# On the Relationship between Track Geometry Defects and Development of Internal Rail Defects

Professor Allan M. Zarembski[1] , Professor Nii Attoh-Okine[2], Daniel Einbinder[3]

[1] University of Delaware, Newark, Delaware, USA

[2] University of Delaware, Newark, Delaware, USA

[3] ENSCO, Inc. Springfield, Virginia, USA

*dramz@udel.edu*

## Abstract

For many years, railway engineers have hypothesized a relationship between track geometry defects and rail defects. The basic reasoning has been that the presence of one of more track geometry defects will generate increased dynamic wheel/rail loads, which in turn cause can accelerated rail damage, leading to the development of rail defects. While some evidence of this relationship has been demonstrated for the development of rail surface defects, little or no evidence is available linking the presence of track geometry defects to the development of internal rail defects. This paper presents the results of a multi-year project, sponsored by the US Federal Railroad Administration, aimed at developing relationships between track geometry defects, as identified by Track Recording Cars and the development of internal rail defects. The project made use of approximately five years of inspection data from 37,000 km of track on a major US freight railroad and used state of the art “Big Data” analysis techniques such as Multivariate Adaptive Regressive Splines (MARS) to develop these relationships. Over 335,000 track geometry defect records and 26,000 rail defect records were analyzed.

The result was the development of a set of analysis algorithms relating track geometry data with the probability of the development of critical rail defects and the corresponding reduction in the rail life. It was found that there was a well-defined statistically significant relationship between geometry defects and rail defects, when the geometry defect preceded the removal of the rail defect. The overall matching of rail defects with track geometry defect at the same location was 11%, of which 38% matched multiple track geometry defects. Looking at detail fractures (TDD) defects only, the largest category of rail defects, 15% of rail defects were preceded by one or more track geometry defects at the same location, of which 41% matched multiple track geometry defects. On curves, the overall matching of rail defects with track geometry defect at the same location was 21%, of which 46% matched multiple track geometry defects. For detail fracture (TDD) defects on curves, the overall matching was 30.5% or almost 1 in 3 TDD defects.

In addition, the analyses developed a strong relationship between the presence of a track geometry defects and the reduction in the life of the rail, as defined by cumulative tonnage (age in MGT) of the rail at the time of failure or detection and removal. This relationship showed that when a rail defect is preceded by a track geometry defect, the age of the rail (in cumulative MGT) is 28% less than that of a rail defect that failed with no track geometry defect at that location.

It is expected that these algorithms can help railroads refine their inspection techniques and programs by identifying locations with the potential of early or premature development of high risk rail defects.

## 1. Introduction

For many years, railway engineers have hypothesized a relationship between track geometry defects and rail defects. The basic reasoning has been that the presence of one of more track geometry defects will generate increased dynamic wheel/rail loads, which in turn cause can accelerated rail damage, leading to the development of rail defects. While some evidence of this relationship has been demonstrated for the development of rail surface defects, little or no evidence is available linking the presence of track geometry defects to the development of internal rail defects. A recent US Federal Railroad Administration

(FRA)[1] sponsored study looked at the relationship between the presence of one of more track geometry defects and the development of rail defects at that same location, after the occurrence of the geometry defects. This is a relationship that basic track engineering theory has suggested but which has never been proven or validated.

Theoretical research has shown that the presence of geometry defects generates increases dynamic wheel/rail loads 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. This includes both bending stresses and contact stresses, both of which have an effect on the development of rail defects.The objective of this study was to validate this theoretical hypotesis and to determine the increased probability ("risk") of a rail defect developing with geometry defects and the associated reduction in rail life.

In order to examine this relationship, the study correlated multiple years of track geometry with a data base of several years of rail defects obtained from a major US Class 1 railroad. The railroad system data represented more than 35,000 track 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 Million Gross Tonnes or 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.

## 2. Discussion of Analysis

An initial correlation was performed between rail defects and geometry defects for the full system (35,000 km) and for the "high tonnage" segments, defined here as having a minimum annual tonnage of 18 Million Gross Tonnes (MGT)[2]. As can be seen from Table 1, for the full system, 11% of all rail defects were preceded by one or more track geometry defects. For track with greater than 18 MGT annual tonnage, this percentage increases to almost 12%.

In contrast, if the relationship between rail defects and geometry defects were purely random 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"). The higher density ( > 20 MGT) track showed a similar behavior.

On curves, 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 (Table 2).

---

[1] INTEGRATION OF MULTIPLE INSPECTION SYSTEM DATA TO IDENTIFY POTENTIALLY UNSAFE TRACK RAIL CONDITIONS, FRA BAA Contract DTFR53-13-C-00066, October 2015

[2] 1 Million Gross Tonnes is equivalent to 1.1 Million Gross Tons.

| | Length in km | Annual MGT | Reported Geo Defects | Unique Geo Defects | Rail Defects | Matches | **% of All Matched** |
|---|---|---|---|---|---|---|---|
| Full System | 35,565 | 19.4 | 334937 | 202341 | 26440 | 2918 | **11.04%** |
| High density segments | 17,090 | 32.6 | 173314 | 104952 | 15428 | 1820 | **11.80%** |

Table 1: Summary of Rail and Geometry Defects and Matches

| | Rail Defects | Rail Defects on Curves | Matches | Matched on Curves | **% of All Matched** |
|---|---|---|---|---|---|
| Full | 26440 | 8870 | 2918 | 1871 | **21.09%** |
| High density | 15428 | 5225 | 1820 | 1113 | **21.30%** |

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

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 this approach, the calculated reduction in rail life for the full system , all tracks, was 31% when geometry defects are present. For high density track (> 18 MGT ) the reduction was 25%.

A series of statistical analyses were performed to relate “age” of the defect, in MGT[3], with the presence of the different track geometry defects. The statistical analyses performed included Regression Analysis and Multivariate Adaptive Regressive Splines (MARS) analysis.

The initial regression analyses looked at all geometry defects and developed a set of defect life equations of the form shown in Equation 1:

$$\boldsymbol{MGT=1750(0.8435+ \Sigma a_i GD_i)}$$

| $GD_i$ | $a_i$ |
|---|---|
| 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%

Equation 1: Regression Analysis- All Defects

---

[3] The regression and MARS equations were calculated in terms of Million Gross Tons.

Application of this equation projected a reduction in defect life of 19%, if an alignment defect is present, a reduction of 44%, if there is a warp defect present, and a reduction of 42% if there is a cant defect present. If there is both an alignment and cross-level defect present, the life is reduced by 57%.

This regression analysis approach has several major limitations, including an inability to handle multiple geometry defects of the same type. In order to address this and several other limitation of the standard regression model, the Multivariate Adaptive Regressive Splines (MARS) analysis approach was used. The MARS analysis is a high end statistical analysis approach for dealing with very large data bases. It identifies and focuses on the most important contributing variables rather than looking at all variables.

The resulting MARS Relationship for all rail defects was calculated as shown in Equation 2:

$$MGT = 477.37 + 86.6031(BF_2) + 28.6544(BF_4) - 47.8978(BF_5)$$

$$BF_2 = \max(0, 6 - WARP)$$

$$WARP = \text{number of WARP Defects}$$

$$BF_4 = \max(0, 18 - RAIL\_CANT)$$

$$RAIL\_CANT = \text{number of rail cant defects}$$

$$BF_5 = \max(0, ALIGNMENT - 0)$$

$$ALIGNMENT = \text{number of alignment defects}$$

$$\text{Max } MGT\ (\text{no defects}) = 1513\ MGT$$

Equation 2: MARS Relations for All Rail Defects vs. Geometry Defects

The MARS analysis was able to identify key geometry defects, to include those geometry defects that are most important in the reduction in defect life for tangent and curve track as well as the full track set shown above. These Key variables were identified as follows:

| **Tangent Track** | **Curve Track** |
|---|---|
| Warp 62 | Warp 31 |
| Rail Cant | Rail Cant |
| Gage | Alignment |
| Cross-Level | Cross-Level |

Probability Analysis examined the probability of a rail defect occurring given a geometry defect preceding it. The Probability analysis approaches used included:

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

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. For the data set used, the probability of the rail defect occurring in any one location randomly was calculated to be 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. However, as seen from the correlation analysis, 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 was calculated using Bayesian Network Analyses. Bayesian network models are graphical probabilistic models used for a set of random variables and their conditional probability. Table 3 presents the results of this probabilistic analysis.

| | Entire Railroad (random = 0.18%) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Alignment | Crosslevel | Gage | Rail Cant | Warp 31 | Warp 62 | P(RD\|GD) Bayesian Network |
| **# of defects** | 1 | 0 | 0 | 0 | 0 | 0 | 1.30% |
| | 0 | 1 | 0 | 0 | 0 | 0 | 1.17% |
| | 0 | 0 | 1 | 0 | 0 | 0 | 1.04% |
| | 0 | 0 | 0 | 1 | 0 | 0 | 1.51% |
| | 0 | 0 | 0 | 0 | 1 | 0 | 1.48% |
| | 0 | 0 | 0 | 0 | 0 | 1 | 1.36% |
| | 1 | 1 | 0 | 0 | 0 | 0 | 8.54% |
| | 1 | 0 | 1 | 0 | 0 | 0 | 7.66% |
| | 1 | 0 | 0 | 1 | 0 | 0 | 10.80% |
| | 1 | 0 | 0 | 0 | 1 | 0 | 10.60% |
| | 1 | 0 | 0 | 0 | 0 | 1 | 9.87% |
| | 0 | 1 | 1 | 0 | 0 | 0 | 6.94% |
| | 0 | 1 | 0 | 1 | 0 | 0 | 9.82% |
| | 0 | 1 | 0 | 0 | 1 | 0 | 9.63% |
| | 0 | 1 | 0 | 0 | 0 | 1 | 8.96% |
| | 0 | 0 | 1 | 1 | 0 | 0 | 8.82% |
| | 0 | 0 | 1 | 0 | 1 | 0 | 8.64% |
| | 0 | 0 | 1 | 0 | 0 | 1 | 8.04% |
| | 0 | 0 | 0 | 1 | 1 | 0 | 12.10% |
| | 0 | 0 | 0 | 1 | 0 | 1 | 11.30% |
| | 0 | 1 | 1 | 0 | 0 | 1 | 38.30% |
| | 0 | 1 | 0 | 1 | 1 | 0 | 49.60% |
| | 1 | 0 | 0 | 1 | 1 | 0 | 52.20% |
| | 0 | 1 | 1 | 1 | 0 | 1 | 85.10% |
| | 1 | 1 | 0 | 1 | 1 | 0 | 88.60% |

**Table 3.** Bayesian Network Model Probability Analyses for Select Geometry Defects

Thus, as can be seen in Table 3, a single geometry defect increases the probability of a rail defect (from random) by 6 to 9 times (to 1 to 1.5%), 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 to12 % 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%.

This latter case is illustrated in the Bayesian network model structure presented in Figure 1, when there is a warp 31, rail cant, cross-level and alignment defect all present ( though not necessarily at the same time) prior to the development of the rail defect. The probability of a rail defect occurring is 88.6% ( about 500 times random).

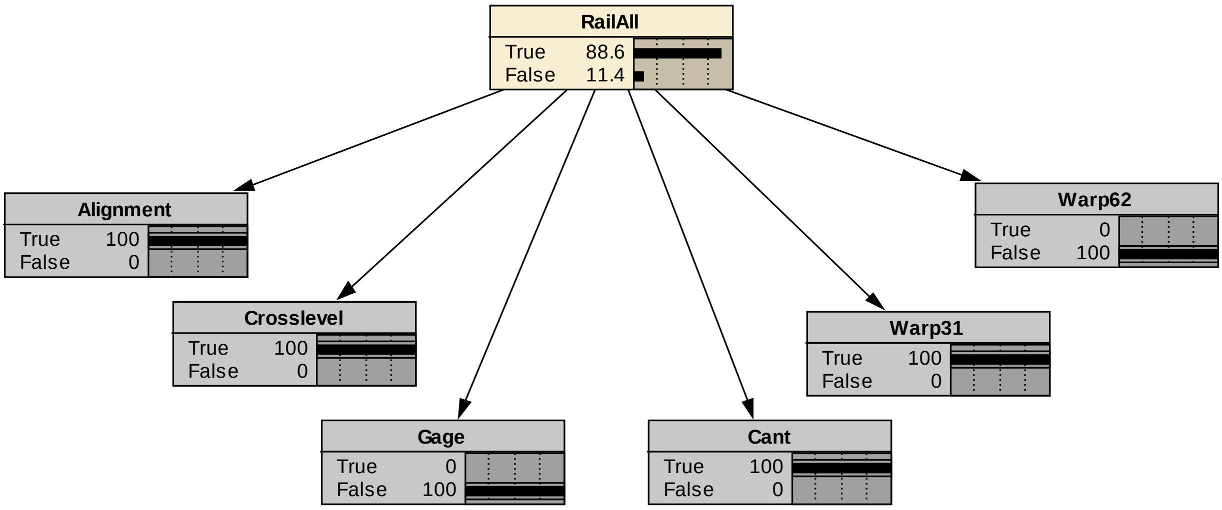

**Figure 1.** Bayesian Network Model results for warp 31, rail cant, cross-level, and alignment defects

From these analyses, it be clearly seen that geometry defects have a significant impact on the probability of development of a rail defect. Furthermore, the occurrence of multiple geometry defects, prior to the development of rail defects is very common, based on the Class 1 data. This is illustrated in Table 4, which shows that of the 1119 occurrences, where two or more geometry defects were present at the same location as (and prior to) the rail defect, 44% 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 4: Number of Multiple Geometry Matches

## 3. Results and Conclusion

The results of the study showed that 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%, of which 38% of these were preceded by two or more track geometry defects. On curves only, the overall matching of rail defects (all defects) with track geometry defect was 21%, with 46% of these preceded by two or more track geometry defects at the same location.

Analysis of the “life” ( in cumulative MGT) of rail defects at the time of failure when matched with a preceding track geometry defect shows that the age of the rail at the time of failure (in cumulative MGT) is approximately 30% less than that of a rail defect that failed with no track geometry defect at the same location.

The key geometry defect drivers for rail behavior as defined by the MARS (and Regression analyses) were found to be:

- For Tangent Track: Cant, Warp 62, Cross-level, and Gage
- For Curve Track: Warp 31, Cant, Alignment and Cross-level

Results of the probability analysis examining the probability of a rail defect occurring given a geometry defect preceding it showed that the presence of a geometry defect has a strong and well defined effect, with the probability of a rail defect occurring at the location where there was one of more preceding geometry defects being strong and significant. 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 600 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 “data science” tools (e.g. MARS, Bayesian analysis, etc.) within the Big Data paradigm 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.