
P(RD|Alignment)	1.41%	7.84
P(RD|Crosslevel)	1.24%	6.91
P(RD|Gage)	1.12%	6.21
P(RD|Rail Cant)	1.59%	8.83
P(RD|Warp 31)	1.57%	8.75
P(RD|Warp 62)	1.47%	8.17
P(RD|GD)	1.44%	8.01

Figure 32 Conditional probabilities for a rail defect to occur given a geometry defect preceded it from Bayes' Theorem

The Bayes' Theorem methodology applied above looks at the condition where one geometry defect occurs prior to a rail defect. The conditional probability of a rail defect occurring after two or more geometry defects must also be considered, since this is a condition that was observed repeatedly in the actual data. To do this a different method must be used. One approach that can be used to determine the

conditional probability of a rail defect after multiple geometry defects is Naïve Bayes. This is an extension of Bayes' Theorem to cover the situation where multiple conditions precede an event and the likelihood and evidence functions used in Bayes' Theorem can be rewritten as shown in Equation 23 and Equation 24.

$$Likelihood = P(GD_1|RD) * P(GD_2|RD) * \ldots * P(GD_n|RD) * P(RD) \quad \text{Equation 23}$$

$$Evidence = P(GD_1|RD) * P(GD_2|RD) * \ldots * P(GD_n|RD) * P(RD) + P(GD_1|\neg RD) * P(GD_2|\neg RD) * \ldots * P(GD_n|\neg RD) * P(\neg RD) \quad \text{Equation 24}$$

Where n is the number of geometry defects present prior to the rail defect. Using the same inputs as already discussed when applying Bayes' Theorem, the results from the Naïve Bayes analysis is presented in Figure 33 for several different geometry defect combinations.

Entire Railroad							
	Alignment	Crosslevel	Gage	Rail Cant	Warp 31	Warp 62	P(RD | GD)
# of defects	1	0	0	0	0	0	1.41%
	0	1	0	0	0	0	1.24%
	0	0	1	0	0	0	1.12%
	0	0	0	1	0	0	1.59%
	0	0	0	0	1	0	1.57%
	0	0	0	0	0	1	1.47%
	1	1	0	0	0	0	9.08%
	1	0	1	0	0	0	8.23%
	1	0	0	1	0	0	11.36%
	1	0	0	0	1	0	11.26%
	1	0	0	0	0	1	10.59%
	0	1	1	0	0	0	7.32%
	0	1	0	1	0	0	10.13%
	0	1	0	0	1	0	10.05%
	0	1	0	0	0	1	9.44%
	0	0	1	1	0	0	9.19%
	0	0	1	0	1	0	9.11%
	0	0	1	0	0	1	8.56%
	0	0	0	1	1	0	12.53%
	0	0	0	1	0	1	11.79%
	0	1	1	0	0	1	39.52%
	0	1	0	1	1	0	50.01%
	1	0	0	1	1	0	53.19%
	0	1	1	1	0	1	85.41%
	1	1	0	1	1	0	88.81%

Figure 33 Results of different geometry defect combinations using Naïve Bayes

The Naïve Bayes analysis shows that multiple geometry defects prior to a rail defect have a dramatic impact on the occurrence of a rail defect. Thus while the probability of a rail defect occurring after a Warp 31 defect is 1.6% (as compared to the random probability of 0.18%) , the probability of a rail defect occurring after a Cross-level and Warp 31 defect increases to 10%, the probability of a rail defect occurring after a Cross-level, Cant and Warp 31 defect jumps to 50%, and the

probability of a rail defect occurring after an alignment, Cross-level, Cant and Warp 31 defect increases to 88.8%.

However, it should be noted that the Naïve Bayes method is still relatively simple and furthermore has some built-in independency assumptions in the interaction of the geometry defects. This can introduce an error into the conditional probability analysis. To further improve the analysis, reduce these errors, and to develop a more accurate conditional probability model for the relationship between multiple geometry defects and associated rail defects, a Bayesian Network was constructed.

Bayesian networks are a graphical probabilistic model used for a set of random variables and their conditional probability. The variables are represented by nodes and the dependencies between variables are represented by an arc or a directed edge. Each node has a conditional probability attached to it. These conditional probabilities are calculated using Bayes' Theorem. Nodes with conditional probability attached to them are called 'Child' nodes. These nodes are conditioned by 'Parent' nodes. A parent node is a node with no arcs or direct edges leading to it. These nodes have a single probability distribution.

In order to apply this Bayesian Network approach, and to develop a model based on the CSX data, a program called NETICA was used. Figure 34 illustrates the structure of the Bayesian Network developed within NETICA. The RailAll node is the only parent node in this network, and represents the conditional probability of a rail defect occurring after one or more geometry defects. The geometry defect nodes are all child nodes of the RailAll node. This means their values are conditioned by the RailAll node. The variables used for this network are discrete variables.

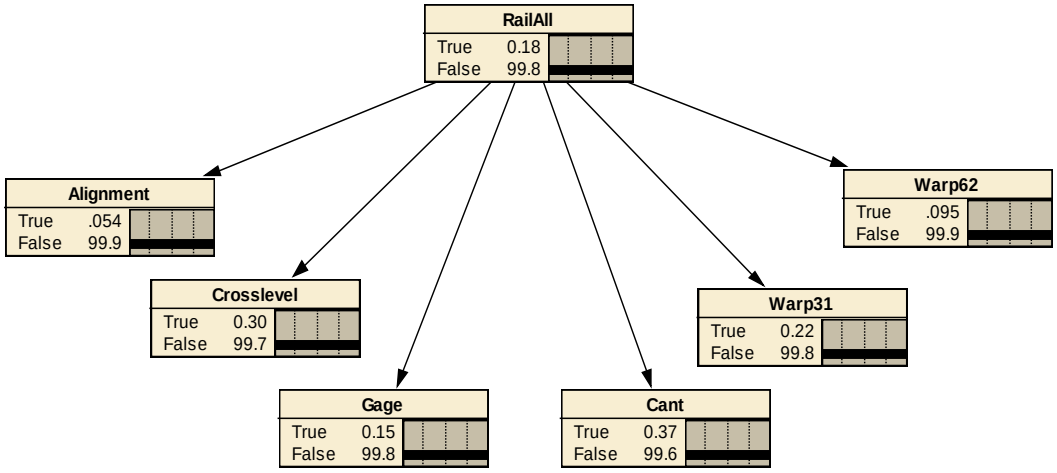


Figure 34 Bayesian network in NETICA; Random Case.

The RailAll node is the pure random probability of a rail defect occurring anywhere on the railroad. Thus, as can be seen in Figure 35, the random probability of having a rail defect is 0.18%, as calculated previously. The other nodes are the conditional probability of a specific geometry defect will occur based on the RailAll node. This value is calculated from a set of inputs placed into a true/false table in NETICA shown in Figure 33.

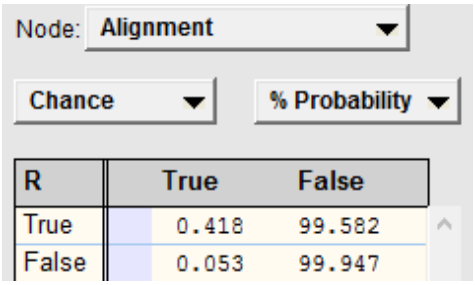


Figure 35 Input table for Alignment Defect in NETICA

The values put into the table in Figure 33 are the conditional probabilities of a geometry defect occurring if followed by a rail defect and a geometry defect occurring if not followed by a rail defect.

With the Bayesian Network model, the conditional probability of a rail defect given a geometry defect can be determined. This is done by setting the geometry defect(s) that occurred as true (100%) and the rest as false (0%). This will calculate the probability of a rail defect occurring given the selected geometry defect(s). The results of several different cases are shown in the following figures.

Figure 36 represents a Warp 31 defect; where the conditional probability of a rail defect following a Warp 31 defect is 1.48%.

Figure 37 represents Warp 31 and Rail Cant defects; where the conditional probability of a rail defect following a Warp 31 and Cant defect is 12.1%..

Figure 38 represents Warp 31, Rail Cant, and Cross-level defects; where the conditional probability of a rail defect following a Warp 31, Cant and Cross-level defect is 49.6%.

Figure 39 represents Warp 31, Rail Cant, Cross-level, and Alignment defects; where the conditional probability of a rail defect following these four defects is 88.6%.

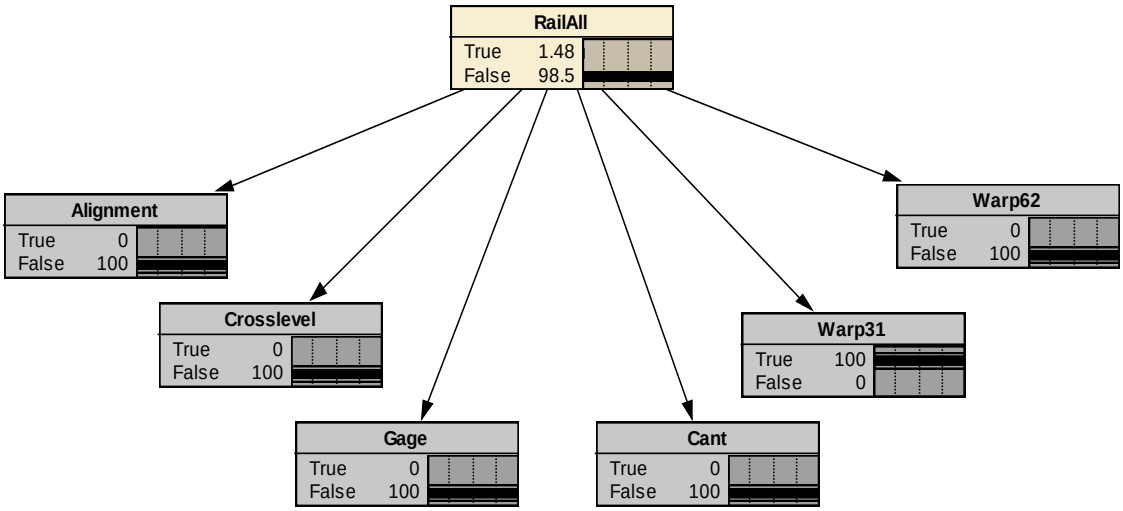


Figure 36 Bayesian Network for a Warp 31 Defect.

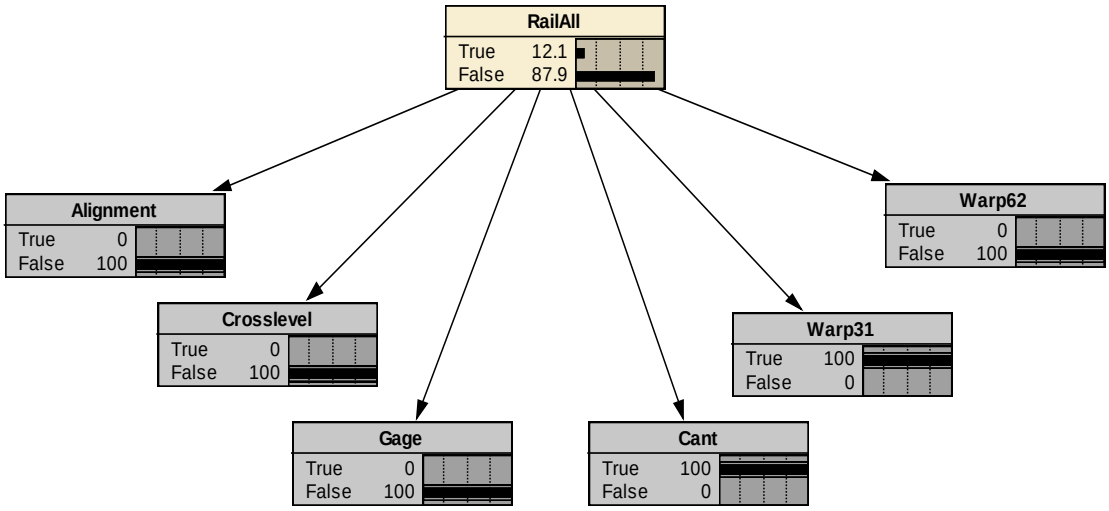


Figure 37 Bayesian Network for a Rail Cant and a Warp 31 Defect.

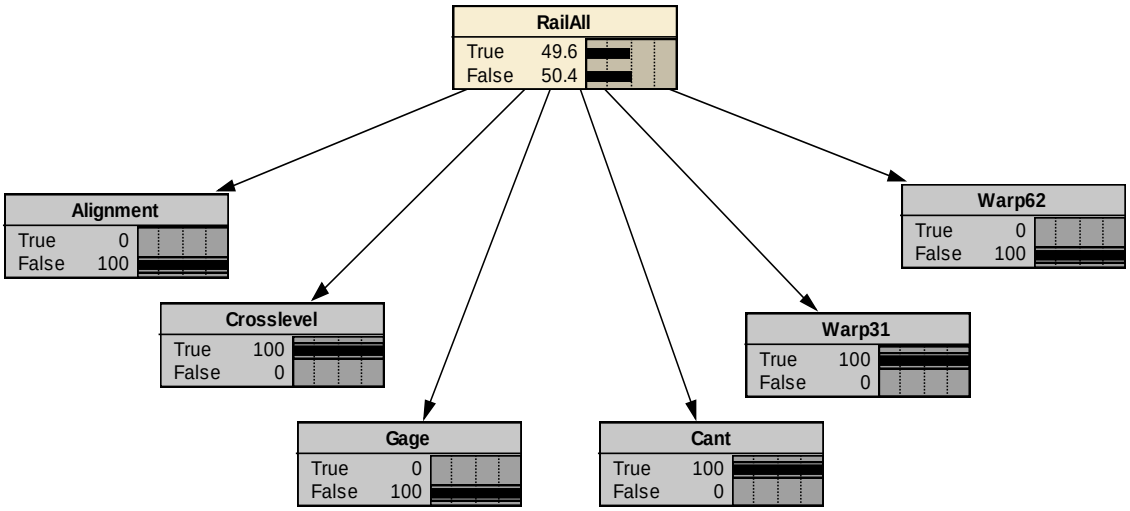


Figure 38 Bayesian Network for a Warp 31, a Rail Cant, and a Crosslevel Defect.

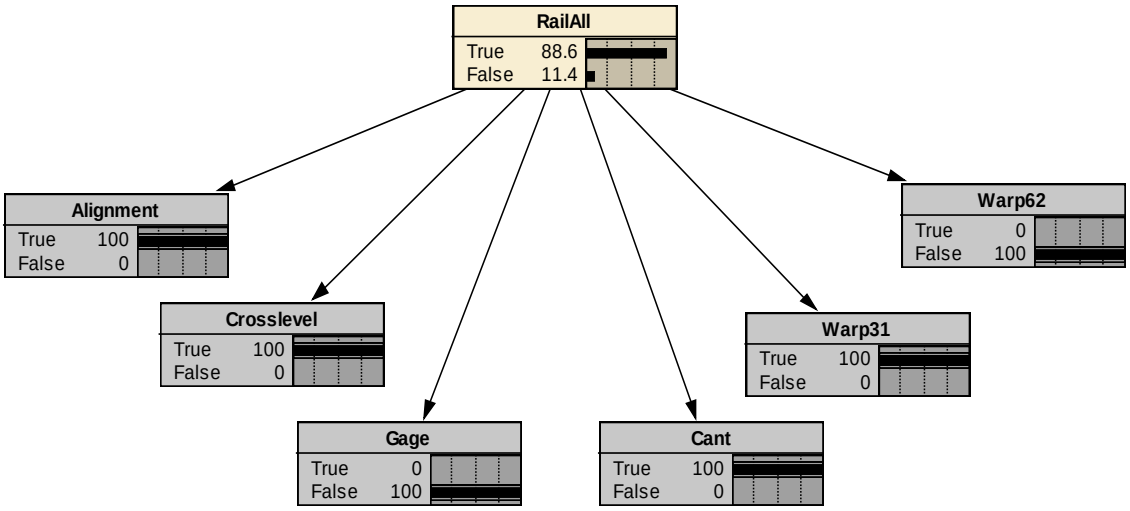


Figure 39 Bayesian Network for a Warp 31, a Rail Cant, a Crosslevel, and an Alignment Defect.

Note, the Bayesian Network model gives a very similar answer to that given by the Naïve Bayes' model shown earlier. The key difference is that the Bayesian

network is assuming that only those two geometry defects had an effect on the development of the rail defect. The Naïve Bayes model assumes that the other geometry defects could have had an effect, increasing the likelihood of a rail defect slightly. A comparison of the two models can be seen below in Figure 40. As can be seen, the differences are relatively small but distinct.

Entire Railroad								
	Alignment	Crosslevel	Gage	Rail Cant	Warp 31	Warp 62	P(RD | GD) Naïve	P(RD|GD) Network
# of defects	1	0	0	0	0	0	1.41%	1.30%
	0	1	0	0	0	0	1.24%	1.17%
	0	0	1	0	0	0	1.12%	1.04%
	0	0	0	1	0	0	1.59%	1.51%
	0	0	0	0	1	0	1.57%	1.48%
	0	0	0	0	0	1	1.47%	1.36%
	1	1	0	0	0	0	9.08%	8.54%
	1	0	1	0	0	0	8.23%	7.66%
	1	0	0	1	0	0	11.36%	10.80%
	1	0	0	0	1	0	11.26%	10.60%
	1	0	0	0	0	1	10.59%	9.87%
	0	1	1	0	0	0	7.32%	6.94%
	0	1	0	1	0	0	10.13%	9.82%
	0	1	0	0	1	0	10.05%	9.63%
	0	1	0	0	0	1	9.44%	8.96%
	0	0	1	1	0	0	9.19%	8.82%
	0	0	1	0	1	0	9.11%	8.64%
	0	0	1	0	0	1	8.56%	8.04%
	0	0	0	1	1	0	12.53%	12.10%
	0	0	0	1	0	1	11.79%	11.30%
	0	1	1	0	0	1	39.52%	38.30%
	0	1	0	1	1	0	50.01%	49.60%
	1	0	0	1	1	0	53.19%	52.20%
	0	1	1	1	0	1	85.41%	85.10%
	1	1	0	1	1	0	88.81%	88.60%

Figure 40 Comparison of Naïve Bayes and Bayesian networks models for select geometry defects.

Tangent and Curve Comparison

The previous analysis included all track, both curve and tangent. A comparison between tangent and curve track conditional probabilities was also performed to see which geometry defects had a larger impact on curve track and tangent track. The same process was carried out with each set of tangent and curve track data. Figure 41 presents a comparison of the random probabilities of rail and geometry defects occurring on tangent, curve, and all tracks. Equations 19 and 20 were both used to calculate these values.

	TANGENT	CURVE	ALL
P(RD)	0.186%	0.152%	0.18%
P(GD)	0.527%	2.743%	1.19%
P(Alignment)	0.012%	0.151%	0.05%
P(Crosslevel)	0.324%	0.242%	0.30%
P(Gage)	0.028%	0.441%	0.15%
P(Rail Cant)	0.048%	1.120%	0.37%
P(Warp 31)	0.017%	0.705%	0.22%
P(Warp 62)	0.098%	0.084%	0.10%

Figure 41 Tangent, Curve, and All random probability of a defect

As can be seen in Figure 41, the curve track has a larger random probability of a geometry defect occurring (for all defects except Warp 62) than the tangent track. The tangent track has a larger random probability of a rail defect occurring (and also of a Warp 62 defect occurring).

Figure 42 presents the conditional probabilities of the geometry defects given a rail defect ,which are used as input into Bayes' Theorem.

	TANGENT	CURVE	ALL
P(Alignment|RD)	0.101%	1.377%	0.42%
P(Crosslevel|RD)	2.287%	1.647%	2.10%
P(Gage|RD)	0.202%	3.144%	0.93%
P(Rail Cant|RD)	0.547%	11.334%	3.23%
P(Warp 31|RD)	0.223%	7.067%	1.93%
P(Warp 62|RD)	0.875%	0.524%	0.78%
P(GD|RD)	4.236%	25.094%	9.38%

Figure 42 Conditional probabilities of geometry defects

With the geometry defect's conditional probabilities calculated, per Figure 42, Bayes' Theorem can be applied to calculate the conditional probability of rail defects occurring after a geometry defect. The results of Bayes' Theorem Conditional probability analyses are shown below in Figure 43.

Bayes' Theorem						
	TANGENT		CURVE		ALL	
	P(RD|GD)	Likelihood more to occur	P(RD|GD)	Likelihood more to occur	P(RD|GD)	Likelihood more to occur
P(RD)	0.186%		0.15%		0.18%	
P(RD|Alignment)	1.50%	8.34	1.37%	9.13	1.41%	7.84
P(RD|Crosslevel)	1.29%	7.18	1.02%	6.81	1.24%	6.91
P(RD|Gage)	1.34%	7.43	1.07%	7.13	1.12%	6.21
P(RD|Rail Cant)	2.06%	11.44	1.51%	10.08	1.59%	8.83
P(RD|Warp 31)	2.44%	13.54	1.50%	9.99	1.57%	8.75
P(RD|Warp 62)	1.63%	9.05	0.94%	6.26	1.47%	8.17
P(RD|GD)	**1.49%**	**8.01**	**1.39%**	**9.26**	**1.44%**	**8.01**

Figure 43 Conditional probability of a rail defect occurring after a geometry defect occurred from Bayes' Theorem

Figure 43 above suggests that over all, geometry defects have a larger impact on the likelihood of the development of a rail defects on curve track compared to tangent track. This was also reconfirmed using the Naïve Bayes method show in Figure 44. Multiple defects on curve track likewise had a much larger impact of the probability of a rail defect on the curve track when compared to the tangent track. Still, tangent track had a higher overall probability then curve track. It should be noted that both Rail Cant and Warp 31 have the largest increase in likelihoods for rail defects for both tangent and curve track.

							TANGENT	CURVE	ALL
	Alignment	Crosslevel	Gage	Rail Cant	Warp 31	Warp 62	P(RD | GD)	P(RD | GD)	P(RD | GD)
# of defects	1	0	0	0	0	0	1.50%	1.37%	1.41%
	0	1	0	0	0	0	1.29%	1.02%	1.24%
	0	0	1	0	0	0	1.34%	1.07%	1.12%
	0	0	0	1	0	0	2.06%	1.51%	1.59%
	0	0	0	0	1	0	2.44%	1.50%	1.57%
	0	0	0	0	0	1	1.63%	0.94%	1.47%
	1	1	0	0	0	0	9.69%	8.61%	9.08%
	1	0	1	0	0	0	10.00%	8.99%	8.23%
	1	0	0	1	0	0	14.69%	12.30%	11.36%
	1	0	0	0	1	0	16.99%	12.20%	11.26%
	1	0	0	0	0	1	11.95%	7.97%	10.59%
	0	1	1	0	0	0	8.71%	6.84%	7.32%
	0	1	0	1	0	0	12.89%	9.45%	10.13%
	0	1	0	0	1	0	14.95%	9.37%	10.05%
	0	1	0	0	0	1	10.44%	6.05%	9.44%
	0	0	1	1	0	0	13.28%	9.85%	9.19%
	0	0	1	0	1	0	15.39%	9.77%	9.11%
	0	0	1	0	0	1	10.76%	6.32%	8.56%
	0	0	0	1	1	0	22.01%	13.33%	12.53%
	0	0	0	1	0	1	15.76%	8.74%	11.79%
	0	1	1	0	0	1	45.91%	31.42%	39.52%
	0	1	0	1	1	0	66.51%	51.09%	50.01%
	1	0	0	1	1	0	69.81%	58.41%	53.19%
	0	1	1	1	0	1	90.56%	82.24%	85.41%
	1	1	0	1	1	0	94.21%	90.51%	88.81%

Figure 44 Naive Bayes analysis of tangent, curve, and all track.

Figures 45 through 48 show the results from the Bayesian Network model to include both the tangent and curve models. The Bayesian Network model results agree

with the previous two analyses showing that geometry defects on tangent track have a larger impact then on curve track in the development of rail defects.

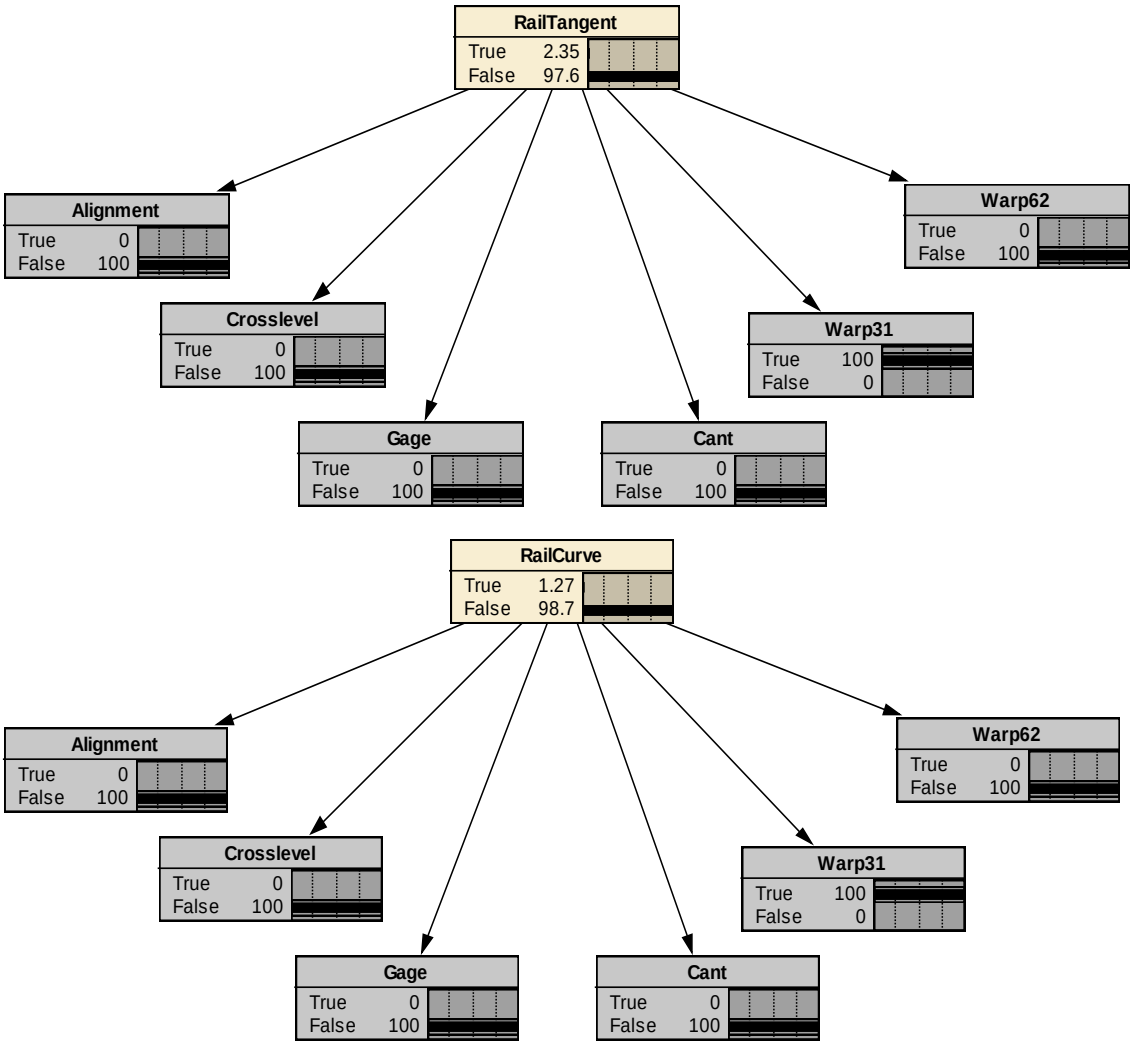


Figure 45 Bayesian Network Model for Warp 31 with Tangent on Top and Curve on Bottom.

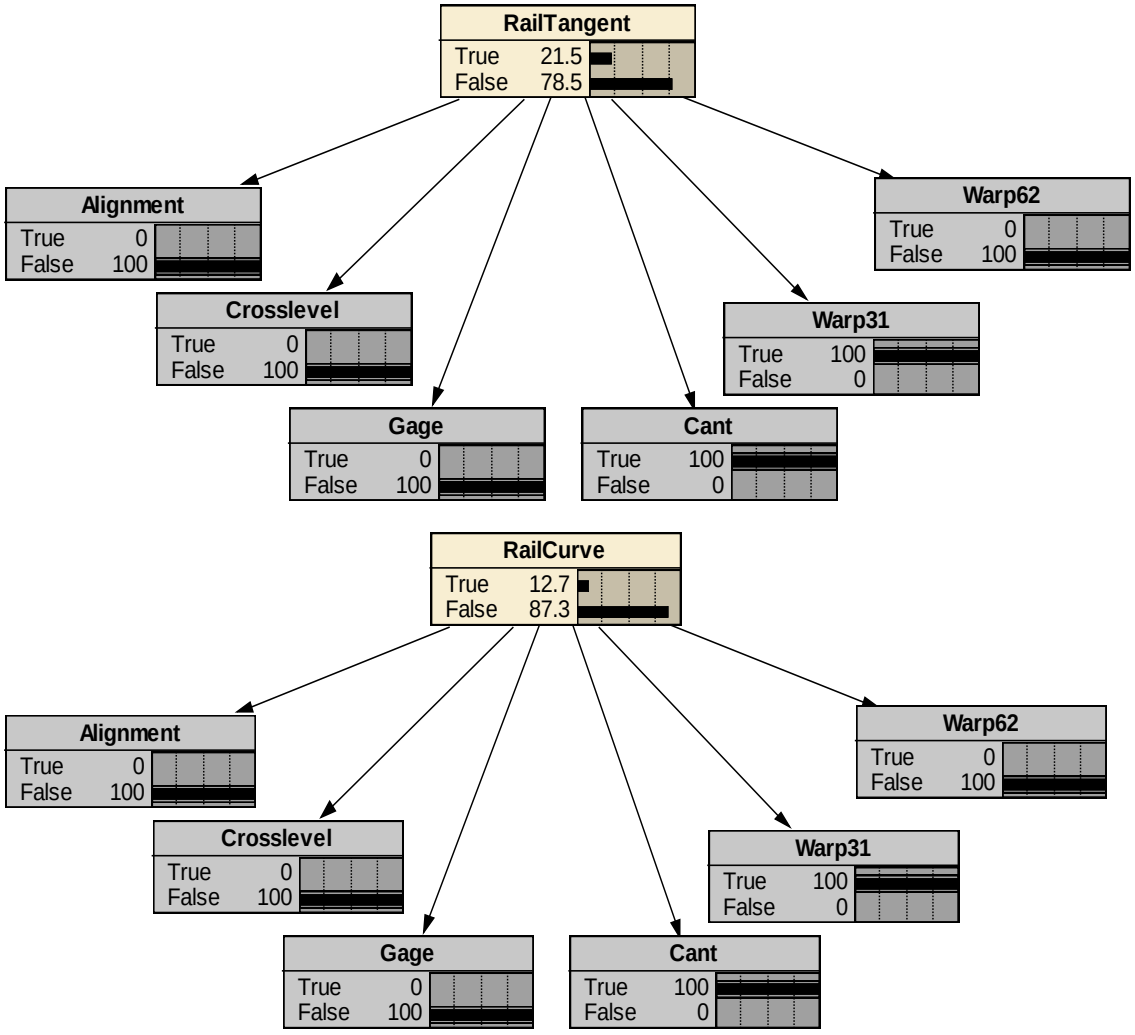


Figure 46 Bayesian Network Model for Warp 31 and Rail Cant with Tangent on Top and Curve on Bottom.

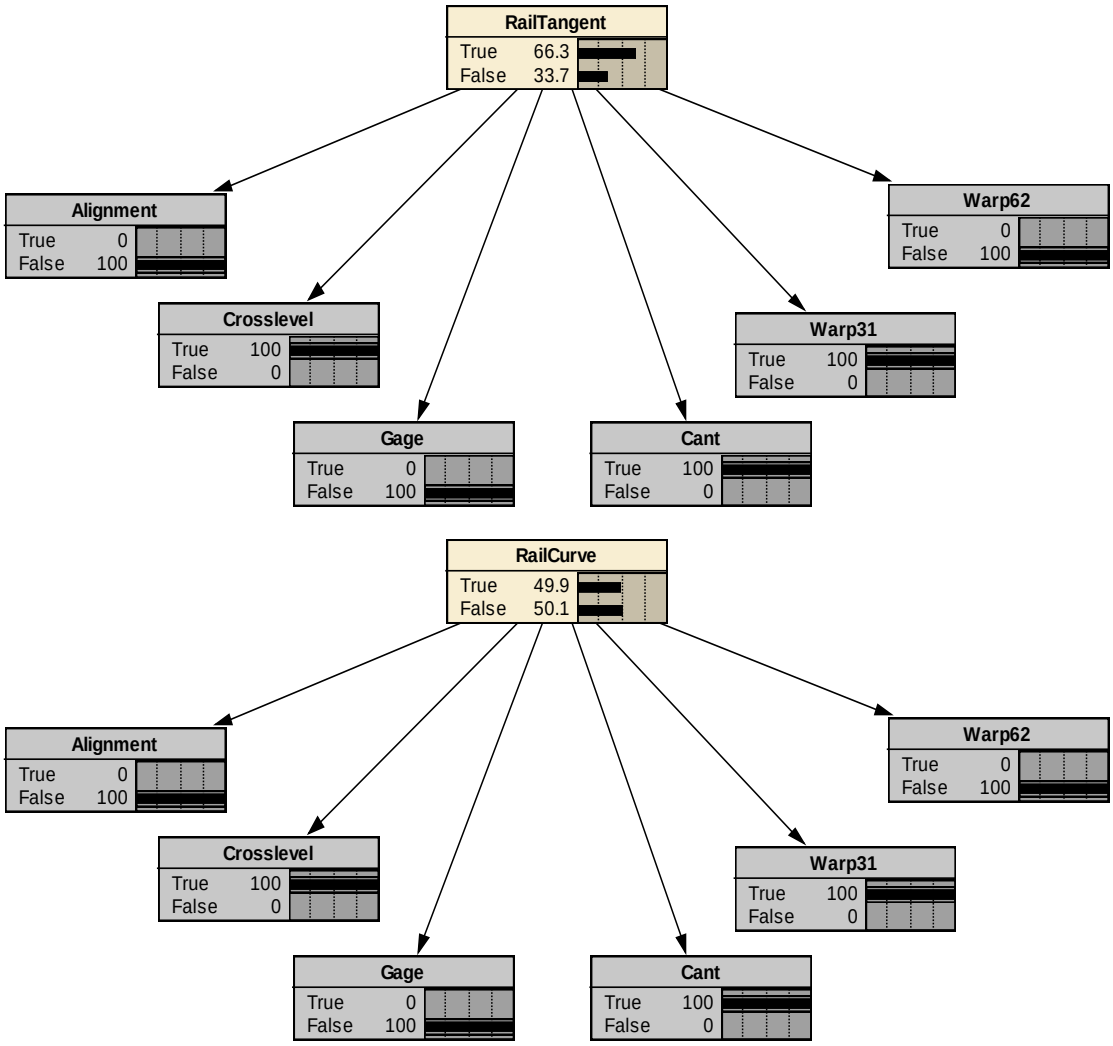


Figure 47 Bayesian Network Model for Warp 31, Rail Cant, and Cross-level with Tangent on Top and Curve on Bottom.

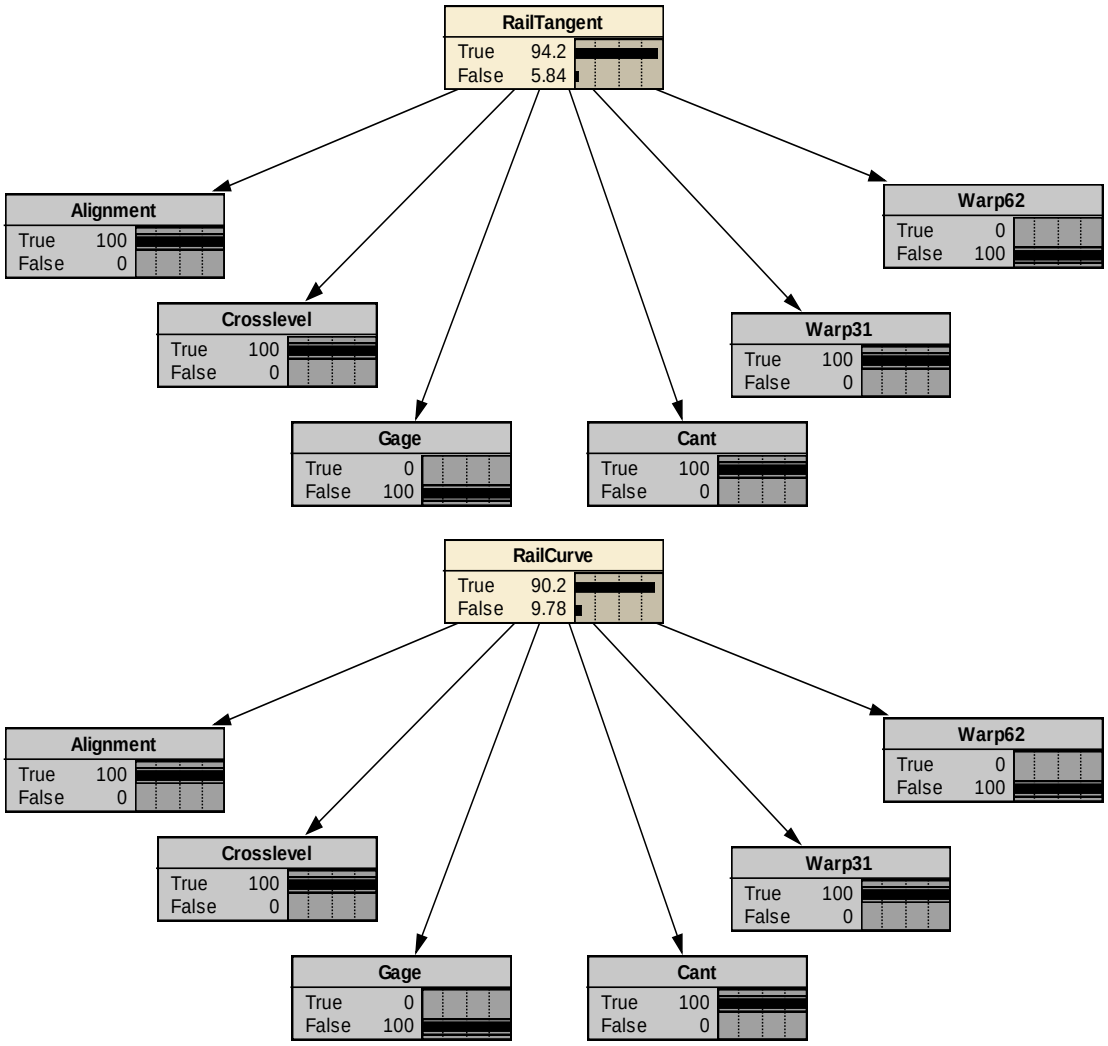


Figure 48 Bayesian Network Model for Warp 31, Rail Cant, Cross-level, and Alignment with Tangent on Top and Curve on Bottom.

Figure 49 below shows the comparison of the results from the Bayesian Network model and the Naïve Bayes model for various geometry defect combinations. Again note the results are close but with a distinct difference. The Bayesian Network model is considered the more "accurate" model because of the way it handles interactions between variables.

							TANGENT		CURVE		ALL	
	Alignment	Crosslevel	Gage	RailCant	Warp31	Warp62	P(RD | GD) Naïve	P(RD|GD) Network	P(RD | GD) Naïve	P(RD|GD) Network	P(RD | GD) Naïve	P(RD|GD) Network
# of defects	1	0	0	0	0	0	1.50%	1.45%	1.37%	1.10%	1.41%	1.30%
	0	1	0	0	0	0	1.29%	1.27%	1.02%	0.82%	1.24%	1.17%
	0	0	1	0	0	0	1.34%	1.29%	1.07%	0.87%	1.12%	1.04%
	0	0	0	1	0	0	2.06%	1.99%	1.51%	1.34%	1.59%	1.51%
	0	0	0	0	1	0	2.44%	2.35%	1.50%	1.27%	1.57%	1.48%
	0	0	0	0	0	1	1.63%	1.58%	0.94%	0.75%	1.47%	1.36%
	1	1	0	0	0	0	9.69%	9.54%	8.61%	7.11%	9.08%	8.54%
	1	0	1	0	0	0	10.00%	9.69%	8.99%	7.52%	8.23%	7.66%
	1	0	0	1	0	0	14.69%	14.30%	12.30%	11.10%	11.36%	10.80%
	1	0	0	0	1	0	16.99%	16.50%	12.20%	10.60%	11.26%	10.60%
	1	0	0	0	0	1	11.95%	11.60%	7.97%	6.51%	10.59%	9.87%
	0	1	1	0	0	0	8.71%	8.59%	6.84%	5.71%	7.32%	6.94%
	0	1	0	1	0	0	12.89%	12.70%	9.45%	8.53%	10.13%	9.82%
	0	1	0	0	1	0	14.95%	14.80%	9.37%	8.14%	10.05%	9.63%
	0	1	0	0	0	1	10.44%	10.30%	6.05%	4.93%	9.44%	8.96%
	0	0	1	1	0	0	13.28%	12.90%	9.85%	9.02%	9.19%	8.82%
	0	0	1	0	1	0	15.39%	15.00%	9.77%	8.61%	9.11%	8.64%
	0	0	1	0	0	1	10.76%	10.50%	6.32%	5.23%	8.56%	8.04%
	0	0	0	1	1	0	22.01%	21.50%	13.33%	12.70%	12.53%	12.10%
	0	0	0	1	0	1	15.76%	15.40%	8.74%	7.83%	11.79%	11.30%
	0	1	1	0	0	1	45.91%	45.70%	31.42%	27.50%	39.52%	38.30%
	0	1	0	1	1	0	66.51%	66.30%	51.09%	49.90%	50.01%	49.60%
	1	0	0	1	1	0	69.81%	69.20%	58.41%	57.30%	53.19%	52.20%
	0	1	1	1	0	1	90.56%	90.50%	82.24%	81.00%	85.41%	85.10%
	1	1	0	1	1	0	94.21%	94.20%	90.51%	90.20%	88.81%	88.60%

Figure 49 Comparison of Naïve Bayes and Bayesian Network Models for Tangent, Curve, and All Track.

As seen in both the Naïve Bayes and Bayesian Network, the presence of multiple geometry defects can be observed to drastically increase the probability of a rail defect occurring up to 80-95%. Figure 50 below shows the increased likelihoods of a rail defect occurring after a geometry defect based off the Bayesian Networks. It is noticed that the likelihood of a rail defect occurring on curve track is reduced much more than on tangent track. This is due to being able to control the independence of the occurrence of geometry defects. In both the Bayes' Theorem and Naïve Bayes analyses, it is assumed that a geometry defect could have occurred even though it was not recorded. This assumption raises the likelihood of a rail defect occurring since multiple geometry defects is shown to increase the probability of a rail defect as indicated in Figure 49. As shown earlier, curve track is much more prone to having repeat geometry defects, making this increase much more noticeable on curve track due to this assumption. By setting the defects to just one defect no matter the situation in the Bayesian Network, a clearer value of the likelihood of a rail defect occurring due to a geometry defect is obtained.

Bayesian Network						
	TANGENT		CURVE		ALL	
	P(RD|GD)	Likelihood more to occur	P(RD|GD)	Likelihood more to occur	P(RD|GD)	Likelihood more to occur
P(RD)	0.186%		0.15%		0.18%	
P(RD|Alignment)	1.45%	7.80	1.10%	7.33	1.30%	7.22
P(RD|Crosslevel)	1.27%	6.83	0.82%	5.47	1.17%	6.50
P(RD|Gage)	1.29%	6.94	0.87%	5.80	1.04%	5.78
P(RD|Rail Cant)	1.99%	10.70	1.34%	8.93	1.51%	8.39
P(RD|Warp 31)	2.35%	12.63	1.27%	8.47	1.48%	8.22
P(RD|Warp 62)	1.58%	8.49	0.75%	5.00	1.36%	7.56
P(RD|GD)	1.44%	7.74	1.10%	7.33	1.31%	7.28

Figure 50 Bayesian Network Probability and Likelihood.

Figure 51 presents a comparison of rail life reduction for tangent and curve track. Note, tangent track shows a reduction in life, when a geometry defect is present, of 28%, twice the life reduction in MGT when compared to curve track (13%). This, may be related to the shorter rail life on curved track with a stronger rail wear effect, but can also be related to this increased probability of a rail defect occurring after a geometry defect on tangent track.

Defect	Match Tangent	Not Matched Tangent	Reduction Tangent	Match Curve	Not Matched Curves	Reduction Curve	Match All	Not Matched All	Reduction All
TOTAL	**503.64**	**703.70**	**-28%**	**413.23**	**477.15**	**-13%**	**446.77**	**646.19**	**-31%**

Figure 51 Average MGT Life of Rail Comparison for Tangent, Curve, and All Track.

From these analyses, it is shown that geometry defects have a significant impact on the development of a rail defect. As it can be seen in Figure 50, geometry defect increases the probability of a rail defect by more than seven times and reduces the life of the rail significantly, as seen in Figure 51. This increase in probability is even higher with the presence of multiple geometry defects, up to a range of 85-95% depending on the geometry defects present

The Bayesian network model can be used to help predict the development of a rail defect given a geometry defect occurred prior. This model is very simple to use and only requires the user to set which geometry defects were true (occurred) and which were false (did not occur). In the future, this model can be extended to other pieces of data, such as Ground Penetrating Radar Data (GPR).

Chapter 5

VALIDATION OF MODELS

Validation of the statistical model was performed using two sets of data while validation of the regression models was done with the original data set. The first set of data was the BA division using two different years of rail data (2013 and 2014). This data represented approximately 5% of the initial data and therefore not the best for validation. Even though it is a small size, it still represents the overall data well since it is using the same exception measurements. The second set of data was provided by BNSF and is approximately 10% of the initial data. This is a far better size for validation than the BA data and helps validate the work done across different exception measurements and track conditions. Figure 52 below has a breakdown of the BNSF subdivision mileage, MGT range, and percent wood tie for a specific MGT range.

Figure 53 presents the breakdown of the defects in all three cases. Defect rates are shown in defects/mile/year. The length in miles for the BNSF is assumed to be 75% tangent and 25% curved rail based on other BNSF data and CSX data. This was seen as a safe assumption. Warp 31 is not reported as such within the BNSF data set and is instead placed into the twist category. The table also contains a comparison of rail life lost. Figures 54 and 55 below have a comparison of the likelihoods and probability of a rail defect occurring after a geometry defect, BA division and BNSF respectively.

Subdivision	Miles of wood ties w/ 20 - 40 MGT	Other miles	% Miles of wood ties w/ 20 - 40 MGT
MENDOTA	220	20	0.92
LAMPASAS	239	0	1
DOUGLASS	30	0	0.99
SLATON	131	78	0.63
HASTINGS	173	54	0.76
BELLINGHAM	44	41	0.51
COLUMBIA RIVER	122	68	0.64
SCENIC	102	82	0.55
SEATTLE	188	161	0.54
AFTON	47	37	0.56
CHEROKEE	102	102	0.5
CREEK	162	12	0.93
CASCO	49	25	0.66
HILLSBORO	74		1
HINCKLEY	124		1
MARSHALL	222		1
WAYZATA	87	0	1

Figure 52 BNSF Subdivision Breakdown.

		Tangent	Curve	All	
All Div	Length in Miles	15725.12	6502.6	22227.72	
BA Val		580	462	1042	
BNSF		2102.805	700.935	2803.74	
All Div	Unique Geo	70287	132054	202341	
BA Val		2442	5262	7704	
BNSF		17898	16221	34119	
All Div	Unique Geo Rate Def/M/Y	0.89	4.06	1.82	
BA Val		0.84	2.28	1.48	
BNSF		1.70	4.63	2.43	
All Div	Rail Defects	19761	6679	26440	
BA Val		880	372	1252	
BNSF		1664	139	1803	
All Div	Rail Defects Rate Def/M/Y	0.50	0.41	0.48	
BA Val		0.76	0.40	0.60	
BNSF		0.79	0.20	0.64	
All Div	Matches	1047	1871	2918	
BA Val		43	33	76	
BNSF		189	57	246	
All Div	Matches Rate Def/M/Y	0.027	0.115	0.053	
BA Val		0.037	0.036	0.036	
BNSF		0.090	0.081	0.088	
All Div	% of Defects Matched	5.3%	28.0%	11.0%	
BA Val		4.9%	8.9%	6.1%	
BNSF		11.4%	41.0%	13.6%	
All Div	Unique Gage Defects	2952	19414	22366	
BA Val		47	811	858	
BNSF		9937	12513	22450	
All Div	Unique Warp 31 Defects	1762	31017	32779	
BA Val		70	1651	1721	
BNSF		174	50	224	
All Div	Unique Warp 62 Defects	10451	3693	14144	
BA Val		274	119	393	
BNSF		351	341	692	
All Div	Unique Crosslevel Defects	34529	10657	45186	
BA Val		1382	425	1807	
BNSF		3456	885	4341	
All Div	Unique Alignment Defects	1312	6629	7941	
BA Val		32	170	202	
BNSF		137	664	801	Below is % Matched for specific Geo Defects (Matches for a geo/Rail Defects)
All Div	Unique Cant Defects	5139	49293	54432	
BA Val		202	1415	1617	
BNSF		786	874	1660	
All Div	Matched Gage Defects	40	210	250	0.9%
BA Val		0	2	2	0.2%
BNSF		83	43	126	7.0%
All Div	Matched Warp 31 Defects	44	472	516	2.0%
BA Val		1	11	12	1.0%
BNSF		2	1	3	0.2%
All Div	Matched Warp 62 Defects	173	35	208	0.8%
BA Val		11	0	11	0.9%
BNSF		3	0	3	0.2%

All Div	Matched	452	110	562	2.1%
BA Val	Crosslevel	18	4	22	1.8%
BNSF	Defects	48	4	52	2.9%
All Div	Matched	20	92	112	0.4%
BA Val	Alignment	0	1	1	0.1%
BNSF	Defects	3	4	7	0.4%
All Div	Matched	108	757	865	3.3%
BA Val	Cant Defects	8	13	21	1.7%
BNSF		6	3	9	0.5%
All Div	Unique Gage	0.038	0.597	0.201	
BA Val	Defects Rate	0.016	0.351	0.165	
BNSF	Def/M/Y	0.945	3.570	1.601	
All Div	Unique Warp	0.022	0.954	0.295	
BA Val	31 Defects	0.024	0.715	0.330	
BNSF	Rate Def/M/Y	0.017	0.014	0.016	
All Div	Unique Warp	0.133	0.114	0.127	
BA Val	62 Defects	0.094	0.052	0.075	
BNSF	Rate Def/M/Y	0.033	0.097	0.049	
All Div	Unique	0.439	0.328	0.407	
BA Val	Crosslevel	0.477	0.184	0.347	
BNSF	Defects Rate Def/M/Y	0.329	0.253	0.310	
All Div	Unique	0.017	0.204	0.071	
BA Val	Alignment	0.011	0.074	0.039	
BNSF	Defects Rate Def/M/Y	0.013	0.189	0.057	
All Div	Unique Cant	0.065	1.516	0.490	
BA Val	Defects Rate	0.070	0.613	0.310	
BNSF	Def/M/Y	0.075	0.249	0.118	
All Div	Not Matched MGT	703.7	477.15	646.19	
BA Val		570.55	490.429	546.39	
BNSF		685.41	536.02	679.08	
All Div	Matched MGT	503.64	413.23	446.77	
BA Val		386.84	359.36	368.52	
BNSF		653.65	499.47	617.78	
All Div	Loss of MGT	28.4%	13.4%	30.9%	
BA Val		32.2%	26.7%	32.6%	
BNSF		4.6%	6.8%	9.0%	

Figure 53 Breakdown of All Three Data Sets.

Bayes Comparision												
	TANGENT				**CURVE**				**ALL**			
	P(RD|GD)	Likelihood more to occur ALL	P(RD|GD) BA Val	Likelihood more to occur BA	P(RD|GD	Likelihood more to occur ALL	P(RD|GD) Val	Likelihood more to occur	P(RD|GD)	Likelihood more to occur ALL	P(RD|GD) Val	Likelihood more to occur
P(RD)	0.186%		0.352%		0.150%		0.119%		0.180%		0.222%	
P(RD|Alignment)	1.50%	8.34	0.00%	0.00	1.37%	9.13	0.73%	6.19	1.41%	7.84	0.62%	2.78
P(RD|Crosslevel)	1.29%	7.18	1.62%	4.60	1.02%	6.81	1.17%	9.89	1.24%	6.91	1.52%	6.83
P(RD|Gage)	1.34%	7.43	0.00%	0.00	1.07%	7.13	0.31%	2.60	1.12%	6.21	0.29%	1.31
P(RD|Rail Cant)	2.06%	11.44	4.90%	13.91	1.51%	10.08	1.15%	9.66	1.59%	8.83	1.62%	7.28
P(RD|Warp 31)	2.44%	13.54	1.78%	5.05	1.50%	9.99	0.83%	7.01	1.57%	8.75	0.87%	3.91
P(RD|Warp 62)	1.63%	9.05	4.97%	14.09	0.94%	6.26	0.00%	0.00	1.47%	8.17	3.47%	15.63
P(RD|GD)	**1.49%**	**8.01**	**2.28%**	**6.46**	**1.39%**	**9.26**	**0.84%**	**7.10**	**1.44%**	**8.01**	**1.30%**	**5.87**

Figure 54 BA Validation Comparison.

Bayes Comparision												
	TANGENT				**CURVE**				**ALL**			
	P(RD|GD)	Likelihood more to occur ALL	P(RD|GD) BNSF Val	Likelihood more to occur BNSF	P(RD|GD)	Likelihood more to occur ALL	P(RD|GD) BNSF Val	Likelihood more to occur BNSF	P(RD|GD)	Likelihood more to occur ALL	P(RD|GD) BNSF Val	Likelihood more to occur BNSF
P(RD)	0.186%		0.308%		0.150%		0.073%		0.180%		0.251%	
P(RD|Alignment)	1.50%	8.34	5.30%	18.14	1.37%	9.13	1.49%	20.38	1.41%	7.84	0.31%	1.31
P(RD|Crosslevel)	1.29%	7.18	3.40%	11.64	1.02%	6.81	1.12%	15.32	1.24%	6.91	2.43%	10.23
P(RD|Gage)	1.34%	7.43	2.06%	7.06	1.07%	7.13	0.85%	11.67	1.12%	6.21	2.02%	8.50
P(RD|Rail Cant)	2.06%	11.44	1.89%	6.46	1.51%	10.08	0.85%	11.66	1.59%	8.83	1.50%	6.32
P(RD|Warp 31)	2.44%	13.54	2.82%	9.67	1.50%	9.99	4.85%	66.28	1.57%	8.75	1.18%	4.98
P(RD|Warp 62)	1.63%	9.05	2.11%	7.22	0.94%	6.26	0.00%	0.00	1.47%	8.17	1.80%	7.56
P(RD|GD)	**1.49%**	**8.01**	**2.41%**	**8.24**	**1.39%**	**9.26**	**0.89%**	**12.18**	**1.44%**	**8.01**	**2.00%**	**8.40**

Figure 55 BNSF Validation Comparison.

The comparison between the BA validation data and the CSX data, shown in Figures 53 and 54, show similar trends in increase likelihood of a rail defect occurring and decrease life of the rail when a rail defect occurs after a geometry defect. When looking at the specific of the increase likelihoods, it can be seen that some geometry defects have a zero in them. This is due to having no matches, which is a result in of the too small sample size. Because of this, looking at specific geometry defects can be somewhat misleading and the all case is what should be observed. In the BA validation case, it is seen that the likelihood of a geometry defect occurring is between 5.87 and 7.1, with curve being the highest. This is slightly lower than the CSX data set, which is from 8 to 9.26 increase likelihood with curve being the highest. Though different it follows the same trends. This difference in likelihoods can be explained with the small sample size and that some geometry defect rates vary greatly.

The comparison between the BNSF validation data and the CSX data, shown in Figures 53 and 54, also show similar trends in increased likelihood and, to a lesser degree, loss of rail life. The rail life loss is a lot less in the BNSF data set than the CSX data set. Even though it is lower, there is still a significant loss of life (40 to 60 MGT). This difference in loss of rail life might be explained by the BNSF data that was used. It was limited to 20-40 annual MGT and had less variance when compared to the CSX data which had a much larger range of annual MGT. The likelihoods for the all geometry cases are very similar for the BNSF and CSX data. The curve data is somewhat skewed by the Warp 31 (twist for BNSF) which can be explained by the fact the two defects being reported are similar but not exactly the same.

Comparing both validation sets of data (BNSF and CSX BA Division) with the CSX data set, it can be seen that the likelihood of a rail defect occurring after a

geometry defect, as defined by the Bayesian analyses, show similar relationships, with all the probability of a rail defect occurring after a geometry defect increasing by approximately a factor of 8 for all three analyses (base and two validations) . Since the random probability of a rail defect differed between track types and track location, the probability of a rail defect following a geometry defect also varied. However, the comparable increase in risk (a factor of 8) strongly supports the validity of this analysis.

The multilinear regression models validity were determined by a k-fold cross-validation method applied to the full CSX data set. This method removes values from the learned data set, refits the model, and then predicts the removed values. Once the removed values are predicted, an error term is made for the model. This error term is shown in the same unit as the predicted value (MGT). These error terms show how well the multilinear model fits. Figure 56 below has the associated error terms from the validation of the five main multilinear equations for rail life reduction in terms of MGT.

K-Fold Cross-Validation	
	SE of Estimation in MGT
Tangent All	76.19
Tangent	73.16
Curve All	44.69
Curve	44.24
ALL TRACK	52.63

Figure 56 Standard Error of Estimation in MGT

The k-fold validation shows that the overall fitness for the curve and all track is adequate but that the overall fitness of the tangent track needs improvement. To improve these error terms the lead coefficients of the equations were adjusted so that if no geometry defect is reported, the equation would equal the average reported MGT of a not matched rail defect. The adjusted lead coefficients for each equation and its new error term can be seen below in Figure 57.

K-Fold Cross-Validation		
	Lead Coefficient in MGT	SE of Estimation in MGT
Tanget All	952	41.46
Tangent	983	41.11
Curve All	614	15.67
Curve	619	15.66
ALL TRACK	824	24.79

Figure 57 Standard Error of Estimation with New Lead Coefficient.

With the new coefficient, the standard error is reduced making the model fit the data much better. Tangent track is still the least fit, even though its overall fitness was increased. The curve and all track models fitness was measurably increased with this change. Since tangent track consists of more track, the difference between MGT values will be greater. This explains the larger standard error for the tangent track, and to a less extent, the all track.

Using these revised lead coefficients results in the use average rail life for tangent, curve and all track as opposed to the maximum rail life of 1750 MGT used

before. The resulting regression equations and sensitivity figures that correspond to these revised coefficients are as follows: It should be noted that the average decrease of rail defect life while using one of the six key geometry defects is approximately 30%.

Rail life for Tangent Track using six key geometry variables as identified by MARS
Equation 25

MGT (Tangent) =952 (0.739+$\Sigma a_i GD_i$)

GD_i	a_i
Alignment	+0.0198
Cant	-0.4576
Cross-level	-0.2583
Warp 31	-0.1579
Warp 62	-0.3583
Gage	-0.2583

Rail life for Tangent Track using four key geometry variables as identified by MARS
Equation 26

MGT (Tangent) =983 (0.7156+$\Sigma a_i GD_i$)

GD_i	a_i
Cant	-0.4902
Cross-level	-0.2909
Warp 62	-0.4902
Gage	-0.2249

Rail life for Curve Track using six key geometry variables as identified by MARS Equation 27

MGT (Curve) =614(0.7777+$\Sigma a_i GD_i$)

GD_i	a_i
Alignment	-0.1985
Cant	-0.1342
Cross-level	-0.1949
Warp 31	-0.4987
Warp 62	-0.0162
Gage	-0.0813

Rail life for Curve Track using four key geometry variables as identified by MARS Equation 28

MGT=619 (0.7704+$\Sigma a_i GD_i$)

GD_i	a_i
Alignment	-0.1661
Cant	-0.2211
Cross-level	-0.1661
Warp 31	-0.5193

Rail life for All Track using six key geometry variables as identified by MARS Equation 29

MGT=824(0.784+$\Sigma a_i GD_i$)

GD_i	a_i
Alignment	-0.2305
Cant	-0.2566
Cross-level	-0.291
Warp 31	-0.3808
Warp 62	-0.1255
Gage	-0.0012

Tangent All	952.23							
Intercept	0.739							
Alignment	0.0198	0	1	0	0	0	0	0
Crosslevel	-0.2582	0	0	1	0	0	0	0
Gage	-0.2583	0	0	0	1	0	0	0
Rail Cant	-0.4576	0	0	0	0	1	0	0
Warp 31	-0.1579	0	0	0	0	0	1	0
Warp 62	-0.3583	0	0	0	0	0	0	1
	MGT	703.7	722.5542	457.8335	457.7383	267.9583	553.3424	362.515
Reduction in Life			2.68%	-34.94%	-34.95%	-61.92%	-21.37%	-48.48%

Figure 58 Rail Life for Tangent Track using Six Key Geometry Variables with New Coefficient.

Tangent Specific	983.37					
Intercept	0.7156					
Gage	-0.2249	0	1	0	0	0
Crosslevel	-0.2909	0	0	1	0	0
Rail Cant	-0.4902	0	0	0	1	0
Warp 62	-0.4239	0	0	0	0	1
	MGT	703.7	482.54	417.6375	221.6517	286.8492
Reduction in Life			-40.65%	-68.50%	-59.24%	-39.96%

Figure 59 Rail life for Tangent Track using Four Key Geometry Variables with New Coefficient.

Curve All	613.5399							
Intercept	0.7777							
Alignment	-0.1985	0	1	0	0	0	0	0
Crosslevel	-0.1949	0	0	1	0	0	0	0
Gage	-0.0813	0	0	0	1	0	0	0
Rail Cant	-0.1342	0	0	0	0	1	0	0
Warp 31	-0.4987	0	0	0	0	0	1	0
Warp 62	-0.01618	0	0	0	0	0	0	1
	MGT	477.15	355.3623	357.5711	427.2692	394.8129	171.1776	467.2229
Reduction in Life			-25.52%	-25.06%	-10.45%	-17.26%	-64.12%	-2.08%

Figure 60 Rail Life for Curve Track using Six Key Geometry Variables with New Coefficient.

Curve Specific	619.3536					
Intercept	0.7704					
Alignment	-0.16614	0	1	0	0	0
Crosslevel	-0.2747	0	0	1	0	0
Rail Cant	-0.2211	0	0	0	1	0
Warp 31	-0.5193	0	0	0	0	1
	MGT	477.15	374.2506	307.0136	340.2109	155.5197
Reduction in Life			-21.57%	-35.66%	-28.70%	-67.41%

Figure 61 Rail Life for Curve Track using Four Key Geometry Variables with New Coefficient.

All Track	824.2325							
Intercept	0.78399							
Alignment	-0.2305	0	1	0	0	0	0	0
Crosslevel	-0.291	0	0	1	0	0	0	0
Gage	-0.0012	0	0	0	1	0	0	0
Rail Cant	-0.2566	0	0	0	0	1	0	0
Warp 31	-0.3808	0	0	0	0	0	1	0
Warp 62	-0.1255	0	0	0	0	0	0	1
	MGT	646.19	456.2044	406.3384	645.2009	434.692	332.3223	542.7488
Reduction in Life			-29.40%	-37.12%	-0.15%	-32.73%	-48.57%	-16.01%

Figure 62 Rail Life for All Track using Six Key Geometry Variables with New Coefficient.

Chapter 6

SUMMARY

The initial results from the correlation analysis for the whole CSX system, which is approximately 22,000 miles of track with an average annual tonnage of 21.3 MGT, showed that there was a significant relationship between geometry defects and the occurrence of rail defects. Overall, approximately 11% of all rail defects were matched to at least one preceding geometry defect, with 38% of these matches (4.2% of total rail defects) have two or more geometry defects preceding it.

When this correlation was done for curved track, the amount of rail defects matched to a preceding geometry defect was increased to 21%, with 46% of these matches having two or more geometry defects.

These same correlation analyses were performed on the high traffic density tracks (annual MGT > 19.5). This consisted of approximately 10,600 miles of track with an annual tonnage of 36 MGT. Looking at the matched rail defects to a preceding geometry defect, there was a slight increase in percent of rail defects match to 12%. Approximately 38% of these matches had repeat geometry defects. The same was done again for curved track. Curve tracks percent match stayed the same at 21% where the amount of those matches that were repeats dropped slightly to 43%.

These correlations indicate that there is a correlation between geometry defects and the occurrence of a rail defect after the geometry defect. The correlation analyses on the high traffic density lines shows that geometry defects on high traffic density tracks have a similar relationship with rail defects.

The next set of correlations was performed between specific geometry defects and rail defects. From this correlation, it was observed that a large portion of all matched rail defects were matched with either a rail cant defect or a warp defect of some sort. Also in this initial correlation it was shown that detail fracture defects (TDD) made up a large majority of the rail defects. From this initial correlation, a compressed version of geometry and rail defects was created. This combined geometry defects of the same type (alignment left and alignment right- combined into alignment) and removed low occurring rail defects (such as pipe defects). With this compressed comparison and with the addition of the cumulative MGT of a matched rail defect, a cross-correlation was performed. This cross-correlation showed that there was a strong negative correlation between the presence of a geometry defect and the cumulative life (in MGT) of a rail defect, which indicates a rail life reduction. With these relationships in mind, several analyses were performed, first on the effects of geometry defects on rail life.

A series of different analyses were performed on the effects of geometry defects have on the life of a rail defect. The first analysis showed that a rail defect that was matched to a geometry defect preceding it had approximately 30% less life (in MGT) then a rail defect that had no geometry defects prior to it. Based on this first analysis of rail life two different regressions were performed to further determine the effects of geometry defects on rail defect life and to determine which geometry defects where the most crucial.

The first regression analysis performed was a general multilinear regression analysis. This analysis showed the importance of warp, alignment, and rail cant defects. To get a better understanding of the effects of specific geometry defects on the

life of a rail defect, a MARS analysis was performed. The MARS analysis was performed on tangent and curve track, to further refine the importance of certain geometry defects. The MARS analysis showed the following geometry defects as key defects in determining the life of a rail defect.

For Tangent:

- Crosslevel
- Gage
- Rail Cant
- Warp 62

For Curve:

- Alignment
- Crosslevel
- Rail Cant
- Warp 31

With these key geometry defects in mind, a new set of multilinear regression analyses were performed. Tables to show the sensitivity of each geometry defect were then produced for each regression equation developed. With the key geometry defects that affect the occurrence of a rail defect determined, probability analyses to determine the increased likelihood of finding a rail defect after a geometry defect were performed.

Various Bayesian statistical analyses were performed to determine the increase likelihoods of a rail defect occurring given the presence of a geometry defect prior. First was determining the random probability of a rail defect occurring. For all track this value was 0.18%, for tangent it was 0.186%, and for curve it was 0.15%. After

determining the random probability of the rail defects, Bayes' Theorem was applied to the data. This showed a serious increase in likelihood of a rail defect occurring. This increase ranged from 6 times to 13 times more likely to occur. If the presence of a geometry defect and rail defect were independent from each other, there would be no noticeable increase in likelihood. Since repeat geometry defects are of some interest, further analyses were performed to determine the effects of multiple geometry defects.

Naïve Bayes and Bayesian Networks were applied to the data to get an understanding of the impact in probability that multiple geometry defects have on the development of a rail defect. With the presence of multiple geometry defect prior to a rail defect, the likelihood a rail defect increased drastically, in some case to increase of over 600 increasing the probability of a rail defect to approximately 90% in the case of 4 key geometry defects occurring. This showed the importance of multiple geometry defects in the development of a rail defect, even though it did not greatly affect the MGT life of the defect.

These statistical analyses were validated with the use of more recent BA division data from CSX and data provided by BNSF. The CSX data was slightly too small for true validation purposes (5% of actual data set) but still provided good results since it used the same measurements as the original data. The BNSF data was of appropriate size (approximately 10% of the data) but used different measurements and naming convention compared to the original data. This led to only looking at the overall data. In both cases of the data, the overall trends were similar with roughly all three sets of data having around an increased likelihood of a rail defect by 8.

The multilinear analyses were validated with the use of a k-fold validation, the standard validation method used for multilinear regression. After using this validation

method it was noted that the error terms were slightly higher than desired. With this in mind, the lead coefficients in the multilinear analyses were changed so that with zero geometry defects present, the rail life would be the “average” reported rail life of a non-matched rail defect. After this change, the error terms were reduced to more acceptable levels.

In conclusion, it can be said that there is a significant relationship between the presence of a geometry defect and the occurrence of a rail defect following the geometry defect.

Chapter 7

FUTURE WORK

There are several recommendations for future work that can be made. One recommendation is to further extend the Bayesian Network model to include additional parameters. Such parameters include, but are not limited to, Ground Penetrating Radar (GPR) data, degree of curvature of track, track class, track tonnage, ballast information, and traffic type. These can be added as more child nodes, with their parent nodes being the geometry defects.

Another recommendation that can be made for future research is to expand the geometry data to beyond just exception reports. By examining the whole geometry history of the rail at the location of a rail defect, a similar set of analyses can be performed. From these analyses, a new set of geometry errors can be found that have a large impact of the development of rail defects, even though these parameters do not create a known geometry exception. Lastly, these approaches can be used to relate other parameters to rail defects, such as rail wear or ballast condition.

REFERENCES

1. Winkler, E. *Der Lehre von der Elasticitat und Festigkeit* [Elasticity and Strength], Verlag von Dominicus, Prague 1867.
2. Hetenyi, M. *Beams on Elastic Foundation*, University of Michigan Press, 1947
3. Hay, W. W. (1982). *Railroad Engineering*. 2nd Ed. John Wiley and Sons.
4. Kerr, A.D. (2004). Fundamentals of Railway Track Engineering. Omaha, Nebraska: Simmons-Boardman Press, Inc.
5. Zarembski, A. M., & Abbott, R. A., "Fatigue Analysis of Rail Subject to Traffic and Temperature Loading", Heavy Hauls Railways Conference, Perth, Western Australia, September 1978.
6. Steele, R.K., and R.P. Reiff (1981). "Rail: Its Behavior and Relationship to Total System Wear." FAST Engineering Conference (November). Denver Colorado.
7. Jenkins, H. H., Stephenson, J. E., Clayton, G. A., Morland, G. W. and Lyon, D., "The Effect of Track and Vehicle Parameters on Wheel/Rail Vertical Dynamic Forces", *Railway Engineering Journal*, January 1974.
8. Ahlbeck, D. R., "An Investigation of Impact Loads Due to Wheel Flats and Rail Joints", *American Society of Mechanical Engineers*, 80-WA/RT-1, 1980.
9. Zarembski, A.M., The Art and Science of Rail Grinding, Simmons-Boardman Books, Inc**.,** Omaha, NE, August 2005
10. Zarembski, A.M., "Engineering of Railway Tracks", Kerr Symposium on Engineering Mechanics, University of Delaware College of Engineering, Newark, DE, April 29-30, 2004.
11. Friedman, J. H., "Multivariate Adaptive Regression Splines", Annals of Statistics, March 1991, 19(1), pp. 1-67
12. Sephton, P., "Forecasting Recessions: Can We Do Better on MARS™?", Federal Reserve Bank of St. Louis Review, 83 (2), 39-50
13. Attoh-Okine, N. O., Cooger, K., and Mensah, S. "Multivariate Adaptive Regression (MARS) and Hinged Hyperplanes (HHP) for Doweled Pavement Performance Modeling," Journal of Construction and Building Materials , Vol 23, pp 3020-3023, 2009
14. Bolstad, W.M., Introduction to Bayesian Statistics, 2nd Ed. John Wiley and Sons, Inc., Hoboken, NJ, 2007
15. Lewis, D.D., "Naïve (Bayes) at Forty: The Independence Assumption in Information Retrieval", Proc. 10th European Conference on Machine Learning (ECML98), 1998
16. Androutsopoulos, I., Koutsias, J., Chandrinos, K.V., Spyropoulos, C.D,"An Experimental Comparison of Naïve Bayesian and Keyword-Based Anti-Spam

Filtering with Personal E-mail Messages", *Proceedings of the 23rd Annual International ACM SIGIR Confrence on Research and Development in Information Retrieval*, 2000

17. Castellettis, A., Soncini-Sessa, R.,"Bayesian Networks and Participatory modeling in water resource management", *Elsevier Journal of Environmental Modeling and Software,* Volume 22, Issue 8, August 2007, pages 1075-1088
18. Bielza, C., Larranaga P. (2014), "Bayesian Networks in Neuroscience: a survey", Frontiers in Computational Neuroscience

Appendix A

RAW DATA FORMAT SAMPLES

A1 Sample Rail Defect Data

DIVISION	SUBDIVISION	PREFIX	MILEPOST	TRACK TYPE	TRACK CODE	SIDE	DEFECT TYPE	SIZE	DATE FOUND	CURVE-TANG	ROLLED YEAR	MILL	WEIGHT
	JAMES	CA	185.7	SG	M	R	TD D	1 0	09/03/2010	T	1988	T HOT	136
HU	RIVER	B	3										
	JAMES	CA	205.3	SG	M	R	TD D	5 0	11/01/2010	H	2001	T HOT	141
HU	RIVER	B	25										
	JAMES	CA	209.4	SG	M	L	TD D	4 0	11/03/2010	H	1989	T HOT	132
HU	RIVER	B	213.7	SG	M	L	BB		11/07/2010	T	1988	T HOT	132
	MONTR		187.				BH		11/09/			B	
AY	EAL	QM	45	SG	M	R	B	4	2010	T	1943	L	100
	M AND		553.				OA	4	09/01/			U	
AT	M	000	92	SG	M	L	W	0	2010	T	1971	T	132
	NORTH		541.				TD	1	09/01/				
HU	ERN	CA	89	1	M	L	D	0	2010	T	1979	UI	122
	NORTH		540.				TD	1	09/01/			K	
HU	ERN	CA	97	1	M	R	D	0	2010	T	1973	L	122
	NORTH		538.					1	09/01/				
HU	ERN	CA	83	1	M	R	CH	2	2010	T	1973	UI	122
	NORTH		536.						09/01/				
HU	ERN	CA	81	1	M	R	EB	0	2010	T	1974	UI	122
	NORTH		535.				TD	1	09/01/				
HU	ERN	CA	51	1	M	R	D	0	2010	T	1974	UI	122
	NORTH		534.						09/01/				
HU	ERN	CA	53	1	M	L	EB	0	2010	H	1974	UI	122
	METRO												
	POLITA		78.5				TD	2	09/01/			U	
BA	N	BA	4	2	M	R	D	5	2010	H	1948	C	140

Table A2: Sample Geometry Defect Data (CSX) [partial 1 of 3]

PREFIX	LOCATION MP	LOCATION OFFSET	EXCEPTION DATE	EXCEPTION TYPE	CLASS_EX_C	CLASS_SV_C	LENGTH	MAX VAL
000	3.00	-2,250	3/19/2008	WIDE GAGE	2	1	1	1
000	3.00	-2,256	3/19/2008	WIDE GAGE	2	1	2	1
000	3.00	-2,247	3/19/2008	ALIGNMENT RIGHT	2	1	1	2
000	3.00	-2,240	3/19/2008	ALIGNMENT RIGHT	2	1	2	3
000	3.00	-2,200	3/19/2008	CROSSLEVEL	2	1	3	1
000	3.00	-2,184	3/19/2008	WARP 31FT	2	0	41	2
000	4.00	-4,954	2/5/2008	CROSSLEVEL	2	1	1	-2
000	4.00	-4,954	2/5/2008	CROSSLEVEL	2	1	2	-2
000	4.00	-4,954	2/5/2008	PROFILE LEFT 62FT	2	1	1	-2
000	4.00	-4,991	2/5/2008	WARP 62FT	2	1	12	2
000	4.00	-4,971	2/5/2008	CROSSLEVEL	2	1	1	-2
000	4.00	-4,971	2/5/2008	PROFILE LEFT 62FT	2	1	1	-3
000	6.00	-896	2/5/2008	TIGHT GAGE	2	0	2	-1
000	7.00	-360	2/4/2008	TIGHT GAGE	1	0	1	-1
000	8.00	-4,802	2/4/2008	TIGHT GAGE	1	0	2	-1
000	8.00	-4,568	2/5/2008	CROSSLEVEL	4	3	1	1
000	8.00	-3,882	2/4/2008	TIGHT GAGE	1	0	10	-1
000	8.00	-3,639	2/4/2008	CLIM	1	0	72	0
000	8.00	-3,603	2/4/2008	CLIM	1	0	36	0
000	8.00	-3,541	2/4/2008	CLIM	1	0	106	0
000	8.00	-3,481	2/4/2008	CLIM	1	0	33	0
000	8.00	-3,463	2/4/2008	CLIM	1	0	109	0

Table A2: Sample Geometry Defect Data (CSX) [partial 2 of 3]

BEGINMP	BEGINOFFSET	ENDMP	END_OFFSET_FEET_I	CURVE	FRIEGHT_MPH_Q	PASSENGER_MPH_Q	TRACK	ACTIONDATE	AUTHORIZEDINSPECTOR	APPROVEDDATE	APPROVEDBY	LATITUDE	LONGITUDE
3.00	-2,250	3.00	-2,251	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21036400000	85.75693900000
3.00	-2,256	3.00	-2,258	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21037700000	85.75692700000
3.00	-2,247	3.00	-2,248	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21035900000	85.75694300000
3.00	-2,240	3.00	-2,242	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21034100000	85.75695900000
3.00	-2,199	3.00	-2,202	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21025300000	85.75704200000
3.00	-2,174	3.00	-2,214	C	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.21021800000	85.75707600000
4.00	-4,954	4.00	-4,955	T	20	20	2	10/23/2008	C7048	10/23/2008	C7048	38.20252300000	85.76041900000
4.00	-4,954	4.00	-4,956	T	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.20252300000	85.76041900000
4.00	-4,954	4.00	-4,955	T	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.20252300000	85.76041900000
4.00	-4,991	4.00	-5,003	T	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.20262400000	85.76044600000
4.00	-4,971	4.00	-4,972	T	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.20257000000	85.76043200000
4.00	-4,971	4.00	-4,972	T	20	20	2	7/8/2008	C7048	7/8/2008	C7048	38.20257000000	85.76043200000

Table A2: Sample Geometry Defect Data (CSX) [partial 3 of 3]

LATITUDE	LONGITUDE	ENGINEER MILEPOST	DIVISION	SUBDIVISION	PRIORITY	EXCEPTION ACTION
38.21036400000	85.75693900000	2.57	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.21037700000	85.75692700000	2.57	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.21035900000	85.75694300000	2.57	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.21034100000	85.75695900000	2.58	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.21025300000	85.75704200000	2.58	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.21021800000	85.75707600000	2.59	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.20252300000	85.76041900000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	CRT	Repaired
38.20252300000	85.76041900000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.20252300000	85.76041900000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.20262400000	85.76044600000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.20257000000	85.76043200000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.20257000000	85.76043200000	3.04	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Repaired
38.16315600000	85.74978800000	5.83	LOUISVILLE	LOUISVILLE TERMINAL	CRT	Repaired
38.14820800000	85.74522900000	6.93	LOUISVILLE	LOUISVILLE TERMINAL	CRT	Repaired
38.14660700000	85.74479900000	7.09	LOUISVILLE	LOUISVILLE TERMINAL	CRT	Repaired
38.14433700000	85.74477400000	7.14	LOUISVILLE	LOUISVILLE TERMINAL	PR1	Exception Not Found

Table A3: Sample MGT Data (CSX)

DIV	SUB	PREFIX	BEGIN_MILEPOST	END_MILEPOST	TRACK	TONNAGE_MEASUREMENT_D	MGT
AP	BIGSANDY	CMG	1	2	1	01/14/2013	25.36
AP	BIGSANDY	CMG	1	2	2	01/14/2013	20.49
AP	BIGSANDY	CMG	2	2	1	01/14/2013	25.46
AP	BIGSANDY	CMG	2	2	2	01/14/2013	20.68
AP	BIGSANDY	CMG	2	2	1	01/14/2013	25.52
AP	BIGSANDY	CMG	2	2	2	01/14/2013	20.80
AP	BIGSANDY	CMG	2	3	1	01/14/2013	24.89
AP	BIGSANDY	CMG	2	3	2	01/14/2013	20.70
AP	BIGSANDY	CMG	3	5	1	01/14/2013	24.20
AP	BIGSANDY	CMG	3	5	2	01/14/2013	20.52
AP	BIGSANDY	CMG	5	5	1	01/14/2013	24.05
AP	BIGSANDY	CMG	5	5	2	01/14/2013	20.59
AP	BIGSANDY	CMG	5	7	1	01/14/2013	23.83
AP	BIGSANDY	CMG	5	7	2	01/14/2013	20.60
AP	BIGSANDY	CMG	7	8	1	01/14/2013	23.51
AP	BIGSANDY	CMG	7	8	2	01/14/2013	20.58
AP	BIGSANDY	CMG	8	9	1	01/14/2013	23.10

Table A4: Sample VTI Data (CSX data)

ID	SUB_DIV_NAME	LINECODE	GPS_SPEED	MPDECIMAL	RUN_DATE	LATITUDE	LONGITUDE	EXCEPTION_VALUE	EXC_TYPE	SEVERITY
32997	KANAWHA	CA	48	448	11/1/2008 1:03	38.28629	-81.5703	1.07	CBV-PEAK	Priority
33095	KANAWHA	CA	47	457	11/1/2008 1:15	38.364	-81.6918	109720.5	AXV2-PEAK	Priority
33827	ST LOUIS LINE	QS	56	41	11/1/2008 2:41	39.64739	-86.8914	126346	AXV2-PEAK	Near Urgent
33338	ST LOUIS LINE	QS	58	55	11/1/2008 2:58	39.59249	-87.1669	112739.1	AXV2-PEAK	Priority
33348	ST LOUIS LINE	QS	58	55	11/1/2008 2:58	39.59231	-87.1684	110556.4	AXV1-PEAK	Priority
33023	FITZGERALD	ANB	46	656	11/1/2008 3:00	31.70806	-83.2181	131454.4	AXV1-PEAK	Urgent
33085	ST LOUIS LINE	QS	47	61	11/1/2008 3:03	39.56319	-87.2622	1.03	CBV-PEAK	Priority
33206	ST LOUIS LINE	QS	19	75	11/1/2008 3:58	39.4699	-87.4524	112088.9	AXV1-PEAK	Priority
33255	ST LOUIS LINE	QS	49	80.7	11/1/2008 4:11	39.46898	-87.5451	118637	AXV1-PEAK	Near Urgent
33330	ST LOUIS LINE	QS	47	91	11/1/2008 4:23	39.40144	-87.7065	122955.9	AXV2-PEAK	Near Urgent
33514	ST LOUIS LINE	QS	57	112	11/1/2008 5:20	39.27107	-88.0774	107259.2	AXV1-PEAK	Priority
33663	FITZGERALD	ANB	41	717	11/1/2008 6:10	32.21425	-83.9124	115711.2	AXV1-PEAK	Near Urgent
33961	FITZGERALD	ANB	46	758.9	11/1/2008 7:08	32.56043	-84.4436	125881.6	AXV1-PEAK	Near Urgent
33195	ST LOUIS LINE	QS	58	217	11/1/2008 7:53	38.71465	-89.8067	104147.7	AXV1-PEAK	Priority
33147	LINEVILLE	ANJ	44	825	11/1/2008 10:05	33.03239	-85.1205	100293.2	AXV2-PEAK	Priority
33868	S & N A NORTH		40	366	11/1/2008 22:28	33.85456	-86.7473	116547.2	AXV2-PEAK	Near Urgent

Table 5: Sample FRA Track Geometry Inspection Data (CSX)

Run ID	Date	Railroad	Unit	MP	MP Foot	Type	Value	Length	Latitude	Longitude	Posted Class	Actual Class	Track
2008010703	1/7/2008	CSXT	DOT X 220	40	666	Twist 31ft*	2.001	8	39.73215	-76.84795	2	0	5
2008010703	1/7/2008	CSXT	DOT X 220	56	9120	Rockoff Hazard	1.997	118	39.80444	-76.98345	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	50	4036	R Cant POS*	5	6	39.82379	-76.89374	2	0	5
2008010703	1/7/2008	CSXT	DOT X 220	47	3730	L Prof 62	2.963	3	39.79923	-76.86313	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	46	3203	Warp 62	2.655	62	39.78915	-76.85311	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	21	9439	L Prof 62	-2.894	2	39.4735	-76.81777	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	10	5035	Warp 62	3.156	19	39.26684	-76.64007	1	0	1
2008010702	1/7/2008	CSXT	DOT X 220	80	2342	Gage Wide	57.988	7	39.83283	-77.23386	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	47	3730	R Prof 62	2.985	3	39.79923	-76.86313	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	39	2426	Crosslevel	-2.393	3	39.35389	-76.70494	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	38	3616	Warp 62	2.579	61	39.70721	-76.82561	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	45	3992	Warp 62	2.429	57	39.77759	-76.83356	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	21	9435	Gage Wide	57.992	13	39.47351	-77.23386	2	1	5
2008010703	1/7/2008	CSXT	DOT X 220	14	1855	Lmt Speed 3	22	396	39.30588	-76.65892	2	0	5

Appendix B

CSX Baltimore Division Rail and Geometry Defects By Type

DEFECTS	TDD	BRO	TDT	TDC	OAW	EFBW	TW	HSH	VSH	SW	PIPE	HW	BHB	BB	FH	CH	SD	TOTAL	% of Total
CROSSLEVEL	14	1	0	0	1	4	3	2	3	1	0	4	0	0	0	2	2	37	18.14%
EXECESS ELEVATION	3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	5	2.45%
ALINGMENT	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0.49%
ALINGMENT LEFT	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	2	0.98%
ALINGMENT RIGHT	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	0.98%
WARP 31	18	0	0	0	0	2	3	1	3	0	0	1	0	0	0	2	4	34	16.67%
WARP 62	7	0	0	0	1	0	1	0	1	0	0	4	0	0	1	1	0	16	7.84%
CLIM	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0.49%
Curve Speed 3IN	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0.98%
Left Vert ACC	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	3	1.47%
PROFILE LEFT	1	0	0	0	0	2	2	0	1	0	0	1	0	0	0	0	0	7	3.43%
PROFILE RIGHT	4	1	0	0	0	0	2	0	0	0	0	0	1	0	0	1	0	9	4.41%
RIGHT RAIL CANT	12	0	0	0	0	0	3	1	2	0	0	0	0	0	0	1	0	19	9.31%
LEFT RAIL CANT	30	0	0	0	0	3	1	0	2	0	0	3	0	0	0	2	1	42	20.59%
TIGHT GAGE	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1.96%
WIDE GAGE	14	0	0	0	0	2	3	0	1	0	0	0	0	0	0	0	0	20	9.80%
TOTAL	111	2	0	0	2	13	22	4	14	1	0	13	2	0	1	10	9	204	
% of Total	54.41%	0.98%	0.00%	0.00%	0.98%	6.37%	10.78%	1.96%	6.86%	0.49%	0.00%	6.37%	0.98%	0.00%	0.49%	4.90%	4.41%		
	TDD	BRO	TDT	TDC	OAW	EFBW	TW	HSH	VSH	SW	PIPE	HW	BHB	BB	FH	CH	SD		

Appendix C

SKELETON CODE

```
clear all
%Input Geometry defect file
[~, ~, raw] = xlsread('Baltimore_Geo.xlsx');
%Asign to geo matrix
geo = raw;
%Input Rail Defect file
[~, ~, raw] = xlsread('BA_Divisions_2010_2012.xlsx');
%Assign to matrix
rail = raw;
%Creat zero matrix to store matchs
matches = mat2cell(zeros(length(rail),39));
n=1;
m=1;
p=1;
o=1;
%For loop to pick row in rail defect
for k=2:length(rail)
    %For loop to pick row in geometry defect
    for i=2:length(geo)
        %Find difference in milepost decmial formate
        if strcmp(rail{k,2}, geo{i,1})
        mileg=geo{i,31};
        miler=str2num(rail{k,3});
        dist = abs(mileg-miler);
        %For loops to for matching
        if dist < .005
            m=m+1;
            if strcmp(rail{k,5},geo{i,22})
                p=p+1;
                if datenum(geo{i,4}) < datenum(rail{k,10})
                    %Store matched rows
                    for w=1:37
                        matches{n,w} = geo{i,w};
                    end
                    for w=1:17
                        matches{n,w+37} = rail{k,w};
                    end
                    n = n+1;
                    break
                end
            end
        end
        end
    end
end
```

Appendix D

MASTER CHARTS

Divisions	Length in Miles	Annual MGT	Reported Geo Defects	Unique Geo Defects	Reported Warp 31 Defects	Reported Rail Cant Defects	Reported Gage Defects	Rail Defects	TDD	Matches	Matched TDD	Matches On Curves	% of Matches on Curves	Matched TDD on Curves	% of Matched TDD on Curves
Baltimore Full	1042.8	24.05	18351	11462	3619	5242	1993	1543	697	204	111	185	90.69%	87	78.38%
Baltimore > 19.5	798.4	29.60	16207	10238	2924	4644	1711	1330	641	182	107	168	92.31%	85	79.44%
Atlanta Full	2042.1	25.68	27536	16044	5862	8280	5120	2285	920	150	67	51	34.00%	29	43.28%
Atlanta > 19.5	1405.9	33.47	22212	12498	4332	7038	4641	1750	789	120	58	48	40.00%	27	46.55%
Albany	1871.1	25.02	22738	14645	3424	5247	2044	1774	504	157	69	71	45.22%	39	56.52%
Albany > 19.5	840.46	47.00	11906	8757	1630	1735	1014	934	292	78	39	46	58.97%	26	66.67%
Appalachian Full	2221.2	17.29	49616	26383	7022	21489	6490	4786	1705	763	394	657	86.11%	363	92.13%
Appalachian > 19.5	952.44	28.17	23018	12958	2690	12411	2139	2348	917	465	255	394	84.73%	233	91.37%
C&O Full	2051.4	19.03	61677	39658	11488	16768	11174	2374	1293	496	322	407	82.06%	269	83.54%
C&O > 19.5	770.04	37.5	25979	13715	4431	8607	4343	1244	760	306	244	246	80.39%	223	91.39%
Chicago Full	1630.7	18.65	16978	10053	3054	1324	640	1464	185	128	19	59	46.09%	13	68.42%
Chicago > 19.5	400.8	51.02	4533	1965	987	437	190	611	87	51	12	25	49.02%	8	66.67%
Florence Full	3056.5	15.37	38785	22457	5227	6259	3811	4019	1292	285	102	136	47.72%	48	47.06%
Florence > 19.5	1125.9	28.35	18514	10086	2204	3661	1902	1835	731	158	69	72	45.57%	31	44.93%
Great Lakes Full	2326.2	32.8	18485	13202	2555	1824	1016	2520	433	195	41	70	35.90%	19	46.34%
Great Lakes > 19.5	1840.2	40.27	15671	11246	2079	1685	859	2304	391	177	39	59	33.33%	18	46.15%
Jacksonville Full	2920.1	14.97	28308	15727	3075	4678	3269	2272	644	178	46	73	41.01%	17	36.96%
Jacksonville > 19.5	784.37	32.66	10217	5784	848	1938	1434	739	189	75	20	33	44.00%	8	40.00%
Louisville Full	1406.8	17.24	31123	17435	2381	7691	2661	1657	443	156	65	82	52.56%	43	66.15%
Louisville > 19.5	559.94	32.67	10285	7042	1150	4667	1467	982	236	100	38	51	51.00%	27	71.05%
Nashville Full	1658.9	30.61	21340	15275	3176	4400	2204	1746	462	206	53	80	38.83%	25	47.17%
Nashville > 19.5	1202.4	39.59	14772	10663	2592	3698	1894	1351	359	108	32	56	51.85%	15	46.88%
Total Full	22228	21.33	334937	202341	50883	83202	40422	26440	8578	2918	1289	1871	64.12%	952	73.86%
Total Filter	10681	35.94	173314	104952	25867	50521	21600	15428	5392	1820	913	1198	65.82%	701	76.78%

MASTER CHARTS

Column	Tangent	Curve	TOTAL FULL
Length in Miles	15725.12	6502.6	22227.72
Annual MGT	21.33	21.33	21.33
Reported Geo Defects	87532	247405	334937
Geo Defect Density	5.57	38.05	15.07
Reported Gage Defects	4306	36116	40422
Gage Defect Density	0.27	5.55	1.82
Reported Warp 31 Defects	2289	48594	50883
Warp 31 Defect Density	0.15	7.47	2.29
Reported Warp 62 Defects	16775	6104	22879
Warp 62 Defect Density	1.07	0.94	1.03
Reported Crosslevel Defects	52613	14393	67006
Crosslevel Defects Density	3.35	2.21	3.01
Reported Alignment Defects	2029	13175	15204
Alignment Defect Density	0.13	2.03	0.68
Reported Cant Defects	7557	75645	83202
Cant Defects Density	0.48	11.63	3.74
Reported Rail Defects	19761	6679	26440
Rail Defect Density	1.26	1.03	1.19
Reported TDD Defects	5487	3091	8578
TDD Defects Density	0.35	0.48	0.39
Matches	1047	1871	2918
Matches Density	0.07	0.29	0.13
Matched TDD	337	952	1289
Matched TDD Density	0.02	0.15	0.06
Repeat Matches	343	784	1127
Repeat Matches Density	0.02	0.12	0.05

Gage Matches	40	210	250
Gage TDD Matches	15	127	142
Repeat Gage Matches	14	94	108
Repeat Gage TDD Matches	5	58	63
Crosslevel Matches	452	110	562
Crosslevel TDD Matches	146	34	180
Repeat Crosslevel Matches	125	43	168
Repeat Crosslevel TDD Matches	43	16	59
Warp 31 Matches	44	472	516
Warp 31 TDD Matches	9	197	206
Repeat Warp 31 Matches	14	168	182
Repeat Warp 31 TDD Matches	5	66	71
Warp 62 Matches	173	35	208
Warp 62 TDD Matches	47	18	65
Repeat Warp 62 Matches	65	15	80
Repeat Warp 62 TDD Matches	17	5	22
Cant Matches	108	757	865
Cant TDD Matches	52	454	506
Repeat Cant Matches	28	310	338
Repeat Cant TDD Matches	18	191	209
Alignment Matches	20	92	112
Alignment TDD Matches	6	41	47
Repeat Alignment Matches	11	63	74
Repeat Alignment TDD Matches	5	28	33