How Track Geometry Defects Affect the Development of Rail Defects

Dr. Allan M Zarembski PE, FASME, Hon. Mbr. AREMA
Professor and Director of the Railroad Engineering and Safety Program
Department of Civil and Environmental Engineering
University of Delaware, Newark DE

Dr. Nii Attoh-Okine
Professor, Department of Civil and Environmental Engineering
University of Delaware, Newark DE

Daniel Einbinder
Associate Engineer
ENSCO, Inc.
Springfield, VA

Hugh Thompson
Program Manager
Federal Railroad Administration, Washington DC

Ted Sussman
Civil Engineer
Volpe National Transportation Systems Center, Cambridge MA

ABSTRACT

Paper presents results of FRA sponsored study on relationship between geometry defects and rail defects and specifically the relationship between presence of one of more track geometry defects and development of rail defects at that same location.

In this study, approximately 335,000 track geometry defects were correlated with approximately 26,000 rail defects on 22,000 miles (36,700 km) of a major US Class 1 railroad. Correlation and statistical analyses were performed looking at

- Relationship between life of rail (cumulative MGT) and presence of geometry defect(s).
- Relationship between probability of rail defect occurring at a given location and presence of one of more geometry defects at that location.

The results showed that there was a statistically significant relationship between geometry defects and rail defects, when the geometry defect preceded the rail defect. Analysis showed that the if a track geometry defect is presents, the cumulative MGT "life" of the rail was approximately 30% less than that of a rail defect with no track geometry defect. Probability analysis showed that the presence of a geometry defect has a strong and well defined effect on the development of a rail defect. For example, a single geometry defect increases probability of a rail defect by 6 to 13 times, while multiple geometry defects will increase probability of a rail defect by factors of up to 600 times; up to a probability of occurrence of 80 to 90%, depending on number and type of geometry defects.

INTRODUCTION

Traditionally, track defects such as rail defects and track geometry defects are treated as independent conditions. However, there has long been a question as to whether there is in fact a relationship between the presence of track geometry defects and the development of rail defects at that same location.

Theory suggests that there may be a relationship. Extensive theoretical research has shown that the presence of geometry defects generates increases dynamic wheel/rail loads [1,2, 3] which in turn can result in earlier development of rail fatigue defects and an associated reduction in fatigue rail life. That is because the defects result in a dynamic effect on every wheel that passes over the rail section, increasing the level of loading and the associated level of stress experienced by the rail [4, 5]. This includes both bending stresses and contact stresses, both of which have an effect on the development of rail defects [6, 7, 8].

In order to examine this relationship, a US Federal Railroad Administration sponsored study[1] correlated multiple years of track geometry with a data base of several years of rail defects obtained from a major US Class 1 railroad [9]. The railroad system data represented more than 22,000 track miles (36,700 km), and included:

- Three years of rail defect data, representing approximately 50,000 defect records, which was subsequently narrowed to approximately 26,000 defects of "interest"
- Five years of track geometry data representing approximately 335,000 defect records
- Tonnage data (annual MGT)

Correlation and statistical analyses were performed and two sets of analyses relationships were developed:

- Relationship between the life of rail (in cumulative MGT) and the presence of geometry defect(s).
- Relationship between the probability of a rail defect occurring at a given location and the presence of one of more geometry defects at that location.

This paper will discuss the correlation analyses and the probability analysis of a rail defect occurring after a geometry defects at the same location.

Table 1 presents a summary of the initial correlation between rail defects and geometry defects for the full system (22,000 miles) and for the "high tonnage" segments, defined here as having a minimum annual tonnage of 20 MGT. As can be seen from Table 1, for the full system, 11% of all rail defects were preceded[2] by one or more track geometry defects. For track with greater than 20 MGT annual tonnage, this percentage increases to almost 12%. Likewise for the full system, 15% of all TDD rail defects[3] were preceded by one or more track geometry defects and for track with greater than 20 MGT annual tonnage, this percentage increases to almost 17%.

In contrast, if the relationship between rail defects and geometry defects were purely random[4] the probability of a match at a given location was calculated to be 1.4% for all defects and 0.6% for TTDs. Thus the actual percentages of matches were of the order of 7 to 20 times that which would occur purely by random chance.

Analysis of the matches between the rail defects and preceding geometry defects, showed that a large percentage of these matches had in fact multiple (two or more) geometry defects preceding the rail defect, at the same location. These repeat matches were either the same type of geometry defect occurring at a different time (corresponding to a different geometry car run) or were a different type of geometry defect at the same location. These results, show that for the full system, 38% of the matches had multiple geometry defect matches ("Repeats"), while for TDD defects 41% of the matches had multiple geometry defect matches ("Repeats"). The higher density (> 20 MGT) track showed a similar behavior

---

[1] Integration of Multiple Inspection System Data to Identify Potentially Unsafe Track Rail Conditions, FRA Contract DTFR53-13-C-00066, August 2013

[2] In this analysis, only geometry defects that preceded the rail defect at the same location was considered. Geometry defects occurring after the date of the rail defect were ignored in the analysis.
[3] Detail fracture (TTD) defects represented the largest single category of rail defects, approximately 44% of all rail defects.
[4] Based on a random analysis, performed using Bayes' Rule and Conditional Probability

	Length in Miles	Annual MGT	Reported Geo Defects	Unique Geo Defects	Rail Defects	Matches	TDD	Matched TDD	% of All Matched	% of TDD Matched
Full System	**22228**	**21.3**	**334937**	**202341**	**26440**	**2918**	**8578**	**1289**	**11.04%**	**15.03%**
High density segments	**10681**	**35.9**	**173314**	**104952**	**15428**	**1820**	**5392**	**913**	**11.80%**	**16.93%**

Table 1: Summary of Rail and Geometry Defects and Matches

	Rail Defects	Rail Defects on Curves	Matches	Matched on Curves	TDD	TDD on Curves	Matched TDD	Matched TDD on Curves	**% of All Matched**	**% of TDD Matched**
Full	26440	8870	2918	1871	8578	3091	1289	942	**21.09%**	**30.48%**
High density	15428	5225	1820	1113	5392	2030	913	635	**21.30%**	**31.28%**

Table 2: Summary of Class 1 Railroad System Rail and Geometry Defect Matching Data on Curves

On curves (Table 2), the results were even more dramatic, with 21% of all rail defects were preceded by one or more geometry defects and approximately 10% of all rail defects were preceded by two or more track geometry defects. For TDD defects only, over 30% of all rail defects were preceded by one or more geometry defects and 15% of all TDD rail defects were preceded by two or more track geometry defects.

When the rail defects were correlated to geometry defects, by type of defect, the analyses showed that the most commonly occurring geometry defects that matched a rail defect were cant, warp, gage and cross-level as illustrated in Table 3:

	All track		Curves only	
	All defects	TDD defects	All defects	TDD defects
Cant	31.6%	39.4%	44.0%	50.2%
Cross-level /CLIM	18.0%	13.8%	6.0%	3.8%
Warp[5]	25.0%	21.2%	28.8%	23.5%
Gage/Track Strength	11.8%	14.7%	16.4%	18.9%

Table 3: Key Geometry Parameters based on Frequency of Matches with Rail Defects

It is again noted that detail fracture (TDD) defects represent the single largest category of matched rail defects, with approximately 44% of all defects. Other rail defects included Bolt Hole Breaks (BHB)- 8%, Electric Flash Butt Welds (EFBW) 7%, Thermit Welds (TW) 12%, Vertical Split Heads (VSH) 5%, Head and Web separation (HW) 8% and Shelly Spots (SD) 9%.

Reduction in Life

In order to determine if there was a reduction in "age" of the defect (tonnage to failure), the Class 1 railroad data was used to calculate the cumulative MGT of a given rail defect, using the rail roll date[6] and annual traffic level (MGT) at the defect location, adjusted[7] to reflect the changes in MGT by year from the roll date to the date of failure..

Using this approach, it was possible to calculate the cumulative MGT of a rail at the time the defect was removed from track (with the exception of relay rails in second position where the details of the time and tonnage in its first position are not known) and to calculate whether there was a reduction in rail life if there was geometry defect present.

Table 4 presents a summary of the rail defect life for matched rail defects vs. not matched rail defects (no geometry defect reported at that location). This table also compares the rail defect life on curve and tangent track. As can be seen from this Table, *the reduction in rail life for the full system , all tracks, was 31% when geometry defects are present. For high density track (> 20 MGT) the reduction was 25%.*

	Tangent			Curve			All Track		
	Matched	No Match	Reduction in Life	Matched	No Match	Reduction in Life	Matched	No Match	Reduction in Life
Full System	504	704	-28%	413	477	-13%	447	646	-31%
High Density	722	895	-19%	562	602	-7%	620	824	-25%

Table 4: Reduction in Life of Defect (in MGT)

[5] Warp is further subdivided in this study into Warp 31, based on a 31 foot chord and Warp 62 based on a 62 foot chord.
[6] The installation date was assumed to be the roll date plus 1 year.
[7] MGT was adjusted using AAR annual traffic growth statistics.

Probability Analysis

Probability Analysis examines the probability of a rail defect occurring given a geometry defect preceding it. The Probability analysis approaches presented here include:

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

As can be seen, Conditional Probability Analyses (Bayesian methods) were used 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, the focus here is on the effect of a geometry exception (geometry defect) on the likelihood (probability) of the development of a rail defect.

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 1 and Equation 2 below.

$$P(RD) = \frac{39ft*Annual\ Rail\ Defects}{Length\ of\ track\ in\ ft*2} \qquad \text{Equation 1}$$

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

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 5 below has the results of these equations for the entire railroad (Class 1 railroad). The geometry defects shown in Table 5 were determined 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%

Table 5. 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 Class 1 railroad data presented previously. Also shown in Table 5 are the probabilities for any specific geometry defect occurring on a random basis.

However, as seen from the correlation analyses above, rail defects do not appear to occur randomly, but appear to 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) [9]

Bayes' Theorem

The most common and basic method of determining Conditional Probability is Bayes' Theorem which is given below in Equation 3 for the probability of a geometry defect given a rail defect follows it [P(GD|RD)].

$$P(GD|RD) = \frac{P(GD \cap RD)}{P(RD)} \quad \text{Equation 3}$$

A complete discussion of the analysis and associated equations is given in Reference 10.The results from the Bayes' analysis are shown below in Table 6. From this table it can be determined 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[8]
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

Table 6. Conditional probabilities for a rail defect to occur given a geometry defect preceded it from Bayes' Theorem

Naïve Bayes

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 [11.12].

Using the same inputs as already discussed when applying Bayes' Theorem [10], the results from the Naïve Bayes analysis is presented in Table 7 for several different geometry defect combinations

[8] Ratio of probability of a rail defect occurring in conjunction with a track geometry defect P(RD|GD) vs random occurrence (0.18%).

	0 = No defect ; 1 = single defect (of type shown in column)						
							ALL
	Alignment	**Crosslevel**	**Gage**	**Rail Cant**	**Warp 31**	**Warp 62**	**P(RD | GD)**
1 Defect	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%
2 or More Defects	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%

Table 7 Results of different geometry defect combinations using Naïve Bayes (all track)

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 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 and 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 Network

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 [13,14].

In order to apply this Bayesian Network approach, and to develop a model based on the Class 1 railroad data, a program called NETICA was used. Figure 1 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.

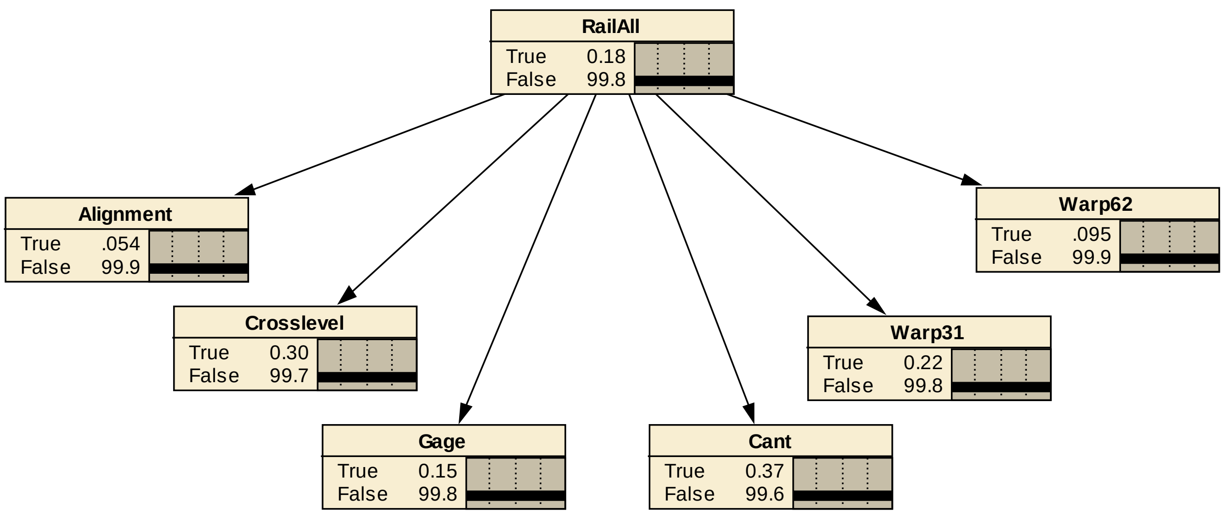

Figure 1. Bayesian network in NETICA; Random case.

The RailAll node is the pure random probability of a rail defect occurring anywhere on the railroad. Thus, as can be seen in Figure 1, 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.

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 result of several different cases are shown in the following figures.

Figure 2 represents a warp 31 defect; where the conditional probability of a rail defect following a Warp 31 defect is 1.48%.
Figure 3 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 4 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 5 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%.

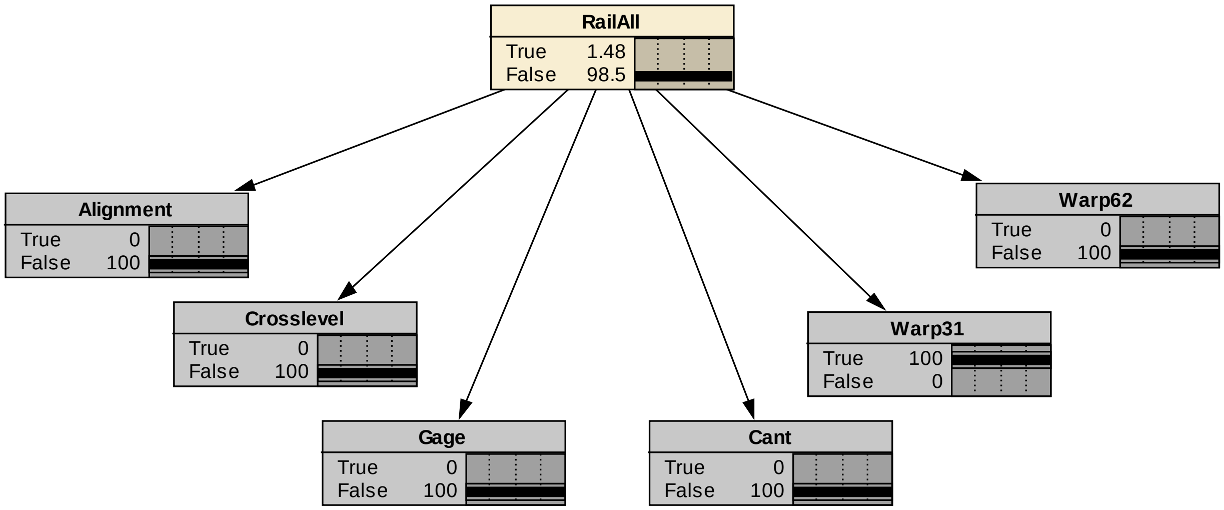

Figure 2. Bayesian network for a warp 31 defect

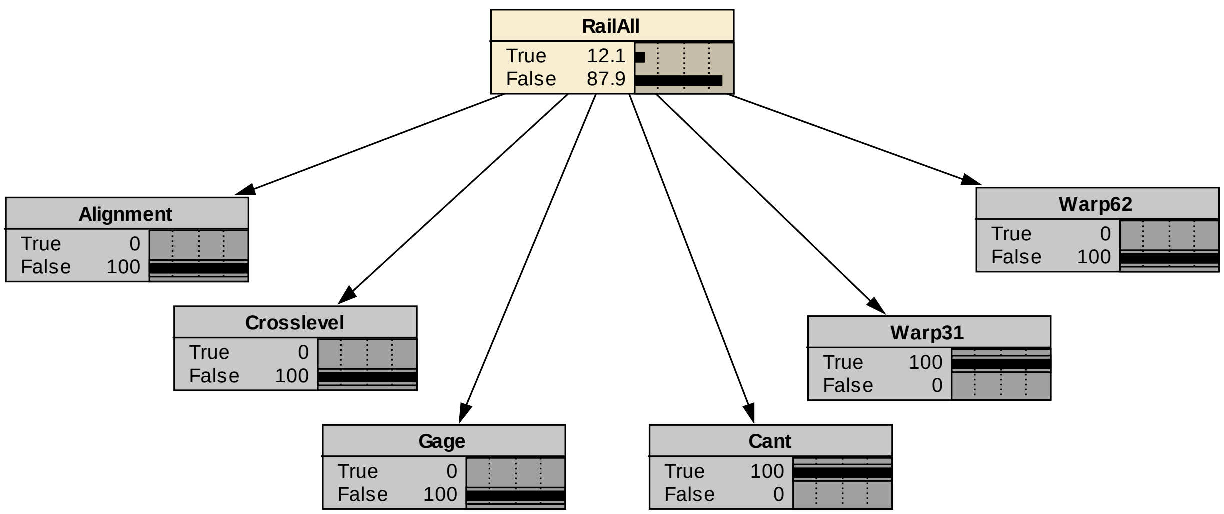

Figure 3. Bayesian network for a rail cant and a warp 31 defect

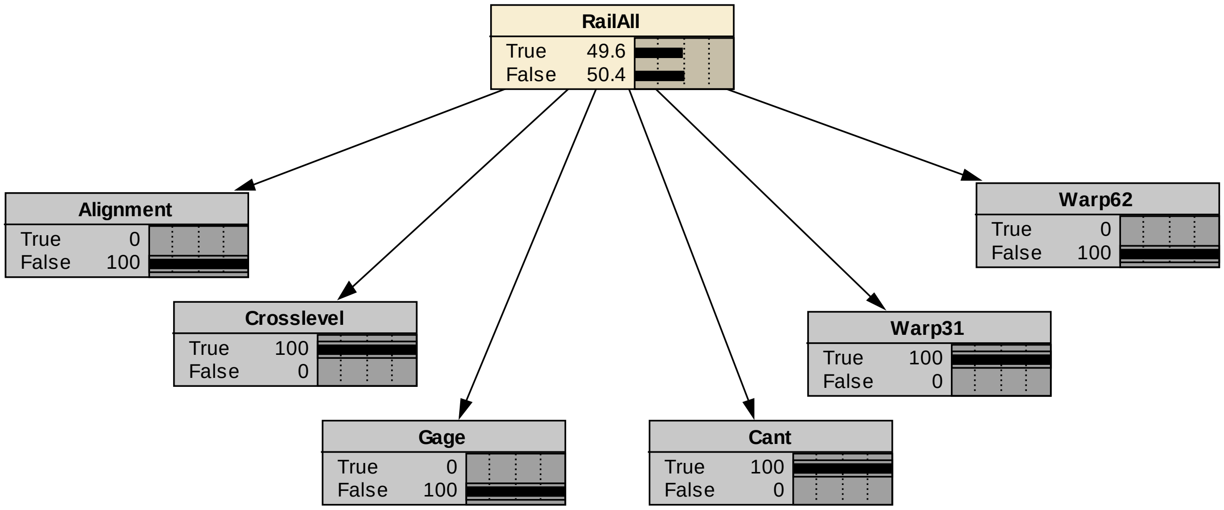

Figure 4. Bayesian network for a warp 31, a rail cant, and a crosslevel defect

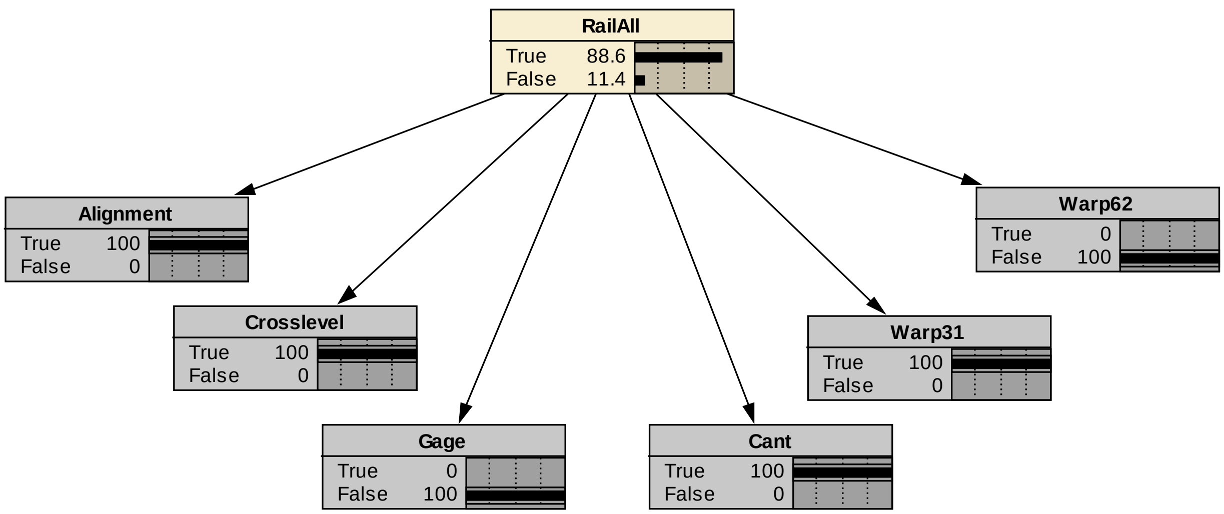

Figure 5. Bayesian network for a warp 31, a rail cant, a crosslevel, and an alignment defect

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 defect 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 Table 8. As can be seen, the differences are relatively small but distinct.

							ALL TRACK	
	Alignment	Cross-level	Gage	Rail Cant	Warp 31	Warp 62	P(RD | GD) Naïve	P(RD|GD) Network
1 defect	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%
2 or More Defects	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%

0 = No defect ; 1 = single defect (of type shown in column)

Table 8. Comparison of Naïve Bayes and Bayesian networks models for select geometry defects

Comparison of Tangent and Curve Track

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 which on tangent track. The same process was carried out with each set of tangent and curve track data. Table 9 presents a comparison of the random probabilities of rail and geometry defects occurring on tangent, curve, and all tracks.

	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%

Table 9. Random Probability of a defect; Tangent, Curve, and All

As can be seen in Table 9, 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).

Bayes' Theorem was used 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 Table 10.

	TANGENT		CURVE		ALL	
	P(RD|GD)	Ratio P/R	P(RD|GD)	Ratio P/R	P(RD|GD)	Ratio P/R
P(RD) Random	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**

Table 10. Conditional probability of a rail defect occurring after a geometry defect occurred from Bayes' Theorem

In examining Table 10, it is necessary to look at both the condition probability P(RD|GD) and the "Ratio P/R" which is the ratio of the Conditional probability/Random probability and shows the increase in risk of developing rail defect over pure random probability. Thus, while in general, tangent track has a higher conditional probability than curve, for the geometry defect categories, it also has a higher random probability, thus the increased "risk" of developing a defect [increase over random] is in several cases higher for the curve track, even though the conditional probability is somewhat lower.

Table 11 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, with the Bayesian Network model being considered the more “accurate” model because of the way it handles interactions between variables.

							TANGENT		CURVE		ALL	
P(RD) Random							0.186%		0.15%		0.18%	
	Alignment	Cross-level	Gage	Rail Cant	Warp 31	Warp 62	P(RD | GD) Naïve	P(RD|GD) Network	P(RD | GD) Naïve	P(RD|GD) Network	P(RD | GD) Naïve	P(RD|GD) Network
1 defect	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%
2 or More Defects	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%

0 = No defect ; 1 = single defect (of type shown in column)

Table 11 Comparison of Naïve Bayes and Bayesian network models for tangent, curve, and all track

As can be seen in Table 11, a single geometry defect increases the probability of a rail defect (from random) by 6 to 13 times, while multiple geometry defects will increase the probability of a rail defect (from random) by factors of up to 500 times, with a Probability of developing a rail defect after four geometry defects being of the order of 80% to 90%.

Finally, analysis of the original rail and geometry defect matches showed that the occurrence of multiple geometry defects, prior to the development of rail defects is very common, based on the Class 1 railroad data. This is illustrated in Table 12, which shows 1119 occurrences where two or more geometry defects were present at the same location as (and prior to) the rail defect. Furthermore, 44% of these had 3 or more geometry defects (495) and 29% (328) had 4 or more geometry defects.

Number of Geo Defects in Match	Number Of Occurrences
>5	172
5	33
4	123
3	167
2	624
Total	1119

Table 12: Number of multiple geometry matches

Summary and Conclusions

The results of the correlation analysis showed that for the Class 1 railroad examined in this study (22,228 miles/37,047 km of track) there was a statistically significant relationship between geometry defects and rail defects (where the geometry defect preceded the removal of the rail defect). The overall matching of rail defects (all rail defects) with track geometry defect at the same location was 11%. Furthermore, 38% of matches (4.2% of all defects) were preceded by two or more track geometry defects at the same location. On curves only, the overall matching of rail defects with track geometry defect was 21%, with 46% of these preceded by two or more track geometry defects at the same location.

Looking at the largest subclass of rail defects, TDD defects -detail fractures, the correlations were even greater, with overall matching of TDD defects with track geometry defect at the same location of 15% and on curves, overall matching of TDD defects with track geometry defect at the same location of 30.5%.

For the high traffic density lines (> 20 MGT) (10,681 miles /17,802 km) of track there was likewise a significant relationship between geometry defects and rail defects. Here the overall matching of rail defects with track geometry defects at the same location was 12%, with 38% of these preceded by two or more track geometry defects at the same location. On curves only, for the high density lines the overall matching of rail defects (all defects) with track geometry defect was 21%, with 43% of these being preceded by two or more track geometry defects at the same location.

Similar results for TDDs were observed on the high density lines, with overall matching of 17%, for all track and 31.3% on curves.

A random behavior analysis showed that this behavior is not random but rather statistically significant. Furthermore, a significant percentage of these matched rails-track geometry defects occurred under repeat geometry defects, with approximately 40% of all matches being with two or more geometry defects at the same location (50% on curves).

Analysis of the age of the rail defects, showed that there was a distinct and well defined reduction in rail defect “life”, defined as MGT to failure. For the full system , all tracks, the reduction in rail defect life was 31% when geometry defects are present. For the high density track (> 20 MGT) the reduction in rail defect life when geometry defects are present was 25%.

Probability Analysis examined the probability of a rail defect occurring given a geometry defect preceding it. The results of the random analysis showed that for all track the probability of a rail defect occurring randomly is 0.18% for all track, 0.15% for curves and 0.19% for tangent track. The Bayesian analyses showed that the presence of geometry defects have a significant impact on the probability of development of a rail defect.

Thus, a single geometry defect increases the probability of a rail defect (from random) by 6 to 13 times (to 1 to 3%), while multiple geometry defects will increase the probability of a rail defect (from random) by factors of up to 500 times. Thus 2 geometry defects increase the probability of a rail defect to approximately 10 to 20 % depending on defect type, 3 geometry defects increase probability of a rail defect to approximately 40 to 50%, and 4 geometry defects increase probability of a rail defect to approximately 80 to 90%.

Finally, on a broader scope, it can be concluded that “Big Data’ analysis techniques used in this study are effective in dealing with large volumes and with large scale data bases where relationships between parameters are not always intuitively obvious. As such these Big Data tools such as MARS and Bayesian modeling can be applied to other areas where large scale data bases are available but have not been used for anything but the most basic exception reporting and data base uses.

Acknowledgements

The authors would like to acknowledge and thank the Federal Railroad Administration, U.S. Department of Transportation for their support of this research. The authors would also like to thank Gary Carr and Cameron Stuart for their support in this project.

References

1. 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.
2. 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.
3. Zarembski, A.M., Bonaventura, C.S., “Dynamic Effects of Track Surface Condition on Vertical Wheel/Rail Forces and Energy Consumption”, ASME/IEEE/AREMA Joint Rail Conference 2010, Urbana, IL, April 2010.
4. Zarembski, A. M., “Effect of Rail Section and Traffic on Rail Fatigue Life”, American Railway Engineering Association, 78th Annual Technical Conference, Chicago, IL, March 1979.
5. Zarembski, A. M., “Effect of Increasing Axle Loads on Rail Fatigue Life”, Bulletin of the American Railway Engineering Association, Bulletin 685, Volume 83, November - December 1981.
6. Kerr, A.D. (2004). Fundamentals of Railway Track Engineering. Omaha, Nebraska: Simmons-Boardman Press, Inc.
7. Steele, R.K., and R.P. Reiff (1981). “Rail: Its Behavior and Relationship to Total System Wear.” FAST Engineering Conference (November). Denver Colorado.
8. Zarembski, A.M., Palese, J.W., “Evaluation of the Effectiveness of Rail Grinding on Reducing Rail Defects on North American Class 1 Railroad”, AREMA 2010 Annual Conference and Exposition, Orlando, FL, August 2010
9. Bolstad, W.M., Introduction to Bayesian Statistics, 2nd Ed. John Wiley and Sons, Inc., Hoboken, NJ, 2007

10. Zarembski, Allan M., Integration of Multiple Inspection System Data to identify Potentially Unsafe Track Rail Conditions, Report submitted to US Department of Transportation, federal Railroad Administration under contract DTFR53-13-C-00066 February 2016.
11. Lewis, D.D., "Naïve (Bayes) at Forty: The Independence Assumption in Information Retrieval", Proc. 10th European Conference on Machine Learning (ECML98), 1998
12. 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 Conference on Research and Development in Information Retrieval,* 2000
13. 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
14. Bielza, C., Larranaga P. (2014), "Bayesian Networks in Neuroscience: a survey", Frontiers in Computational Neuroscience