Rail Fatigue Life Forecasting Using Big Data Analysis Techniques

UTC Spotlight Conference: Rebuilding and Retrofitting the Transportation Infrastructure

Dr. Allan M Zarembski PE, FASME, Hon. Mbr. AREMA

Dr. Nii Attoh-Okine, PE

John Cronin, MCE, EIT

RAIL TEAM

Overview

- US Railroads spend on the order of $11 Billion on Track and Property
- Rail Fatigue is one of the Primary causes of rail failure
- Weibull Analysis allows determination of the rate of failure
- Recent advances in “Big Data” present the option of refining the Weibull methodology
- Better prediction means less waste, saving money or allowing more to be repaired

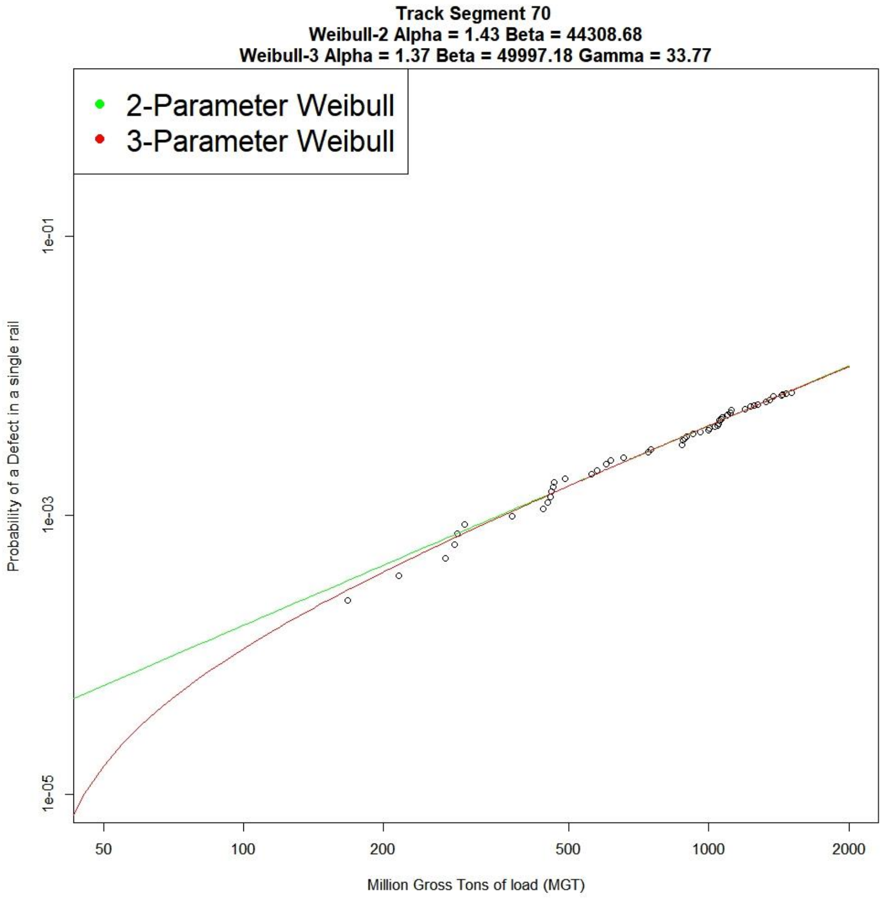
RAIL TEAM

Data and Baseline Analysis

- 10 years of data from a Class 1 railroad
  - 11,000 miles of track
  - 98,000 defects, of which 41,000 are fatigue defects
  - Rail Grinding, Curvature and Location data included
- Traditional 2-parameter Weibull Analysis has been used by railroads to forecast rate of fatigue defect development
- Data does not always follow traditional Weibull expectations
  - Various reasons why: data collection methods, errors in recording
- Enhanced 3-Parameter Weibull Analysis applied
  - 3rd Parameter is an intercept/"failure free" period
  - Often results in better fits, but justification behind the use is still in debate

RAIL TEAM

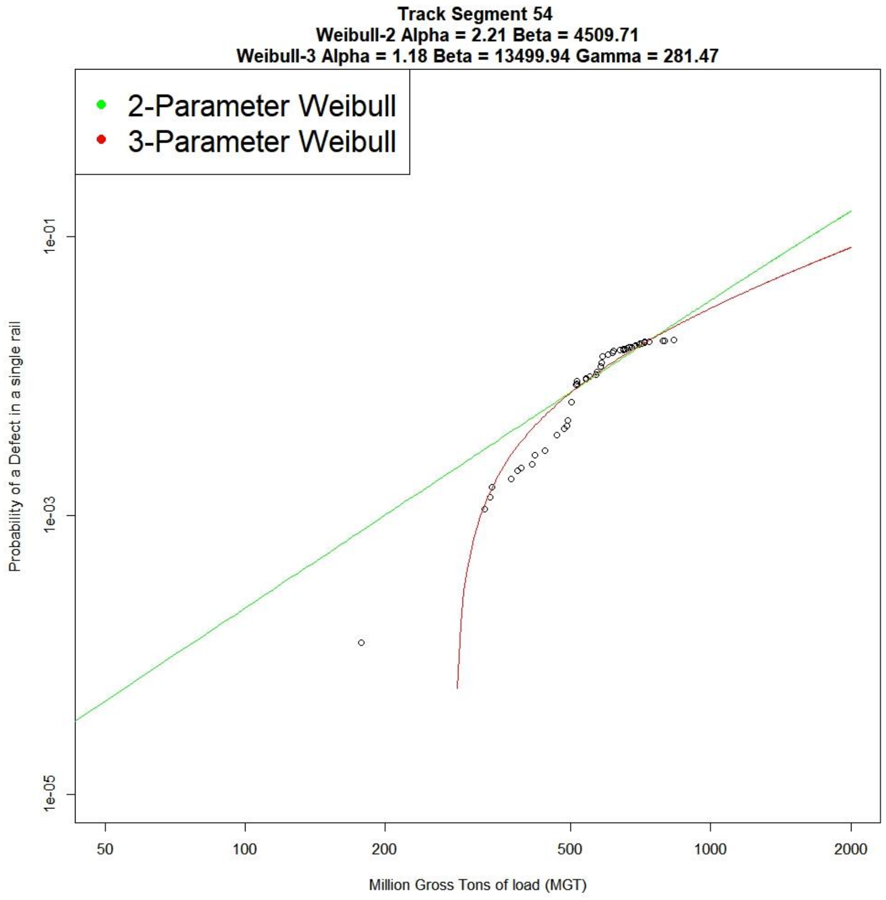
Track Segment A



- Shows a good relationship to the 2-parameter Weibull Function
- 3-parameter Weibull overlaid with similar results
- Several small “jumps” observed; corresponding to several defects occurring at the same time (tonnage level of MGT)

RAIL TEAM

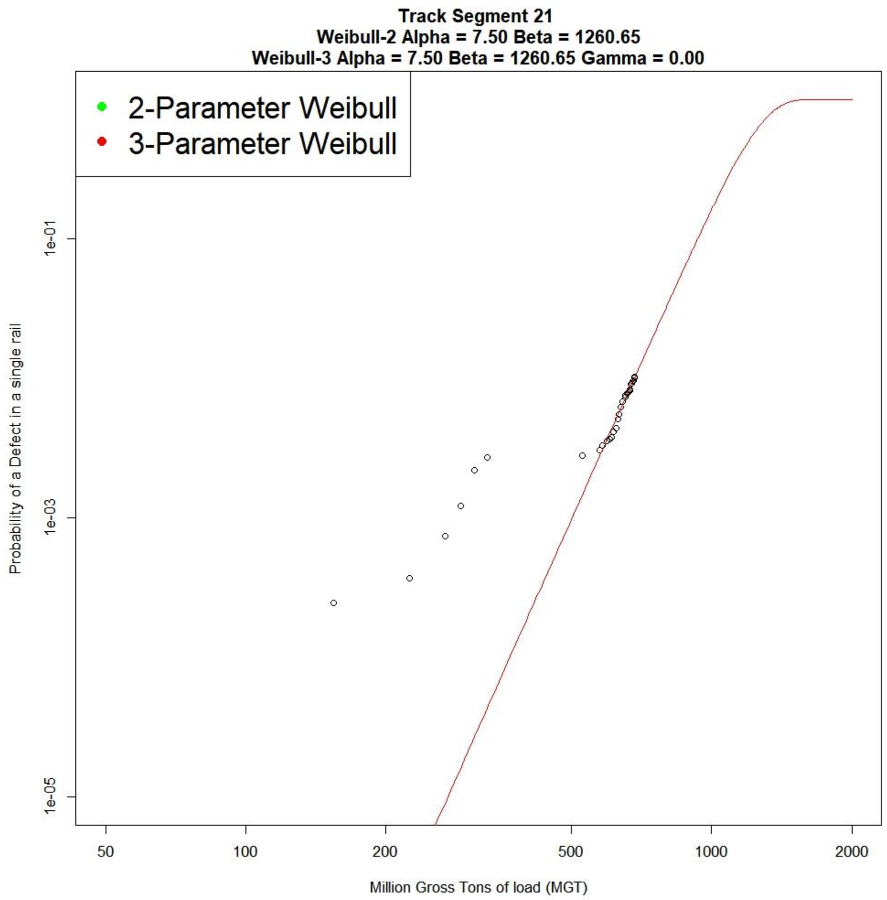
Track Segment B



- “Steps”/”Jumps” in data are more noticeable
- 3-Parameter Weibull fits better, but cannot be rationalized

RAIL TEAM

Track Segment C



- 2-parameter and 3-parameter Weibull curves are identical
- "Early"/Low MGT defects do not appear to follow Weibull
- "Late"/High MGT defects follow Weibull much better
- Possibly a result of two different defect types

RAIL TEAM

Literature Search

- To date, over 60 papers were reviewed
- One approach looked at the applications of hazard functions in the Weibull equation
  - Allows a better fit when two defect populations result in “jumps” or “steps” similar to Segments B and C.
- Rare-Event prediction with Big-Data may allow more precise pinpointing of future defects
  - Instead of track segments on the order of 30 miles, could reduce down to <1 mile
- Expanding the Weibull function to account for situational variations
  - Follows other work that generalized the Weibull function up to a 5-parameter method

RAIL TEAM

Early Machine Learning Tests

- K-Nearest Neighbors was applied to the data
  - Looked at predicting the state of individual track segments approx. 1 mile in length vs. Weibull's ~30 mile segments
  - Included tests to see if increasing years of historical data improved results
  - Due to majority of null-responses, accuracy was based on only positive predictions
  - Not an effective fit

RAIL TEAM

KNN Results Pt.1

Trial 1	Output 0	Output 1		Trial 2	Output 0	Output 1
Input 0	21739	93		Input 0	21832	0
Input 1	420	1		Input 1	421	0
False Positive Rate		0.426%		False Positive Rate		0.000%
False Negative Rate		99.762%		False Negative Rate		100.000 %

- Trial 1 was a basic KNN using a single year of historical data
  - Only one track segment was correctly determined
- Trial 2 expanded K to 6
  - Overabundance of null-responses resulted in everything being classed as Null

KNN Results Pt.2

Trial 3	Output 0	Output 1
Input 0	21751	261
Input 1	229	12
False Positive Rate		1.186%
False Negative Rate		95.021%

Trial 4	Output 0	Output 1
Input 0	21761	251
Input 1	225	16
False Positive Rate		1.140%
False Negative Rate		93.361%

- Trial 3 included Milepost as an additional location parameter
- Trial 4 dropped MGT as a parameter

RAIL TEAM

Future Work

- Apply methodologies found in Literature Search
- Expand Machine Learning tests to include other methodologies, such as Neural Networks
- Apply Big-Data methodologies to generate Weibull functions that are sensitive to key railroad variables
- Provide a detailed method of forecasting rail life due to fatigue and the improvements to maintenance efforts

RAIL TEAM

Acknowledgements

The authors would like to acknowledge the funding and support of the US Department of Transportation and the University Transportation Center program. They would also like to acknowledge its UTC team, RAILTEAM, with lead University of Nevada at Las Vegas and partner Virginia Tech.