

# Modeling the Useful Residual Life of Railroad Grease

Dr. Doug Timmer, Thania Martinez, Dr. Robert Jones,  
Dr. Constantine Tarawneh

University of Texas – Pan American

# Project Description

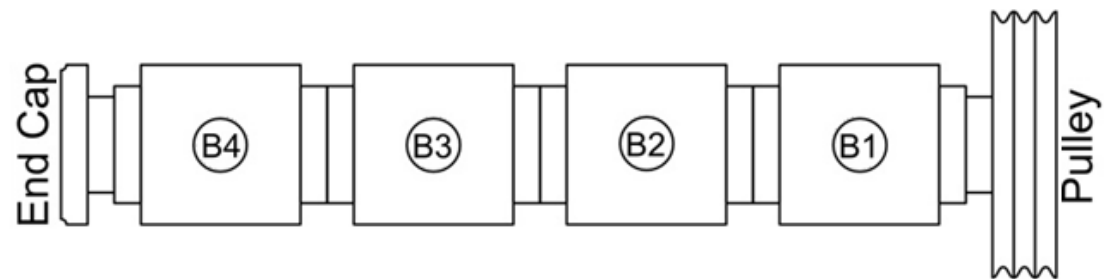
- The degradation of grease used to lubricate railroad bearings is believed to occur due two processes:
  - Mechanical processes occurring within the bearing,
  - Oxygen diffusion.
- Appropriate lubrication of the bearings is critical during railroad service operation.
- This study focuses on the development of empirical models that can accurately predict the residual useful life of railroad bearing grease.
- Employed Modeling Techniques:
  - Linear Regression Analysis
  - Regression Trees
  - Split Plots

## Project Description (cont.)

- The data set used in the development of the model consists of more than 100 samples of grease taken from the railroad bearings which were observed in a laboratory setting.

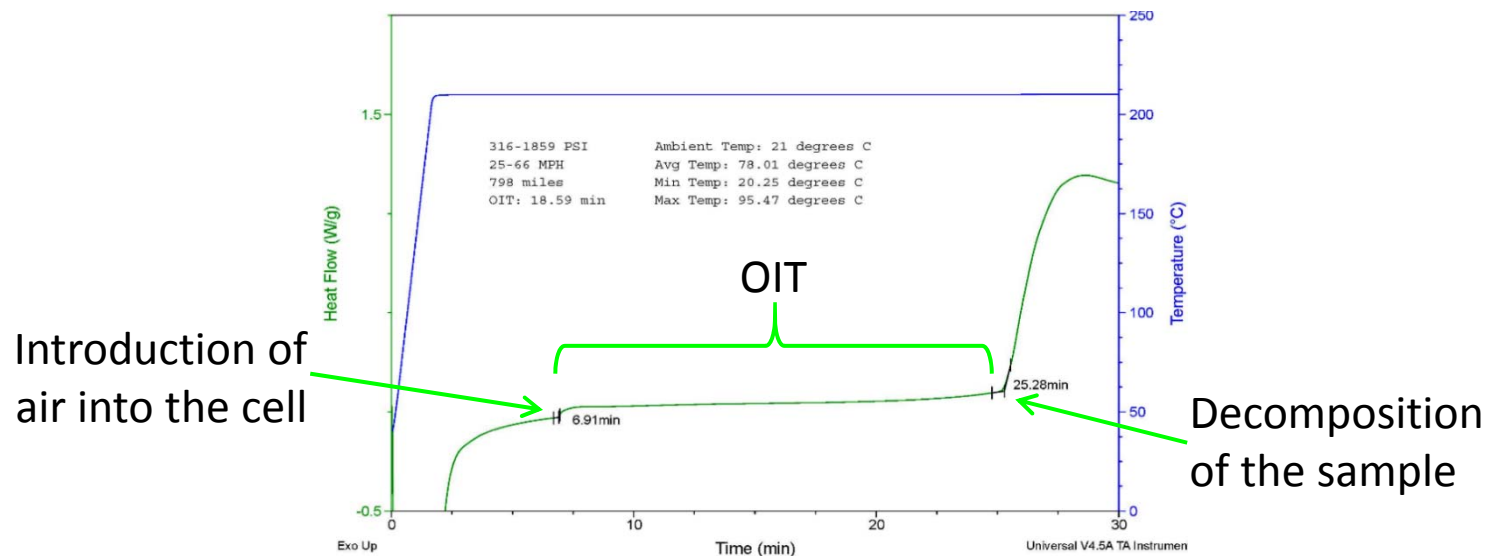
# Laboratory Bearing Tester

- Four bearings on the axle are subjected to the following experimental variables:
  - Load Conditions
  - Rotational Speed
  - Mileage
  - Temperature



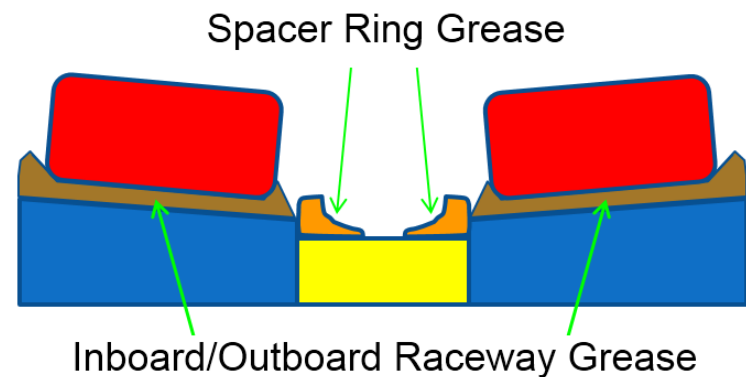
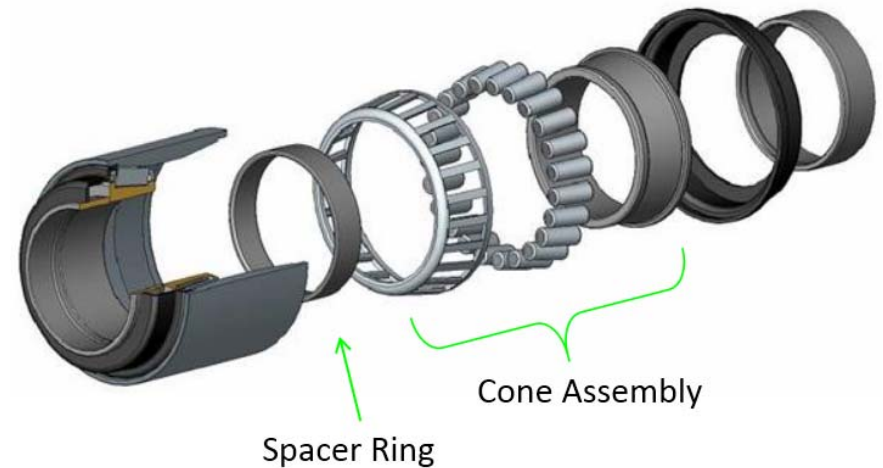
# Oxidation Induction Time

- **Oxidation Induction Time (OIT)** is a test performed in a Differential Scanning Calorimeter (DSC) which measures the level of thermal stabilizers in the material.
- The DSC produces a graph of heat flow vs time.
- The time elapsed between the introduction of air into the cell and the decomposition of the sample reveals the time to oxidation which is then recorded as OIT.

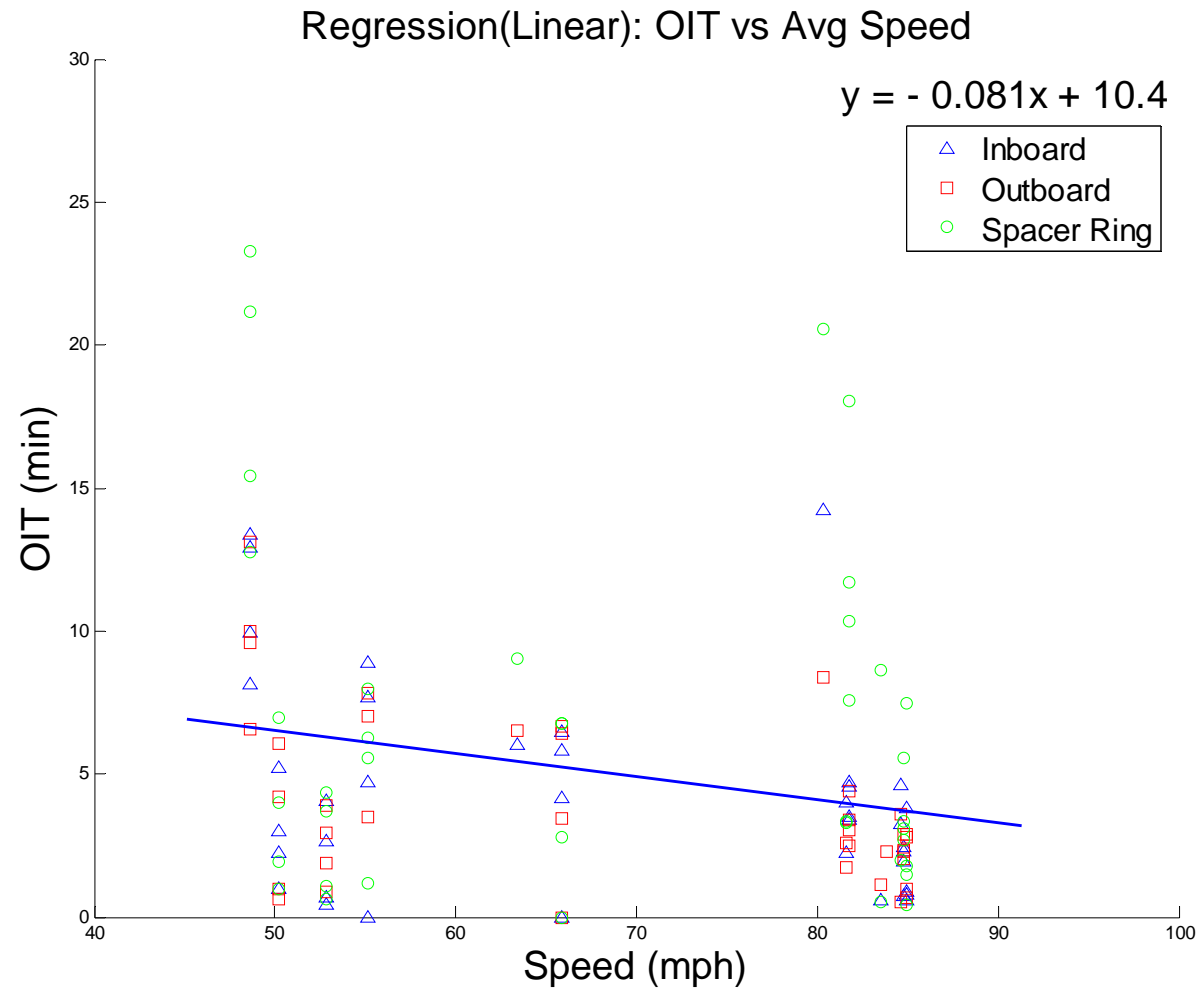


# Bearings

- Three samples come from each bearing, giving a total of twelve possible samples from each axle.
- Grease is sampled from the three critical locations of the bearing:
  - Inboard Cone Assembly Raceway
  - Outboard Cone Assembly Raceway
  - Spacer Ring Area



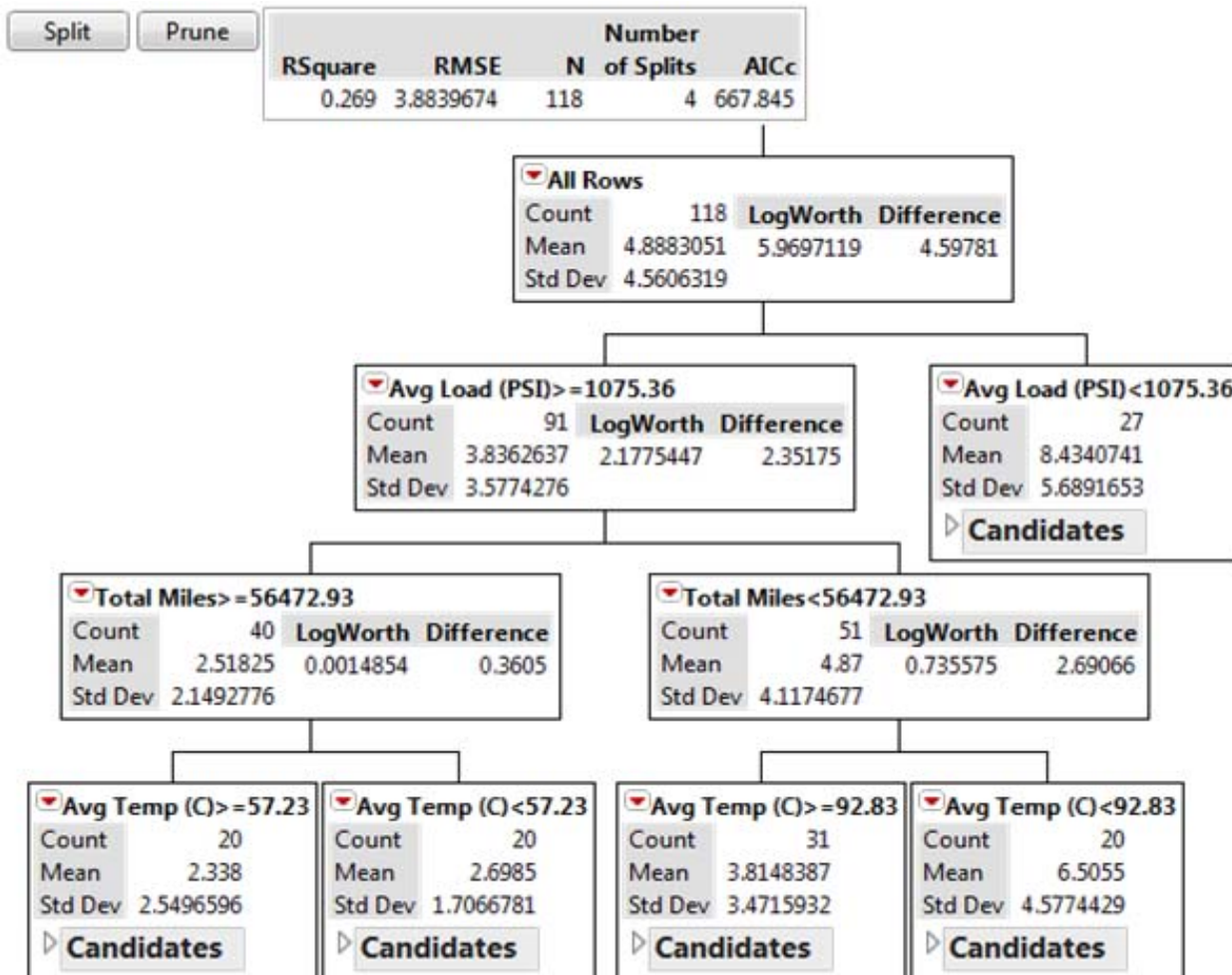
# Linear Regression Plot for OIT vs Speed



# Regression Tree

Min size split 20

- Ad



# Experimental Design

- Split, Split-plot Design
  - Whole plot: axle-setup
  - Sub plot: each bearing on axle
  - Sub, sub plot: sample location within each bearing
- Single replicate
- Unbalanced design

# Unbalanced Data

The screenshot displays the Microsoft Excel interface with the title bar "data - Excel". The ribbon includes tabs for FILE, HOME, INSERT, PAGE LAYOUT, FORMULAS, DATA, REVIEW, VIEW, and a custom tab labeled "INK TOOLS". The user's name "Douglas Timmer" is visible in the top right corner. The spreadsheet contains data organized into columns and rows, with some cells highlighted in yellow and others in blue. The data appears to be unbalanced, with varying numbers of observations across different categories. The status bar at the bottom indicates the spreadsheet is "READY" and shows a zoom level of 20%.

Category	Value	Count
1	1000	1000
2	1000	1000
3	1000	1000
4	1000	1000
5	1000	1000
6	1000	1000
7	1000	1000
8	1000	1000
9	1000	1000
10	1000	1000
11	1000	1000
12	1000	1000
13	1000	1000
14	1000	1000
15	1000	1000
16	1000	1000
17	1000	1000
18	1000	1000
19	1000	1000
20	1000	1000
21	1000	1000
22	1000	1000
23	1000	1000
24	1000	1000
25	1000	1000
26	1000	1000
27	1000	1000
28	1000	1000
29	1000	1000
30	1000	1000
31	1000	1000
32	1000	1000
33	1000	1000
34	1000	1000
35	1000	1000
36	1000	1000
37	1000	1000
38	1000	1000
39	1000	1000
40	1000	1000
41	1000	1000
42	1000	1000
43	1000	1000
44	1000	1000
45	1000	1000
46	1000	1000
47	1000	1000
48	1000	1000
49	1000	1000
50	1000	1000
51	1000	1000
52	1000	1000
53	1000	1000
54	1000	1000
55	1000	1000
56	1000	1000
57	1000	1000
58	1000	1000
59	1000	1000
60	1000	1000
61	1000	1000
62	1000	1000
63	1000	1000
64	1000	1000
65	1000	1000
66	1000	1000
67	1000	1000
68	1000	1000
69	1000	1000
70	1000	1000
71	1000	1000
72	1000	1000
73	1000	1000
74	1000	1000
75	1000	1000
76	1000	1000
77	1000	1000
78	1000	1000
79	1000	1000
80	1000	1000
81	1000	1000
82	1000	1000
83	1000	1000
84	1000	1000
85	1000	1000
86	1000	1000
87	1000	1000
88	1000	1000
89	1000	1000
90	1000	1000
91	1000	1000
92	1000	1000
93	1000	1000
94	1000	1000
95	1000	1000
96	1000	1000
97	1000	1000
98	1000	1000
99	1000	1000
100	1000	1000

# Parameter Estimation

- Restricted Maximum Likelihood (REML)
- Implemented in Matlab
- Degrees of Freedom are approximate due to unbalanced data

# Representation of Bearing Location in Regression Model

- The bearing location was recorded as a nominal value (1, 2, 3, 4)
- Modeled using three indicator variables

	Dummy Variables		
Bearing	X4	X5	X6
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

# Representation of Grease Location in Regression Model

- The grease location was recorded as a nominal value (1, 2, 3)
- Modeled using two indicator variables

	Dummy Variables	
Grease	X7	X8
1	0	0
2	1	0
3	0	1

# Initial Model

	Term	Coef	se(Coef)	t-statistic	approx p-val		Variance Componer	Estimate
wp terms (approx error df = 6)	Intercept	1.0737	3.7641	0.2853	0.7850			
	load	1.5122	7.1931	0.2102	0.8405		wp	15.33
	mileage	-13.2989	4.3152	-3.0819	0.0216 **		sp	1.87
	speed	1.7393	4.5440	0.3828	0.7151		ssp	5.59
	load*mileage	-9.3852	7.4705	-1.2563	0.2557			
	load*speed	-4.6215	7.1840	-0.6433	0.5438		Variance Ratios	Estimates
	mileage*speed	16.6564	7.5960	2.1928	0.0708 *		eta1	2.742
sp terms (approx error df = 24)	x4	2.8713	0.9978	2.8776	0.0083 **		eta2	0.334
	x5	0.6385	1.0011	0.6378	0.5296			
	x6	1.6466	0.9917	1.6604	0.1099			
ssp terms (approx error df = 72)	x7	2.9221	0.6333	4.6141	0.0000 **			
	x8	-0.3409	0.5432	-0.6276	0.5323			
	temperature	-8.7969	3.2085	-2.7417	0.0077 **			
	load*temperature	-3.9283	5.1582	-0.7616	0.4488			
	mileage*temperature	-7.8488	3.6546	-2.1476	0.0351 **			
	speed*temperature	2.7801	4.1258	0.6738	0.5026			
Analysis of coded variables								
	Obs	Approx DF	Approx Error DF					
WP	13	13	6					
SP	40	27	24					
SSP	118	78	72					

# Model 2

		Term	Coef	se(Coef)	t-statistic	approx p-val
wp terms	(approx error df = 9)	Intercept	0.6999	1.5747	0.4445	0.6672
		mileage	-8.0003	3.1156	-2.5678	0.0303 **
		speed	1.6769	1.8711	0.8962	0.3935
		mileage*speed	6.0180	3.7871	1.5891	0.1465
sp terms	(approx error df = 26)	x4	1.9142	0.7417	2.5808	0.0159 **
ssp terms	(approx error df = 75)	x7	2.9134	0.5294	5.5032	0.0000 **
		temperature	-6.6029	2.4734	-2.6696	0.0093 **
		mileage*temperature	-6.2692	2.4603	-2.5481	0.0129 **

Variance Components	Estimate
wp	13.680
sp	2.052
ssp	5.497

Variance Ratios	Estimates
eta1	2.4885
eta2	0.3733

## Analysis of coded variables

	Obs	Approx DF	Approx Error DF
WP	13	13	9
SP	40	27	26
SSP	118	78	75

# Model 3

		Term	Coef	se(Coef)	t-statistic	approx p-val
wp terms	(approx error df = 11)	Intercept	1.8964	1.3727	1.3815	0.1697
		mileage	-3.9511	1.9252	-2.0523	0.0423 **
sp terms	(approx error df = 26)	x4	1.8173	0.7385	2.4608	0.0208 **
		x7	2.7981	0.5002	5.5940	0.0000 **
ssp terms	(approx error df = 75)	temperature	-5.0227	1.9157	-2.6219	0.0106 **
		mileage*temperature	-3.5885	3.0580	-1.1735	0.2443

Variance Components	Estimate
wp	14.714
sp	2.089
ssp	5.500

Variance Ratios	Estimates
eta1	2.6753
eta2	0.3799

## Analysis of coded variables

	Obs	Approx DF	Approx Error DF
WP	13	13	11
SP	40	27	26
SSP	118	78	75

# Final Model

		Term	Coef	se(Coef)	t-statistic	approx p-val
wp terms	(approx error df = 11)	Intercept	2.3872	1.1846	2.0152	0.0690
		mileage	-3.8116	1.7281	-2.2057	0.0496 **
sp terms	(approx error df = 26)	x4	1.7551	0.7521	2.3336	0.0276 **
ssp terms	(approx error df = 76)	x7	2.7443	0.4986	5.5040	0.0000 **
		temperature	-3.7388	1.5485	-2.4145	0.0160 **

## Analysis of coded variables

	Obs	Approx DF	Approx Error DF
WP	13	13	11
SP	40	27	26
SSP	118	78	76

Variance Components	Estimate
wp	11.311
sp	2.231
ssp	5.599

Variance Ratios	Estimates
eta1	2.0202
eta2	0.3984

# Final Model

- $\widehat{OIT} = 2.3872 - 3.8116 * mileage' + 1.7551 * x_4 + 2.7443 * x_7 - 3.7388 * temperature'$
- Where
  - $mileage' = \frac{mileage - 53396}{45687}$
  - $temperature' = \frac{temperature - 80.16}{32.71}$
  - $x_4$  is 1 if bearing 2 location, 0 for other bearing locations
  - $x_7$  is 1 if grease sampling location is the spacer ring and 0 for the inner or outer raceway

# Future Research

- Model Diagnostics
  - Residual analysis
  - $R^2$
  - VIF
- Model Refinement
  - Why is bearing 2 statistically different?
  - Is temperature a covariate (function of load, mileage and speed)?
  - Developing second response variable related to length of grease molecule
  - Alternative Model: neural network or ensemble of neural networks

# Acknowledgements

- University Transportation Center for Railway Safety (<http://www.utrgv.edu/railwaysafety>) for their support of this research
- The Matlab code was provided by Dr. Marcus Perry, Associate Professor of Statistics, Culverhouse College of Commerce, University of Alabama