A BAYESIAN SURVIVAL APPROACH TO ANALYZING THE RISK OF RECURRENT RAIL DEFECTS

Faeze Ghofrani, Reza Mohammadi, Qing He, Abhishek Pathak, Amjad Aref

INFORMS 2018

Phoenix, Arizona
November 6th, 2018

University at Buffalo The State University of New York

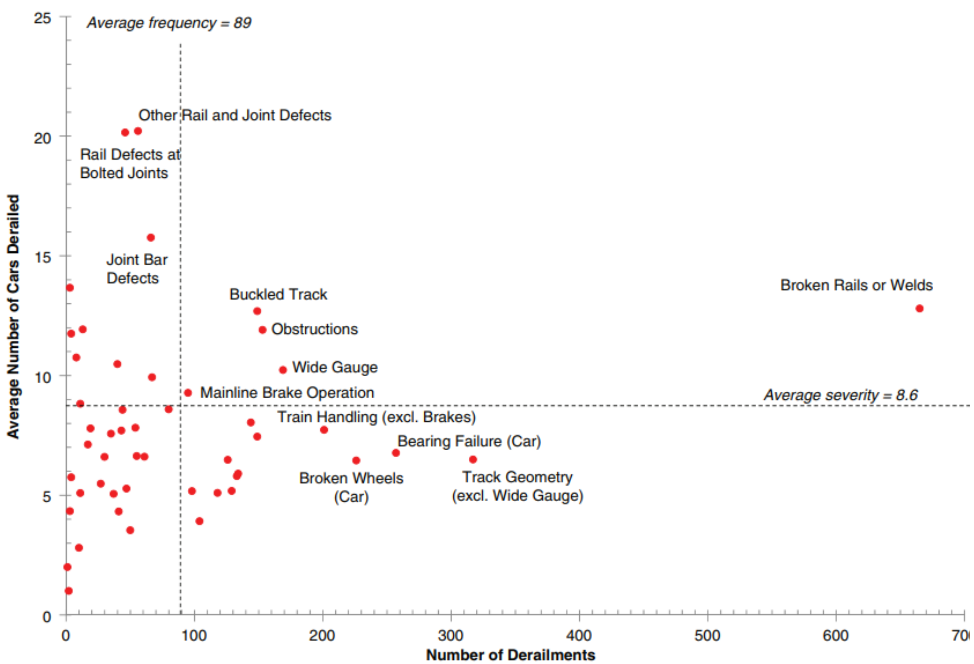
informs

CONTENTS

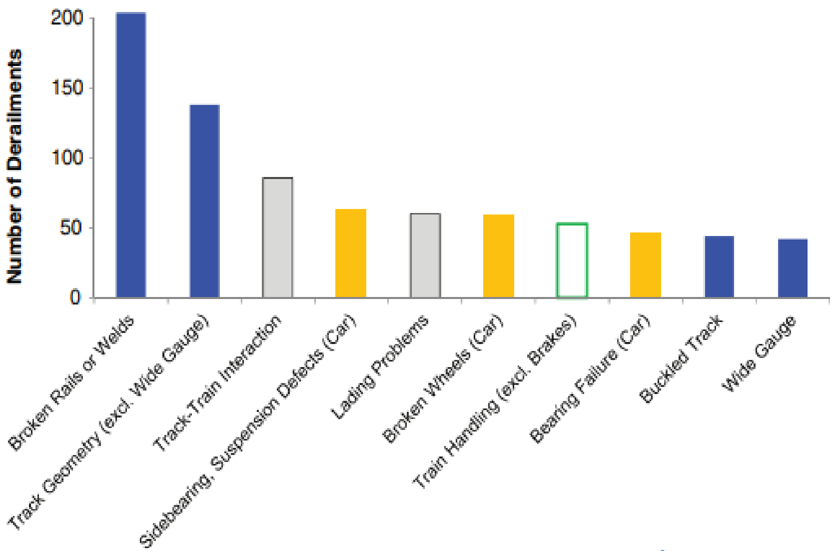
University at Buffalo The State University of New York

Background

- ➢ Rail defects (and more specifically broken rails) account for the largest portion of total train derailments causes
- ➢ Many of rail defects occur in the same location multiple times due to the characteristics of the location.
- ➢ Analysis of recurrent rail defects can assist in identifying network hotspots and providing on-time responsive maintenance.



Frequency and severity graph for causes of Class I main-line freight train derailments, 2001–2010 (Liu et al. 2012).



Number of freight train derailments by accident cause on Class I main lines (Liu et al. 2012).

University at Buffalo The State University of New York

Contributions of the Study

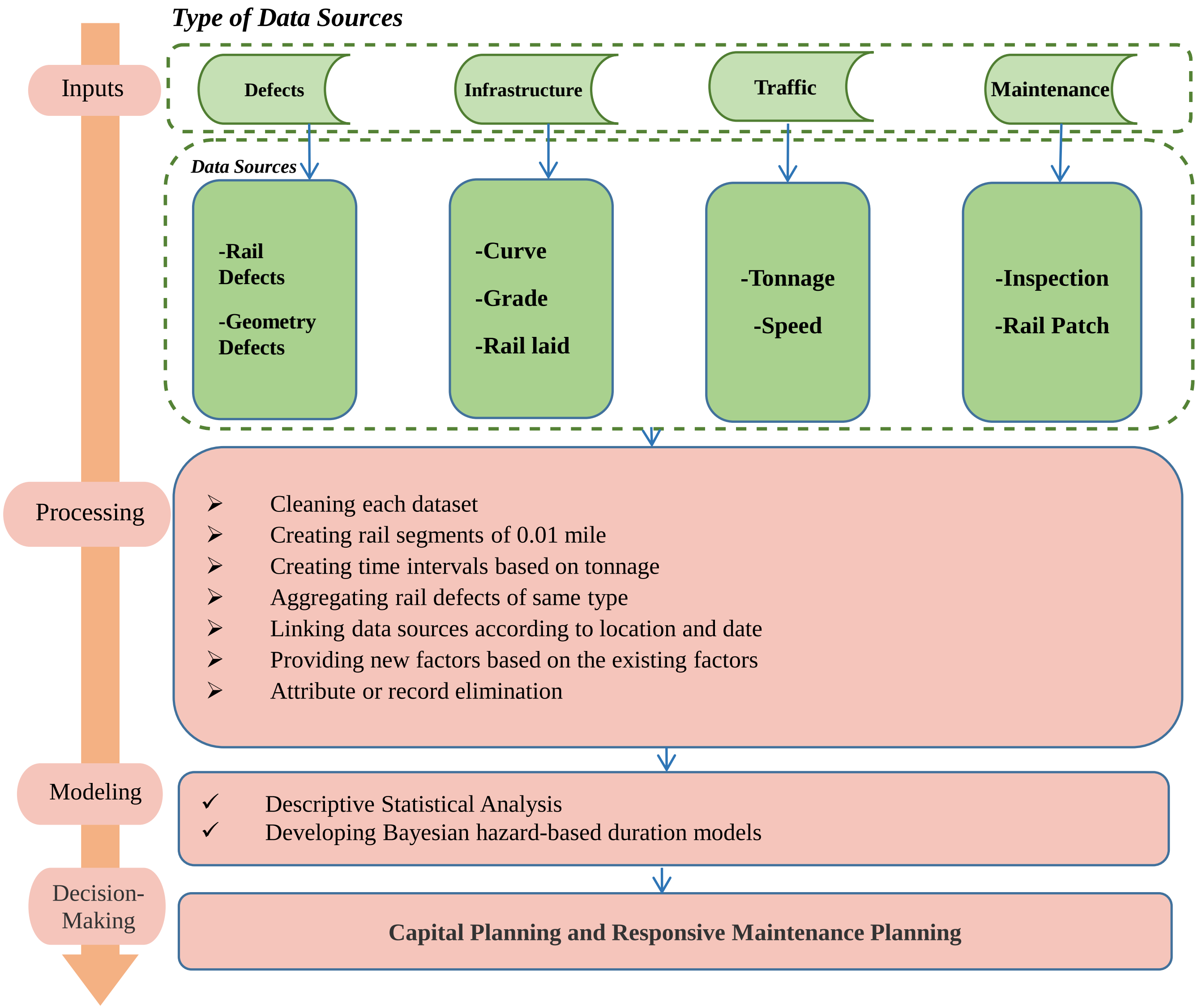
- Designing a comprehensive logical methodology framework for data collection, pre-processing, and modeling based on a collection of datasets from different resources in a Class I railroad

- Applying the correlated event times of survival analysis in the context of railway transportation for recurrent rail defects

- Developing a Bayesian framework by performing Markov Chain Monte Carlo (MCMC) simulation for optimizing the parameters of the model

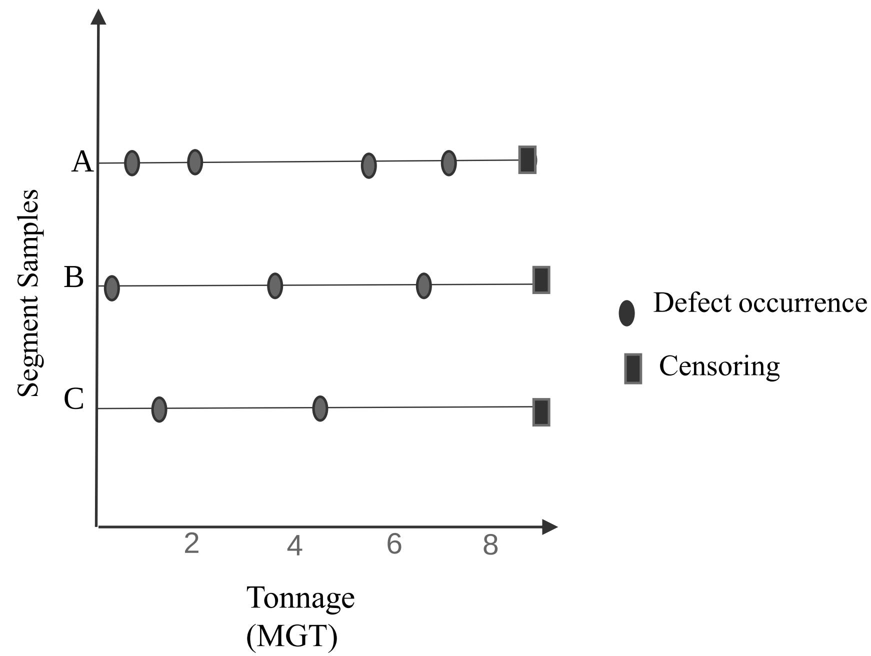
- Verifying the fit of the model by using Cox-Snell residual plot

The methodology framework



Data Pre-processing

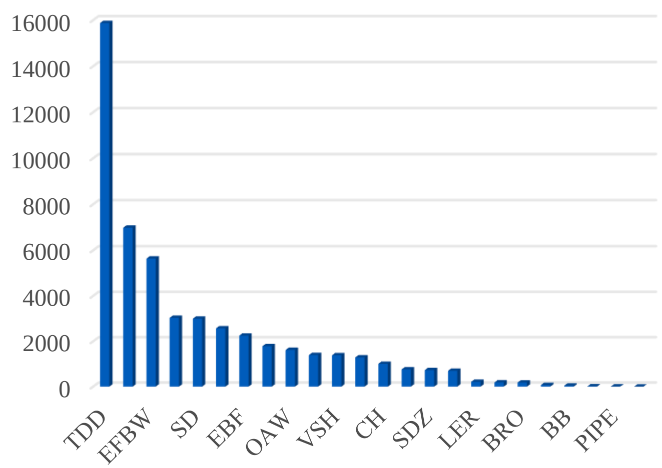
- Dividing the network into short sections (0.01 mile)
- Creating tonnage intervals for each segment
- Creating new variables for each interval
- Linking all datasets based on spatial units with rail defect dataset
- Record elimination



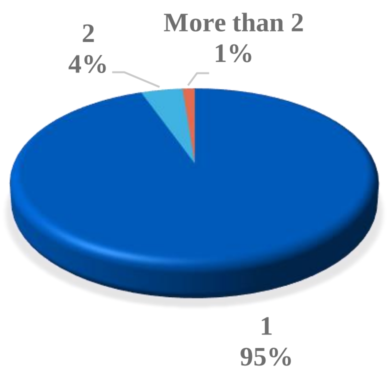
Prefix	Track Type	Milepost	Defect Type	ID	Interval_Number	Time_from	Time_to	Status	Start_Tonn	End_Tonn	Tonn_Gap
0	1	390.89	TDD	351	1	1/3/2011	2/12/2015	1	0	798768707.1	798768707.1
0	1	390.89	TDD	351	2	2/12/2015	5/14/2015	1	798768707.1	873199155	74430447.9
0	1	390.89	TDD	351	3	5/14/2015	5/14/2015	1	873199155	875232759.1	2033604.1
0	1	390.89	TDD	351	4	5/14/2015	12/30/2016	0	875232759.1	1221722965	346490205.9
00C	SG	249.14	CH	7791	1	1/3/2011	8/8/2011	1	0	18186814.97	18186814.97
00C	SG	249.14	CH	7791	2	8/8/2011	10/13/2011	1	18186814.97	25629467.96	7442652.99
00C	SG	249.14	CH	7791	3	10/13/2011	12/13/2011	1	25629467.96	35059404.49	9429936.53
00C	SG	249.14	CH	7791	4	12/13/2011	2/17/2012	1	35059404.49	43349418.76	8290014.27
00C	SG	249.14	CH	7791	5	2/17/2012	12/30/2016	0	43349418.76	240865278.6	197515859.8
0	1	2.023	SD	101	1	1/3/2011	9/9/2016	1	0	75235419.02	75235419.02
0	1	2.024	SD	101	2	9/9/2016	10/7/2016	1	75235419.02	77930334.73	2694915.71
0	1	2.024	SD	101	3	10/7/2016	12/30/2016	0	77930334.73	81864102.31	3933767.58

Descriptive Statistical Analysis

Attribute	Explanation
Def_Type	Type of defects (BHB, EBF, FBW, TDD, etc.)
Geo_Def	Number of geometry defects in the last three years
Freight_Speed	Freight train speed at the location of the defect
Def_Freq	The frequency of defects occurred in the same location
Inspection_Freq	Number of inspections in the last three years
Curve/Tang	Whether the occurred defect is on the low side of a curve, high side of a curve or on tangents
Weight	The weight of rail at the location of the defect
MGT	The tonnage of freight (in MGT) at the location of the defect



Type of defects

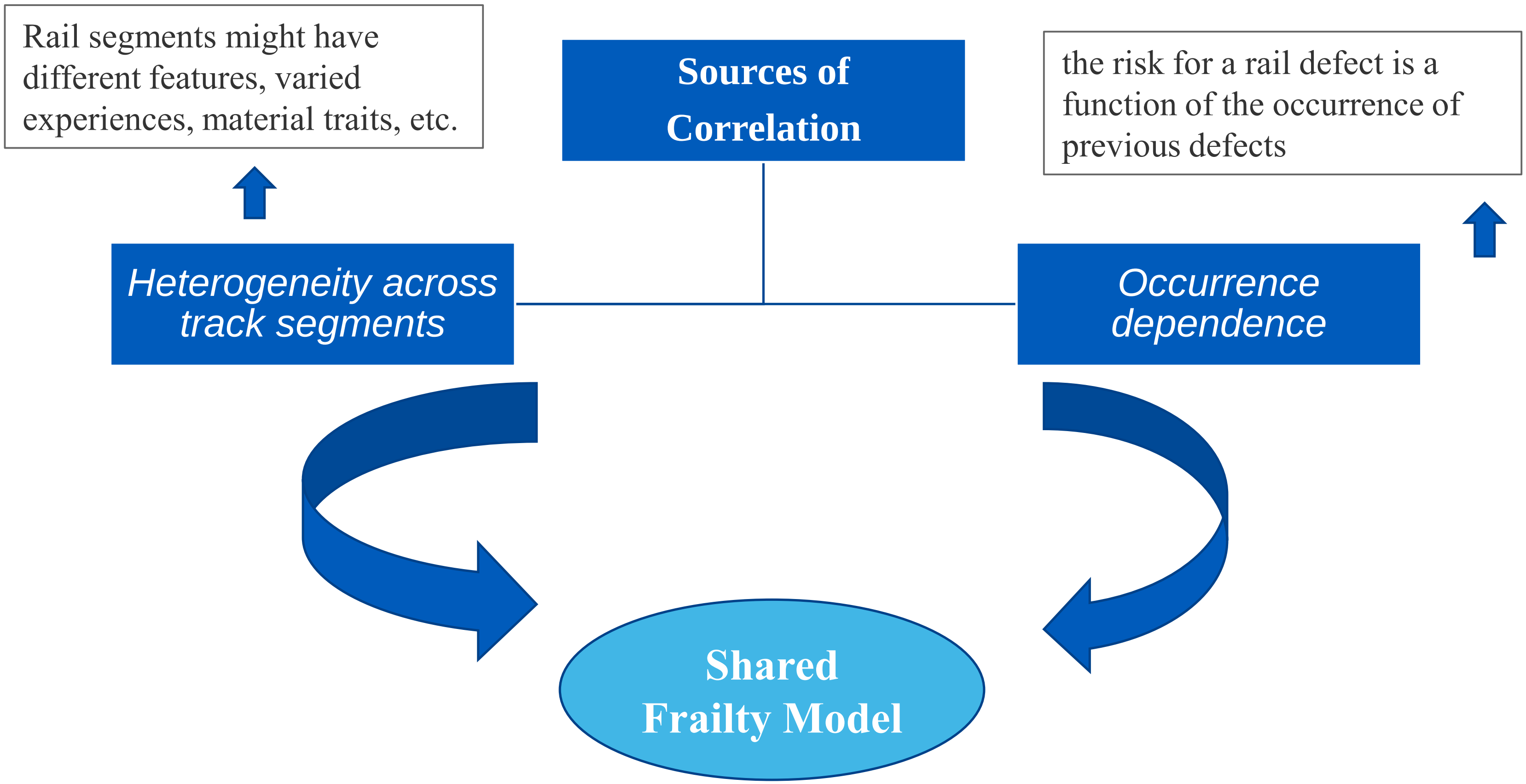


Segments with one, two and more than two recurrent defects

Modeling Approach

Model Types	When Used?	Intensity Function	Examples
Event Counts Models	▪ Individuals frequently experience the events of interest ▪ Events are incidental	$\lambda(t\|H(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{\Delta N(t) = 1\|H(t)\}}{\Delta t}$	Markov Processes, Poisson Models and their extensions
Gap Time Models	▪ Events are relatively infrequent ▪ Some type of individual renewal occurs after an event ▪ Prediction of the time to the next event is of interest.	$\lambda(t\|H(t)) = h(t - T_{N(t^-)})$	Renewal Processes and their extensions

Correlated event times



Shared Frailty Model Specification

Baseline hazard function : ***Weibull*** distribution with parameter μ and γ

$$h(t) = \mu\gamma t^{\gamma-1}$$ *hazard function*

$$S(t) = e^{-\mu t^{\gamma}}$$ *survival function*

$$f(t) = \mu\lambda t^{\gamma-1} e^{-\mu t^{\gamma}}, \mu > 0, \gamma > 0$$ *density function for Weibull distribution*

Likelihoods for Proportional Hazard (PH) model under the frailty approach

$$L_{PHFM} = \prod_{i=1}^{n} \left[\mu\gamma t_i^{\gamma-1} e^{\beta' x + z_i}\right]^{\delta_i} e^{-\mu t^{\gamma} e^{\beta' x + z_i}}$$

z_i: shared frailty variable with normal prior ($z_1, \ldots, z_m \sim N(0, \tau^2)$) and density function:

$$f(Z) = \frac{1}{\tau\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{Z}{\tau}\right)^2} \quad ; -\infty<Z<\infty \text{ , } \tau>0$$

Prior Specification

$\tau^{-2} \sim \Gamma(a_\tau, b_\tau)$, $\beta \sim N(0,m)$, $\mu \sim \Gamma(\rho,\rho)$, $\gamma \sim \Gamma(a,b)$

Posterior Calculation and parameter configuration

$$P_{PH} = \prod_{i=1}^{n} [\mu\gamma t_i^{\gamma-1} e^{\beta X + Z_i}]^{\delta_i} e^{-\mu t^{\gamma} e^{\beta X + Z_i}} \pi(z)\pi(\mu)\pi(\gamma)\pi(\beta)\pi(\tau^2)$$

MCMC calculations have been implemented with burn-in period of 5,000 iterations and the Markov chain subsample to get a final chain size of 4,000 iterations

Results

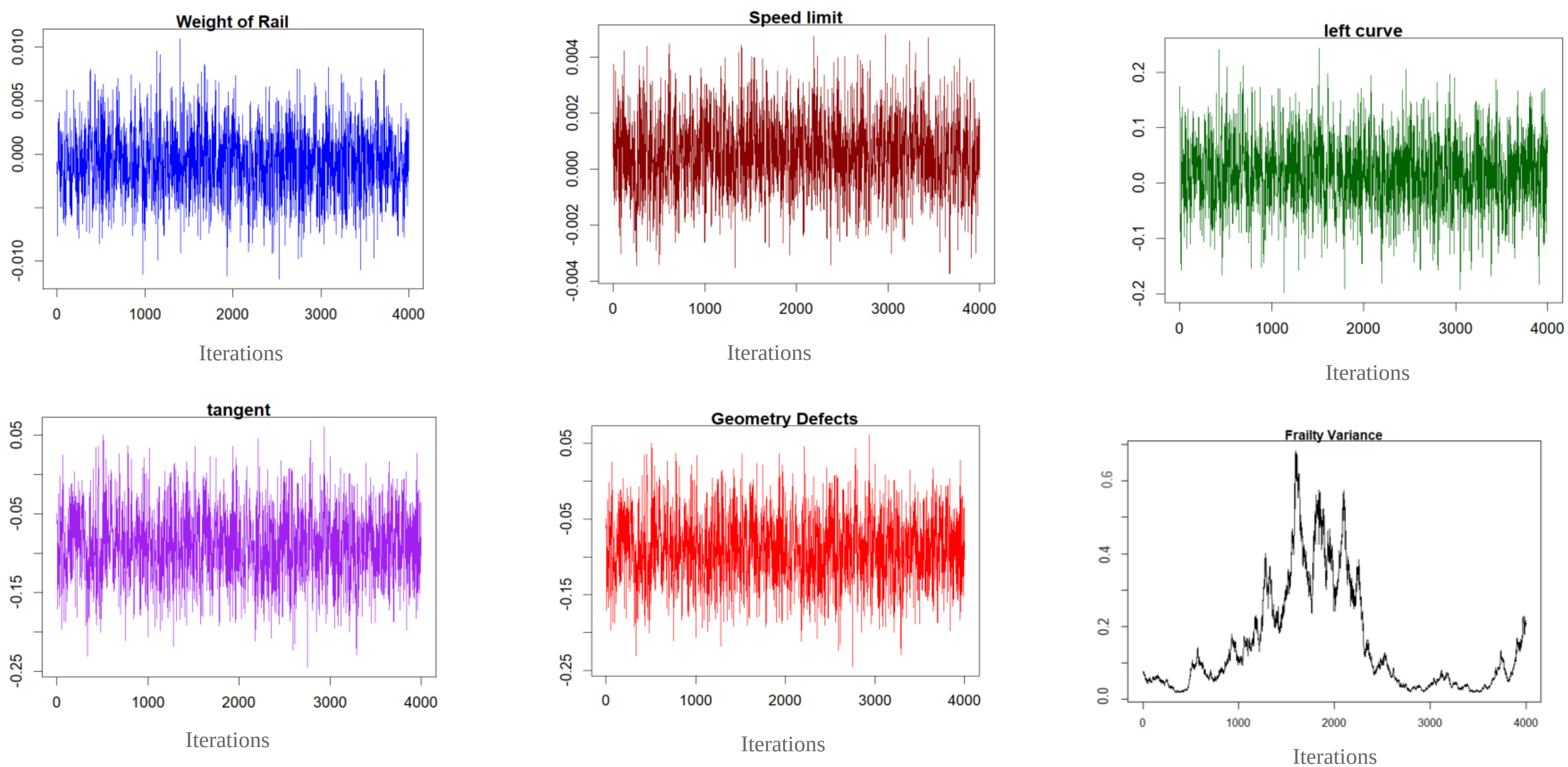
Posterior Inference of Regression Coefficients and frailty variance

Variables	Mean	Median	Std. Dev.	95% CI-Low	95% CI-Upp
Weight of Rail	-0.0006	-0.0005	0.0026	-0.0058	0.0044
Frequency of Geometry Defects	0.0123	0.0124	0.0082	-0.0036	0.0283
Location of Rail (L)	-0.0320	-0.0306	0.0527	-0.1426	0.0689
Location of Rail (T)	-0.0551	-0.0546	0.0320	-0.1187	0.0083
Freight Speed Limit	0.0014	0.0015	0.0012	-0.0010	0.0038
Type of Defect (TDD)	0.1693	0.1691	0.0674	0.0369	0.3074
Type of Defect (TW)	0.1999	0.1994	0.0721	0.0603	0.3370
Type of Defect (VSH)	0.2956	0.2976	0.1202	0.0549	0.5222
Posterior inference of frailty variance	**0.491**	**0.271**	**0. 040**	**0.006**	**0.891**

There is a positive relationship between the occurrence of the defects on the same rail segment;

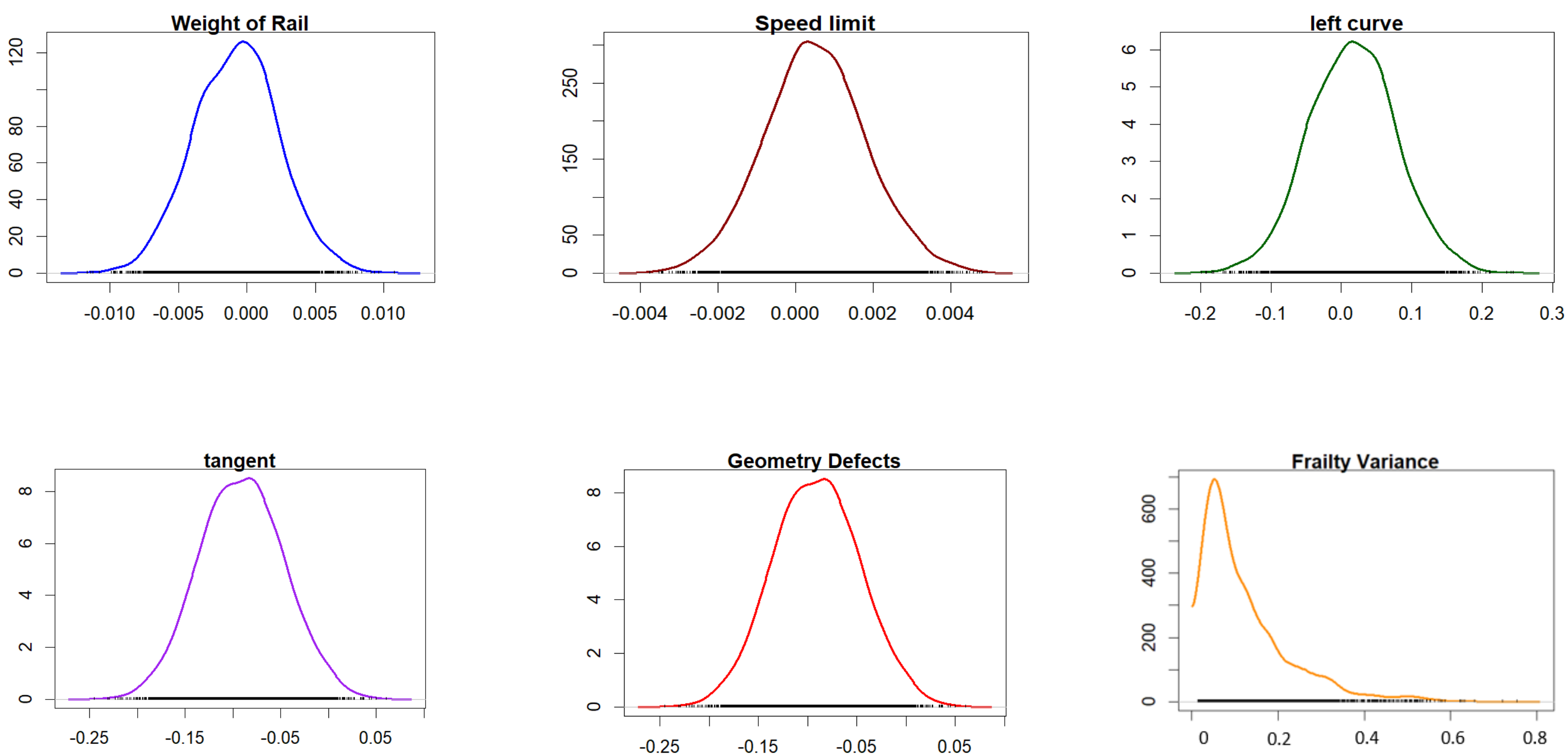
Trace plots for the regression coefficients and frailty variance

- The parameters have sufficient state changes as the MCMC algorithm runs
- This implies that priori distribution is well calibrated



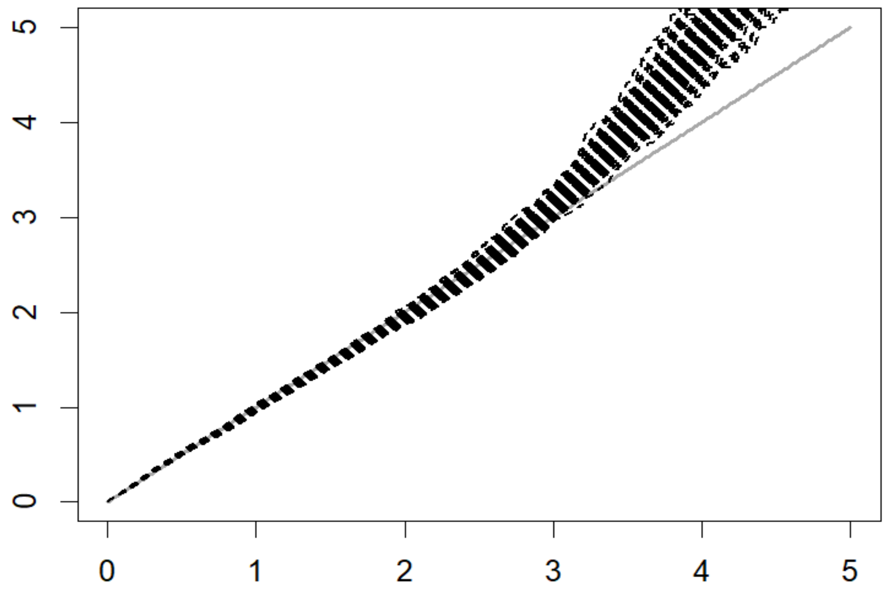
Density plots for posterior distribution of the regression coefficients and frailty variance

- The estimates of the posterior marginal distribution for the coefficients have smooth and unimodal shapes



Model Diagnostics

Estimated cumulative hazard function versus the Cox and Snell residual



Cox Snell Plot for the fitted model

Residuals from a correctly fitted model follow unit exponential distribution

Findings

The following factors increase the risk interval for defect recurrences (decrease the hazard):

- Weight of rail
- Location of defect (if low side of the curve or tangent) compared to high side of the curve

The following factors decrease the risk interval for defect recurrences (increase the hazard):

- Number of geo defects in the last three years
- Freight train speed limit
- “TDD, TW and VSH” types of defect compared to “SD”

Conclusions

- Data collection from a Class I railroad
- Cleaning, fusion, pre-processing and restructuring of multiple datasets
- Study of models for recurrent survival analysis
- Provide a model for predicting the tonnage until the re-occurance of a rail defect
- According to findings railroad could record the data on lighter segments of rail as well as on the high side of the curves on rail segments to be more certain on the risk of defect recurrence.
- The impact of past geometry defects can be reduced by minimizing their occurrence either by accurately predicting the defects or by decreasing the inspection intervals
- The impact of the speed can be reduced by decreasing the speed limit on certain segments.
- The rail segments with higher risk of frequent rail defects shall be treated more frequently.

Thank you!

faezegho@buffalo.edu

References

Cook, R. J., & Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

Schafer, D. H., & Barkan, C. P. L. (2008). A Prediction Model for Broken Rails and an Analysis of their Economic Impact. *Proceedings of the American Railway Engineering and Maintenance-of-Way Association Annual Conference*, *252*(847).

Chattopadhyay, G., & Kumar, S. (2009). Parameter estimation for rail degradation model. *International Journal of Performability Engineering*, *5*(2), 119–130.

Zarembski, A. M., Einbinder, D., & Attoh-Okine, N. (2016). Using multiple adaptive regression to address the impact of track geometry on development of rail defects. *Construction and Building Materials*, *127*, 546–555.

Sadeghi, J., & Askarinejad, H. (2010). Development of improved railway track degradation models. *Structure and Infrastructure Engineering*, *6*(6), 675–688.

Box-Steffensmeier, J. M., & De Boef, S. (2006). Repeated events survival models: the conditional frailty model. Statistics in medicine, 25(20), 3518-3533.