UNIVERSITY of DELAWARE

Big Data in Railway Infrastructure Maintenance Managing The Infrastructure

Dr. Allan M Zarembski, PE
Professor
Director of the Railroad Engineering and Safety Program
University of Delaware
dramz@udel.edu

Introduction

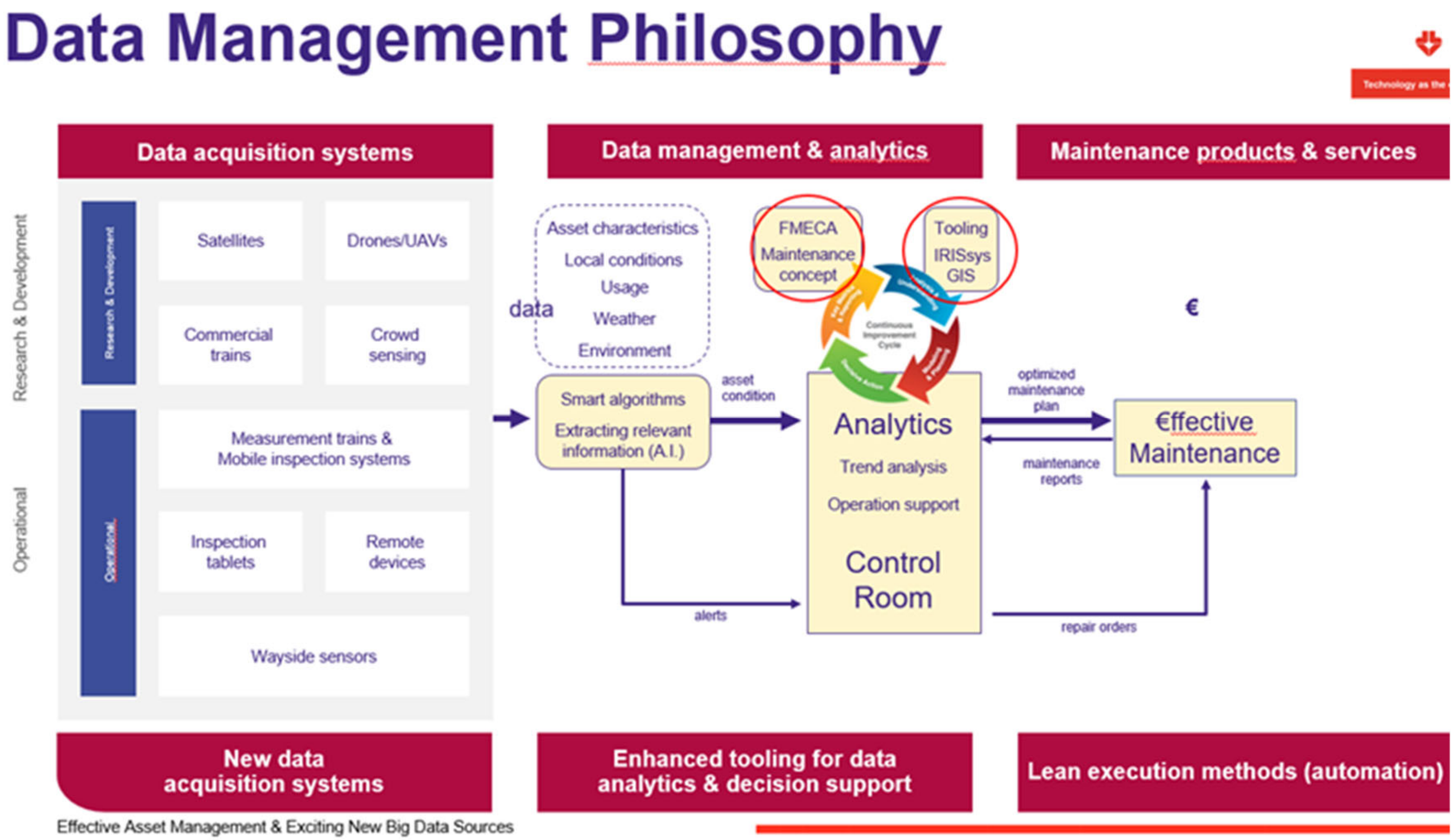
- Railroad industry is an infrastructure intensive industry that relies on significant amounts of information and data for operations and maintenance.
- In US, railroad data collection encompasses the full range of railroad activities
  - Monitoring over 30,000, 000 car loads (shipments) per year,
  - Managing railroad fleet of over 1.3 Million rail cars and 24,000 locomotives
  - Managing the infrastructure of over 330,000 km (200,000 miles) of track, which is owned and maintained by the railroads themselves.
    - Focus of this presentation
  - US railroad industry's annual revenues are of the order of $60 Billion
    - Annual capital program over $15 Billion a year.
- US represents approximately 25% of worldwide RR industry

Evolution of Infrastructure Inspection and Data Analysis

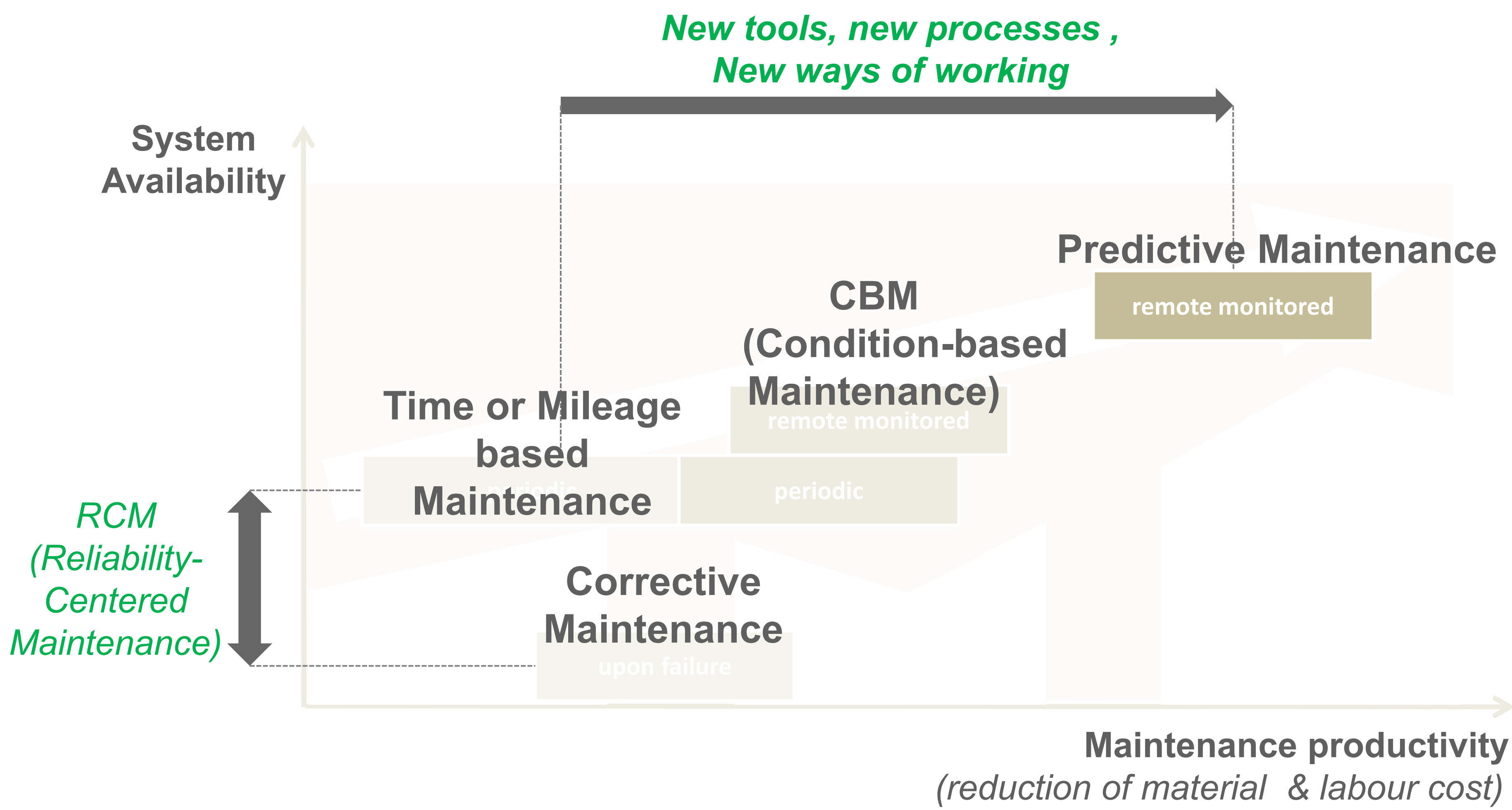
- RR management of the infrastructure has evolved from a subjective activity performed by a large labor force geographically distributed along the railroad lines, to an objective, technology active, data focused centrally managed activity.
- Current inspection makes use of a broad range of inspection vehicle to collect data
- New generation of maintenance management software systems analyzes and interprets this data
- Railroads represent an industry that is starting to make extensive use of its "big data"
  - to optimize its capital infrastructure and safely manage its operations while keeping costs under control.

UNIVERSITY of DELAWARE

Integration of Data Acquisition Systems and Data Analytics in Maintenance Actions

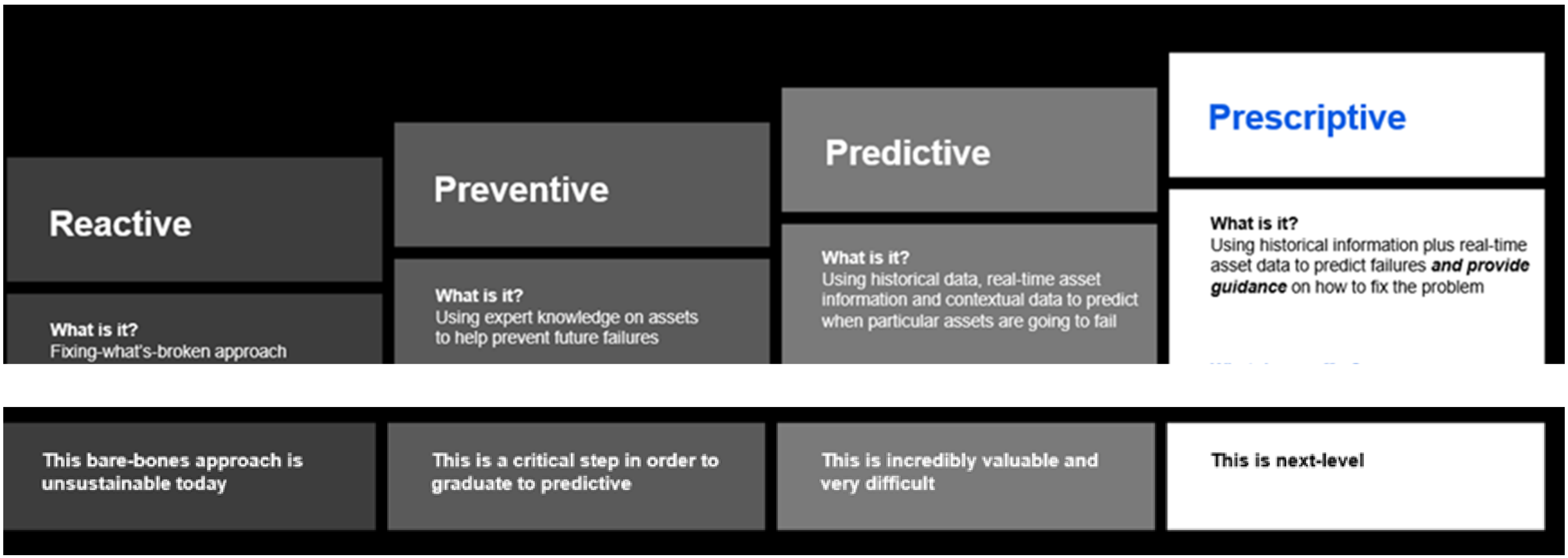


Predictive Maintenance for Rail Vehicles



Flix, Nicolas, "HealthHub, Shaped for Best and Easiest Control of Railways System Operations", Maintenance Engineering Director, Alstom Transport, Big Data in Railroad Maintenance Planning Conference, December 2017, University of Delaware, Newark DE.

Use of Data to Move from Reactive to Prescriptive Maintenance



Big Data in Railroad Track Infrastructure Inspection

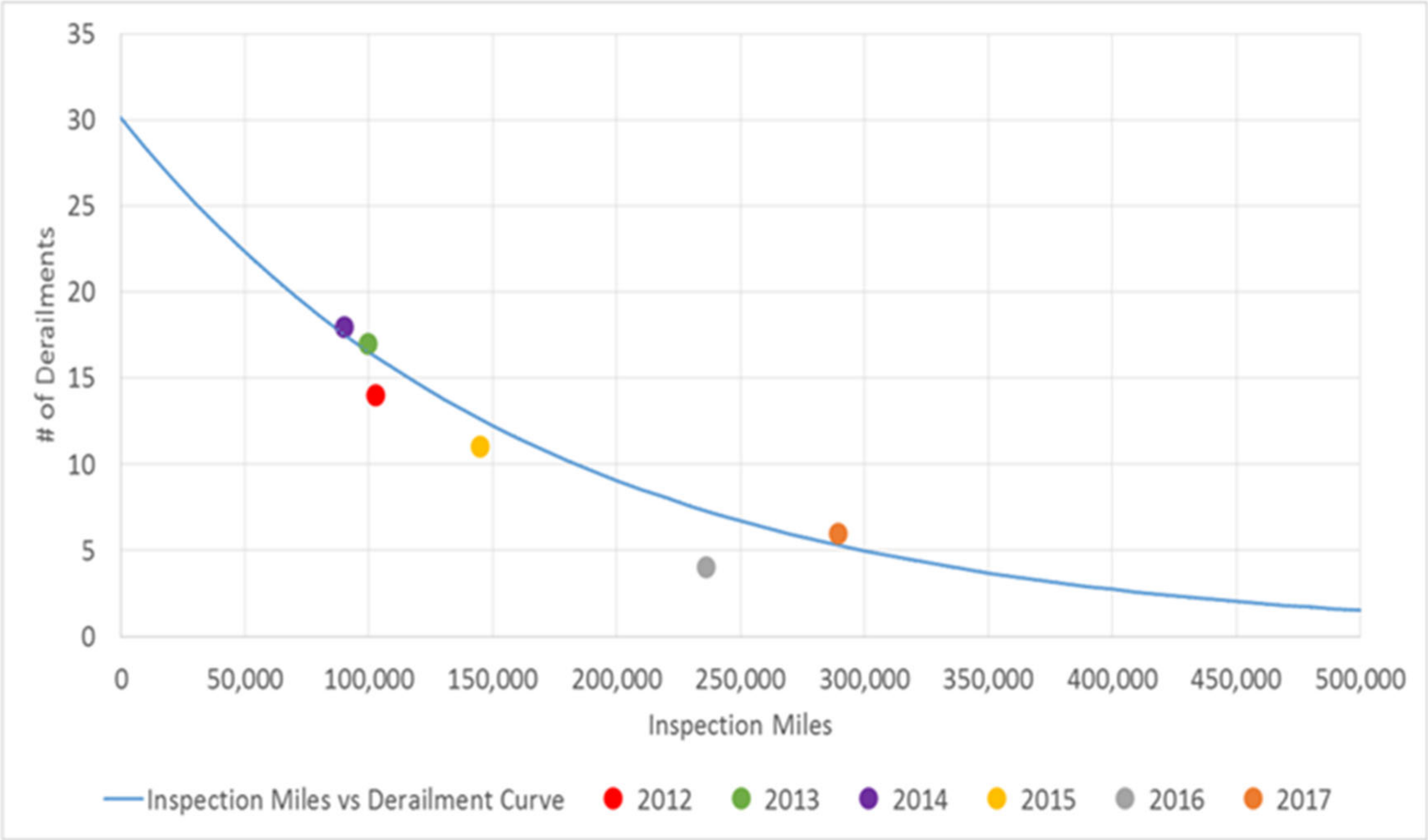
- Most infrastructure inspection is performed from rail inspection vehicles
  - High Speed track geometry inspection vehicles
    - Attended
    - Unattended
  - Ultrasonic rail test vehicles
  - Rail wear inspection vehicles (laser wear measurement)
  - Gauge restraint measurement vehicles
  - Ballast profile and subsurface inspection vehicles (LIDAR and GPR)
  - Tie (sleeper) inspection systems (Aurora)
  - Dynamic load measurement systems (V/TI, etc.)
  - Ride Quality measurement systems
  - Track stiffness measurement systems (M-Rail)
  - Video inspection data
  - Clearance measurement systems (LIDAR)

Example: Track Geometry Car Data

- Track geometry inspection vehicles, operate at track speeds of up to 130 kph (80 mph) for freight railroads and 200 kph (125 mph) for passenger railroads
- Main line tracks being measured from one to up to 12 times a year.
- Collect 10 to 12+ channel of data with a measurement taken as often as every foot.
- Represents over 2,500,000,000 data measurements per year of geometry only (Terrabytes of data)

UNIVERSITY of DELAWARE

Reduction in Track Caused Derailments as a Function of Increased Track Inspection



Supplemental Infrastructure Inspection

- Supplemented by track based measurements of vehicle condition such as:
  - Wheel load/impact detectors
  - Lateral force detectors
  - L/V detectors
  - Wheel profile measurement systems
  - Overheated bearing detectors
  - Dragging equipment detectors
- On a busy mainline a detector would measure over 3 Million wheels a year
- Helps identify rolling stock that causes disproportionate damage to the infrastructure

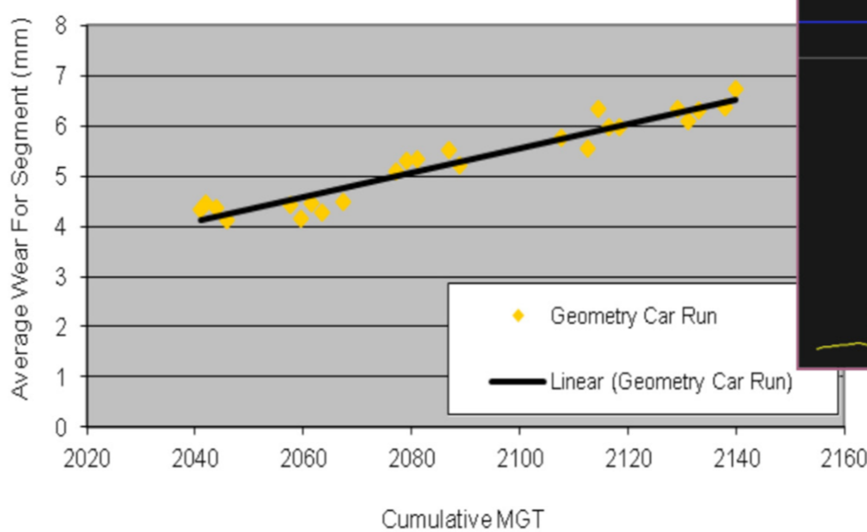
Case Studies

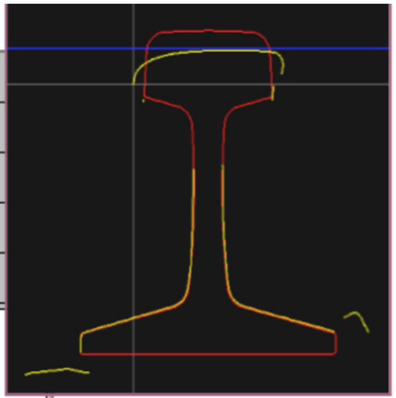
- Big Data Analysis of Track Degradation Behavior
  - Rail Wear
  - Track Geometry vs GPR measured Ballast Condition

UNIVERSITY of DELAWARE

Infrastructure Degradation: Rail Wear

- Rail continues to be one of the track's most vital assets
  - Key part of railway infrastructure
  - Replacement cost $500,000 to $1,000,000 per mile
- Currently inspected using laser/machine vision technology
  - Multiple times per year
  - Every 5'-15' (1.5 to 4.5 meters)
  - Cartesian coordinates of rail profile
- Generates terabytes of data
- Traditional modeling
  - Linear regression

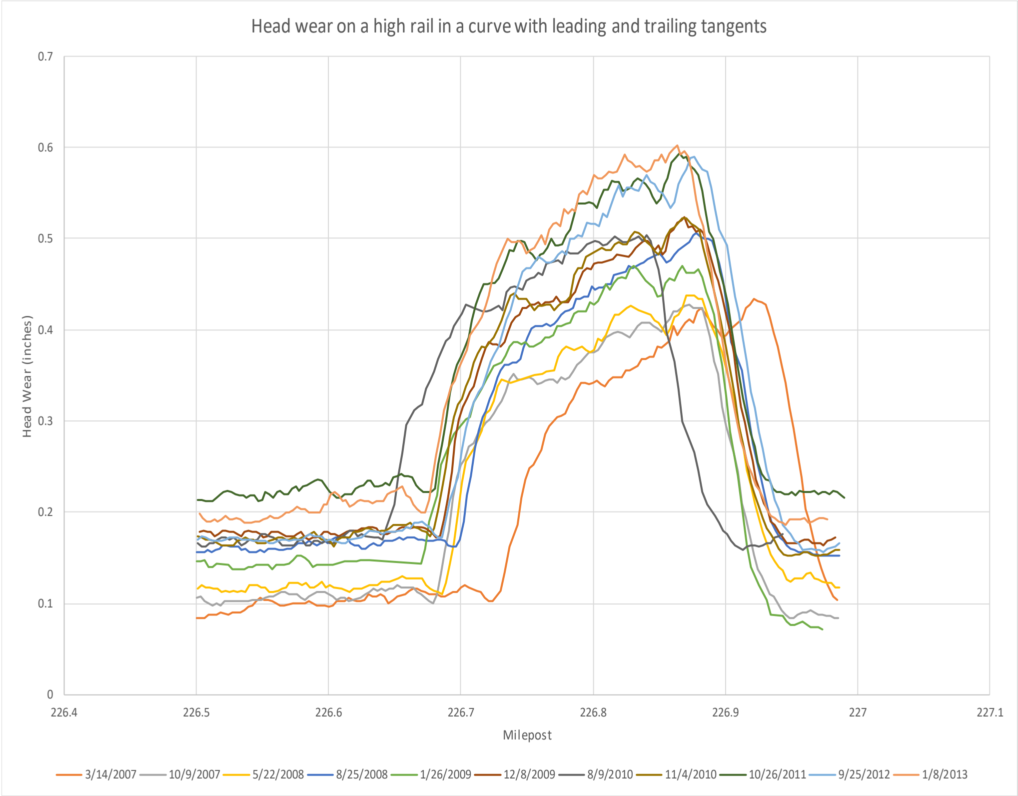




UNIVERSITY of DELAWARE

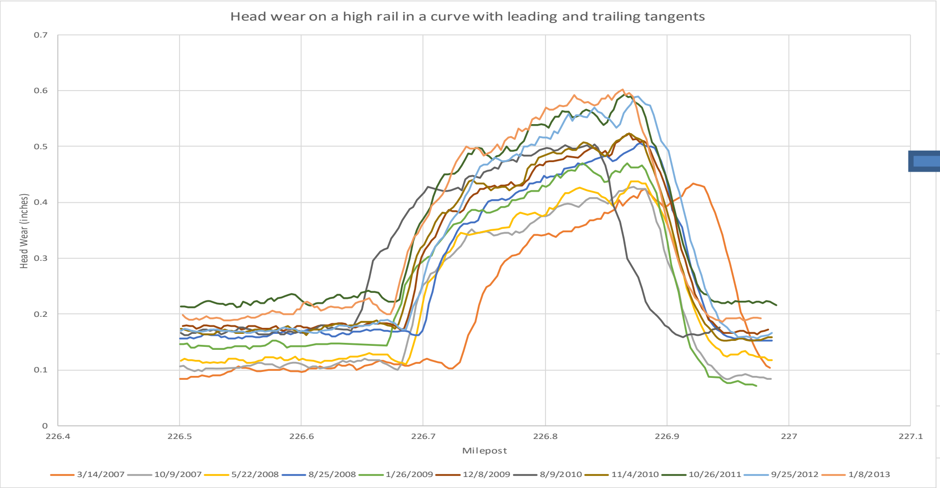
Rail Wear in Curve (over time)

- High rail head wear
- 11 measurements in 6 years
- Longitudinal misalignment
- Easily see increase in wear
  - Over time/MGT
  - Non uniform in curve

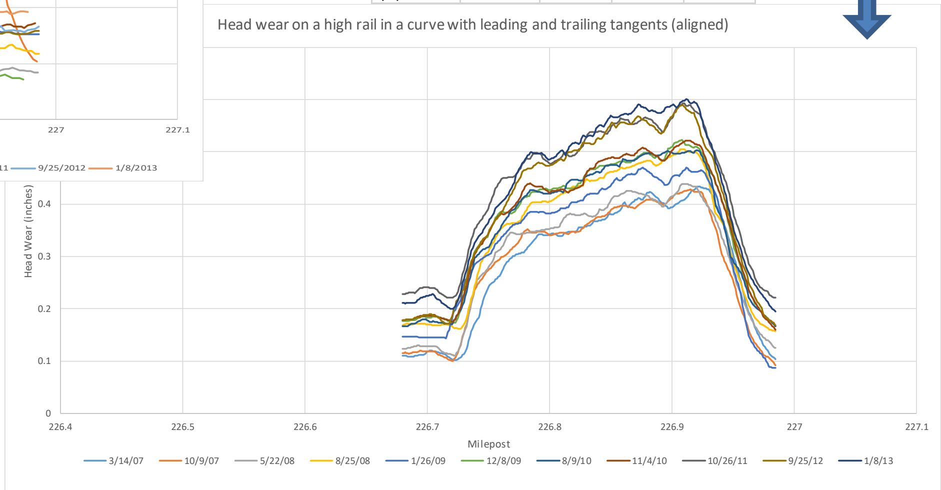


UNIVERSITY *of* DELAWARE

Aligned Data

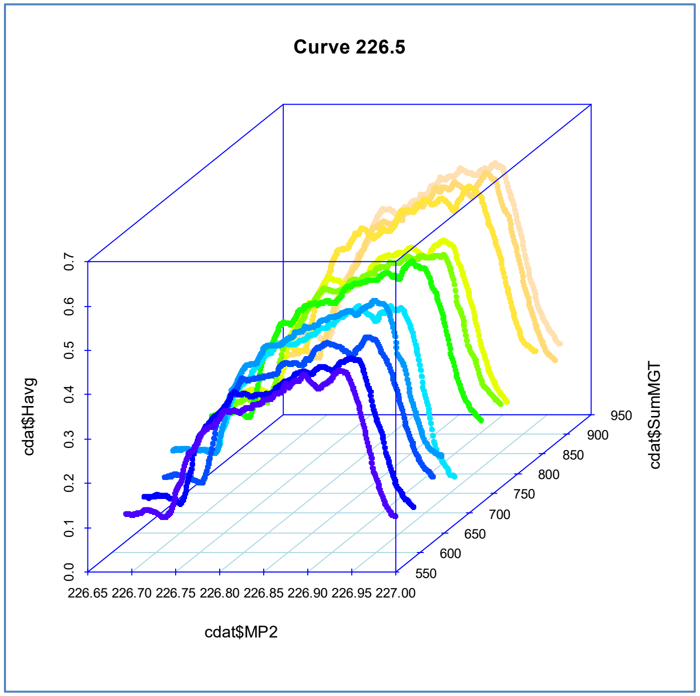


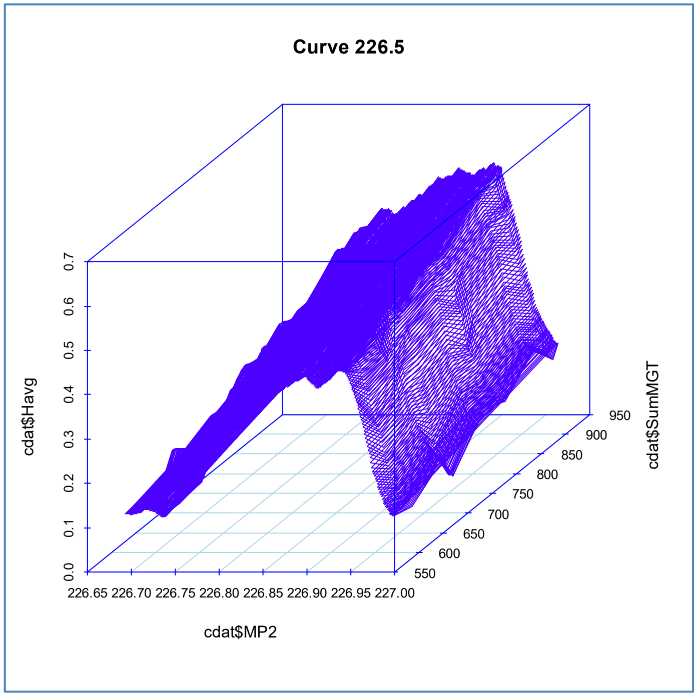
			Shift	
Insp Date	Corr Coef	Lag	Miles	Feet
3/14/07	Reference	0	0	0
10/9/07	0.82	42	0.042	222
5/22/08	0.86	37	0.037	195
8/25/08	0.87	31	0.031	164
1/26/09	0.78	45	0.045	238
12/8/09	0.84	39	0.039	206
8/9/10	0.69	80	0.080	422
11/4/10	0.82	44	0.044	232
10/26/11	0.83	44	0.044	232
9/25/12	0.86	32	0.032	169
1/8/13	0.80	49	0.049	259



Results in 400 samples for each inspection at same milepost locations with consistent distance interval

3-D Plot: Head Wear, MP, Sum MGT





ARIMA – Auto Regressive Integrated Moving Average

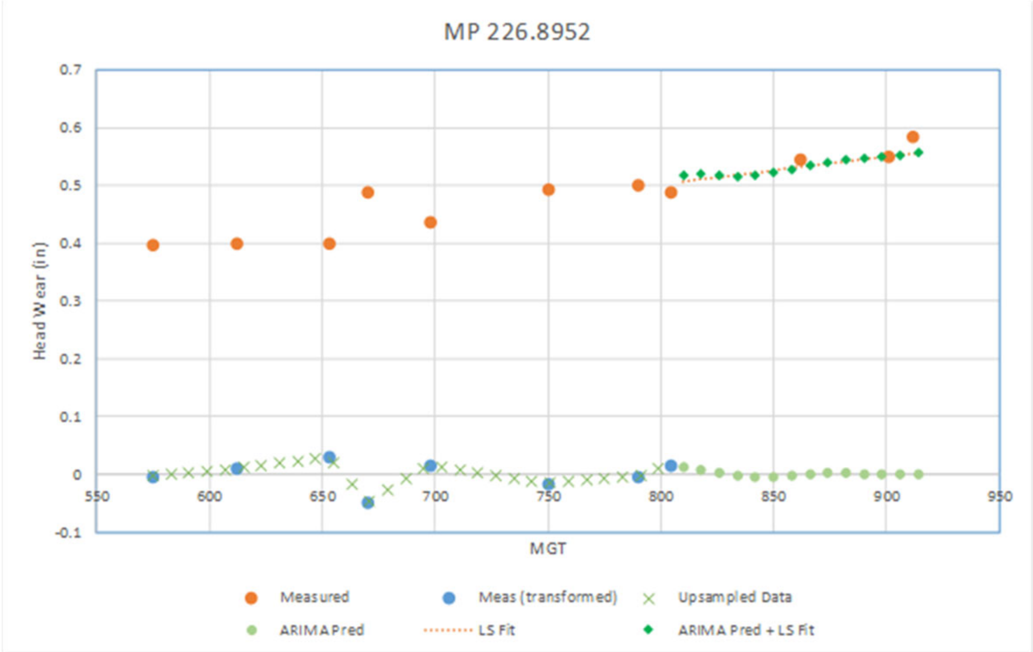
Make data stationary – remove linear trend

Fit function – linear interpolation

Up-sample – Common MGT interval

Perform ARIMA modeling

Predict next MGT intervals

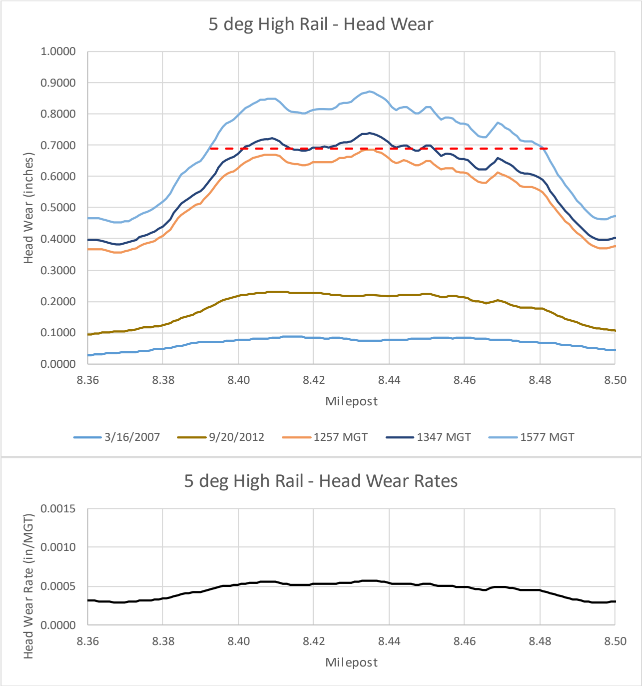


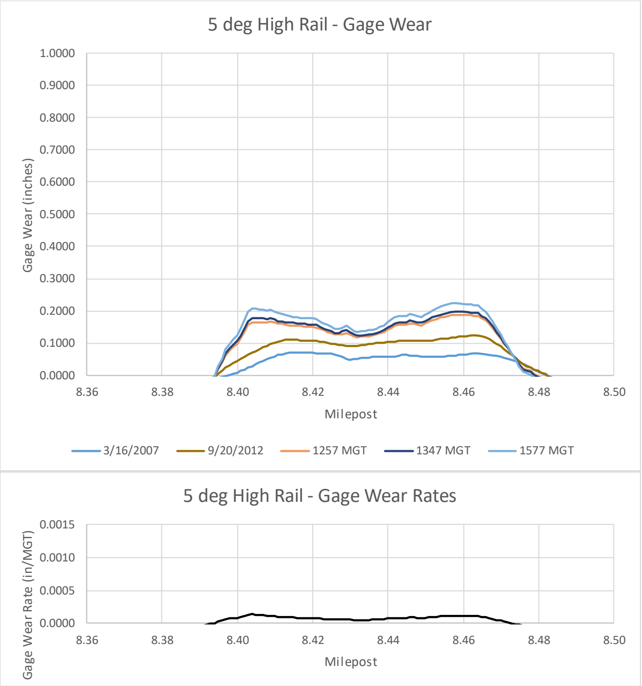
Handles calibration and measurement errors
Converges to linear trend in some cases
Model developed from 570-804 MGT
Forecast to 910 MGT to compare against next 3 measurements

UNIVERSITY of DELAWARE

ARIMA Forecast Example

5 Degree Curve

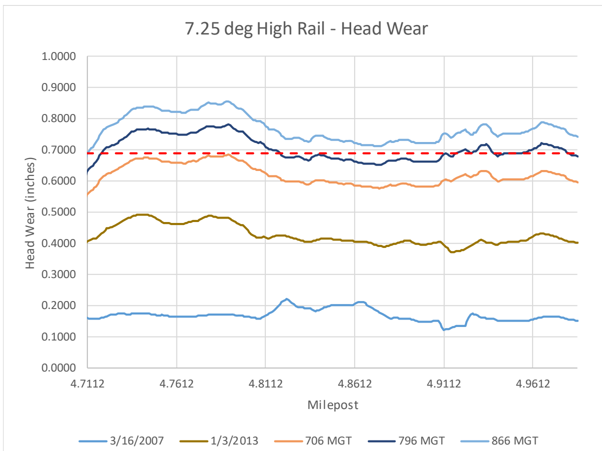


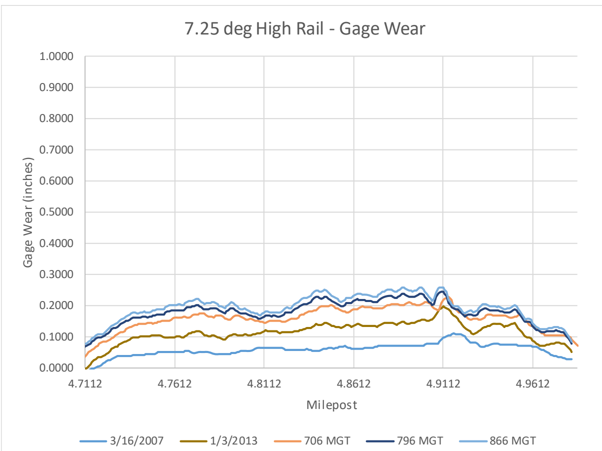


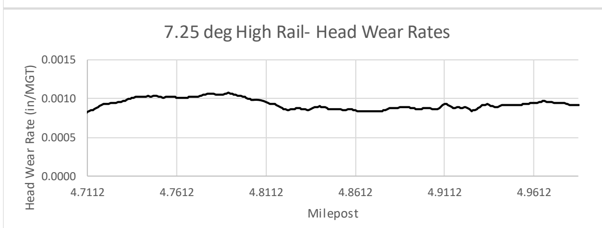
UNIVERSITY of DELAWARE

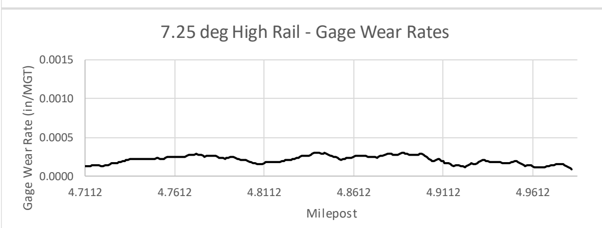
ARIMA Forecast Example

7.25 Degree Curve









Forecasting Replacement

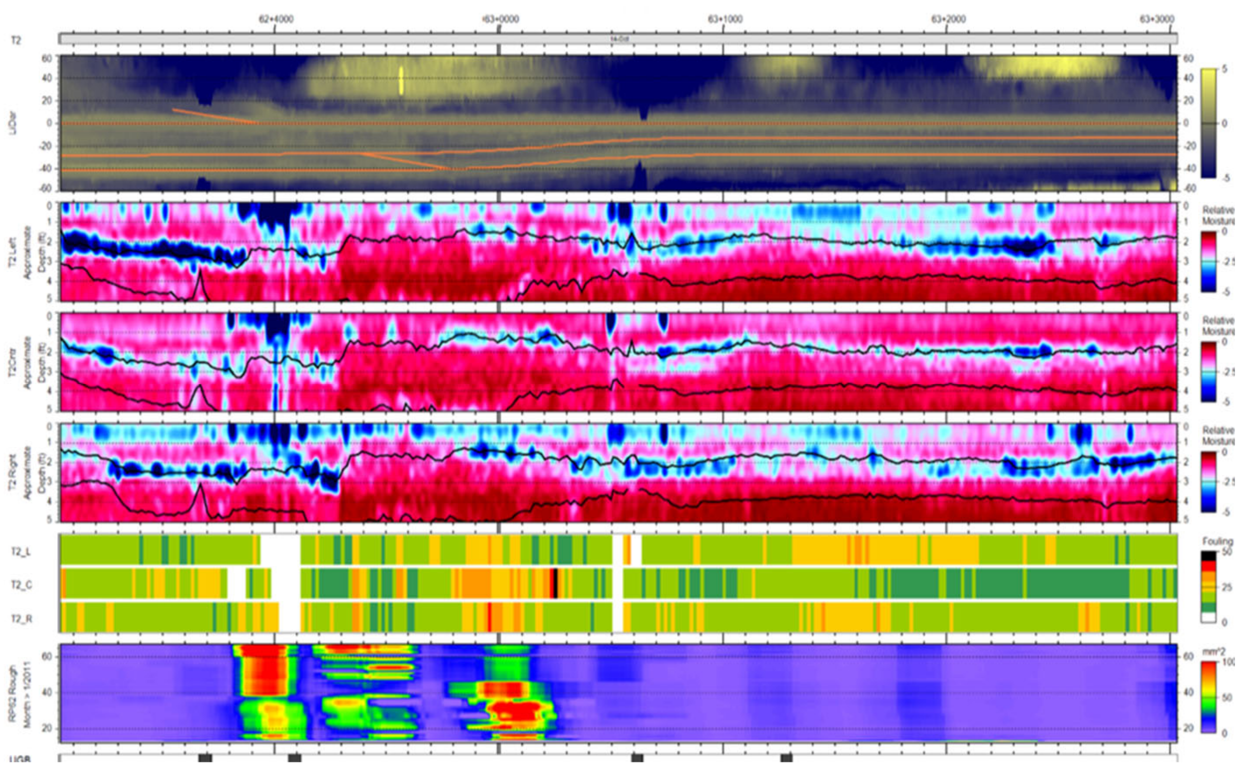
- 7.25 degree curve wear faster than 5 degree curve
  - 2x for head
  - 3x for gage
- Significant time (MGT) difference to replacement based on first hit, 50% of curve, 100% of curve
  - 6 to 10 year difference in remaining "life"

	Wear Rates (in/100MGT)			
	5 Deg Curve		7.25 Deg Curve	
	Head	Gage	Head	Gage
Minimum	0.0427	0.0003	0.0823	0.0092
Maximum	0.0574	0.0137	0.1074	0.0311
Average	0.0515	0.0081	0.0932	0.0211
Std Dev	0.0036	0.0028	0.0069	0.0053
	MGT to Head Wear Limit			
First Hit	1257		706	
50%	1347		796	
100%	1577		866	
	Replacement Date (30 MGT/Yr)			
First Hit	4/21/2038		12/9/2019	
50%	4/20/2041		12/8/2022	
100%	12/17/2048		4/8/2025	

UNIVERSITY of DELAWARE

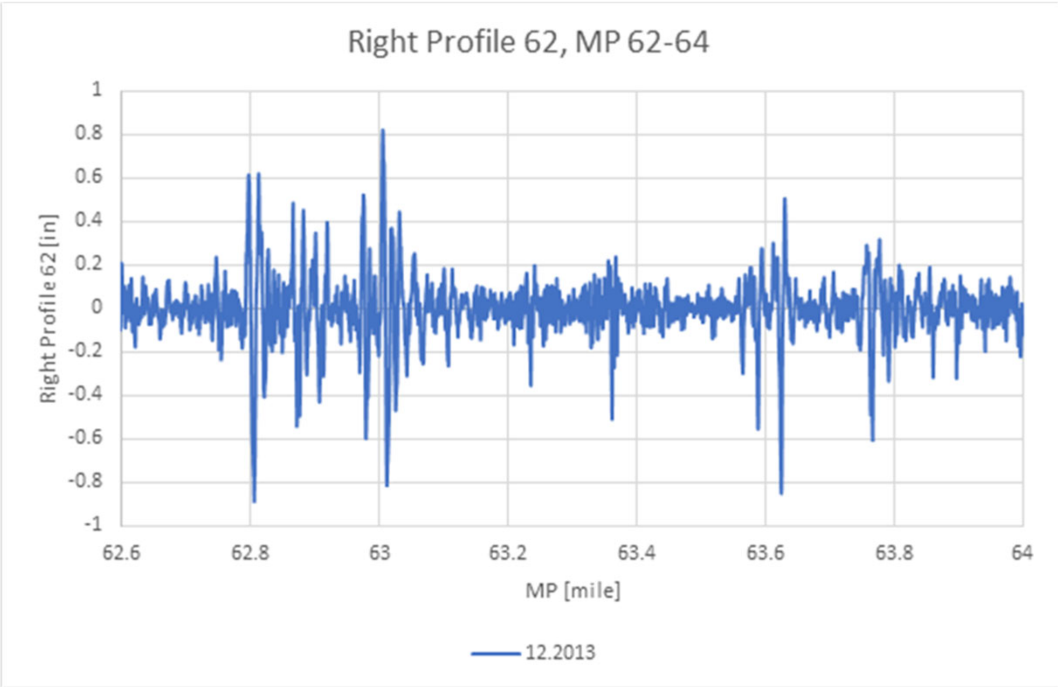
Predicting Development of Track Geometry using GPR measured Ballast Condition

- How does Ballast and subgrade condition as measured by ground penetrating radar (GPR) influence the probability of a track geometry anomaly occurring
- GPR measures:
  - Ballast Fouling
  - Ballast layer thickness
  - Ballast Moisture content
- Focus of profile (surface)



Track Geometry Data

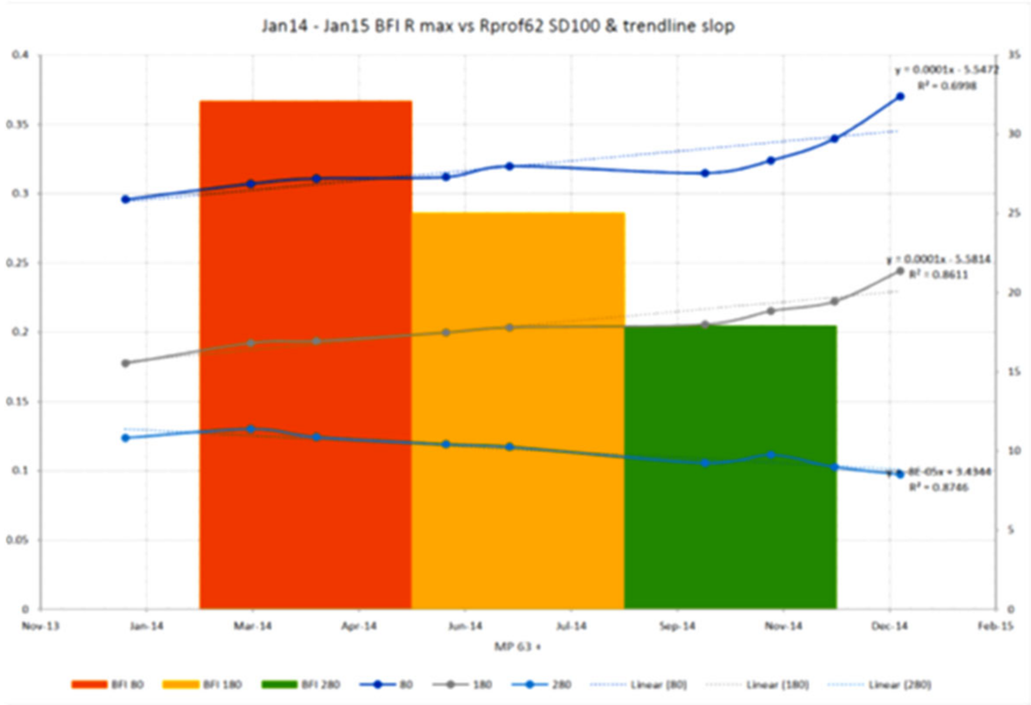
- Right Profile, 62' (20 meter) Chord
- Data every foot
- Inspection date associated with degraded track section



UNIVERSITY of DELAWARE

Preliminary Analysis

- Examine three track segments
  - Highly fouled (red)
  - Moderately fouled (yellow)
  - Relatively clean (green)
- Develop Track Quality Index (TQI) for Right Profile, 62’ chord
  - One year trend in TQI for each track segment
- Magnitude and trend of TQI show strong correlation to BFI



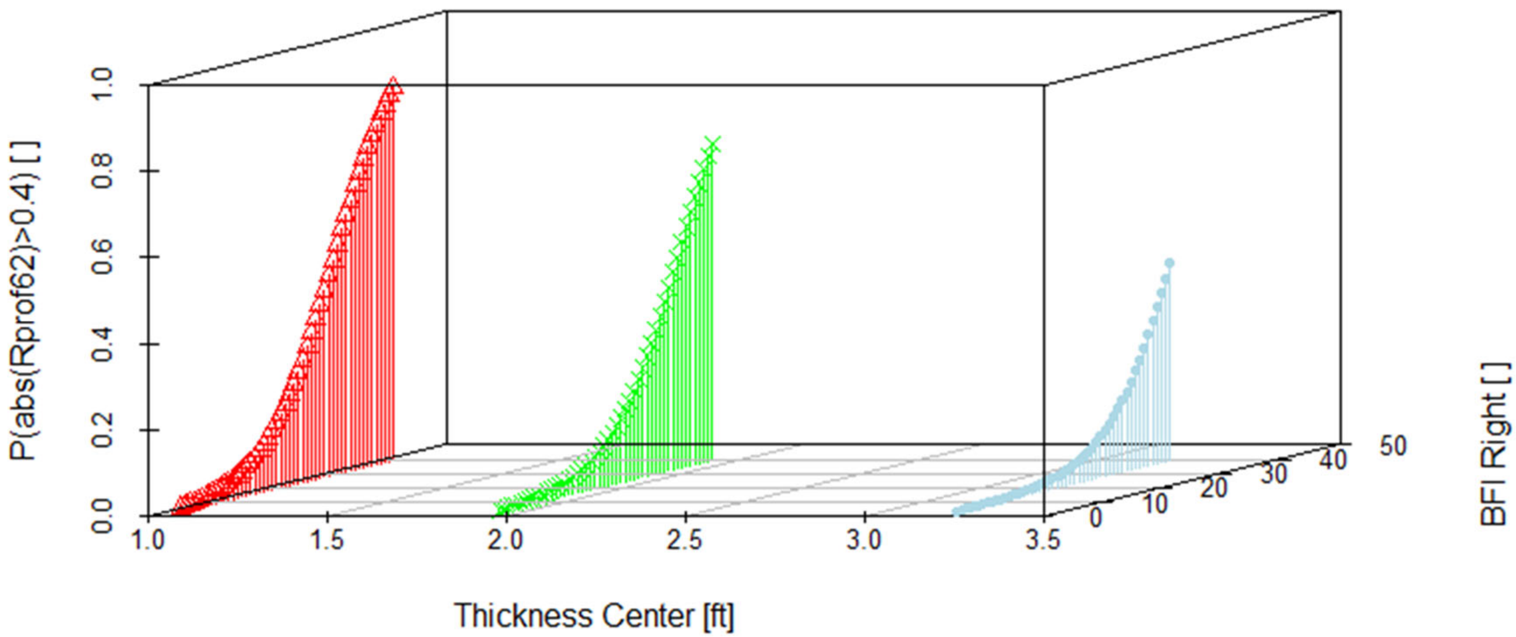
Logistic Regression Model

- Stochastic process for categorizing dependent on/off or binary (0/1) conditions to primary independent variables
- Gives the probability of the binary event as a function of independent variable
- Probability geometry defect > 0.4"
- Probability of developing geometry defect as function of:
  - Ballast Fouling Index (BFI)
  - Ballast Layer thickness (BLT)

$$logit(P) = \ln\left(\frac{P}{1-P}\right) = -4.98 + 0.04 \cdot BFI_{center} + 0.18 \cdot BFI_{right} - 0.92 \cdot BLT_{center}$$

$$\hat{P}_{geometry} = \frac{e^{-4.98+0.04 \cdot BFI_{center} + 0.18 \cdot BFI_{right} - 0.92 \cdot BLT_{center}}}{1 + e^{-4.98+0.04 \cdot BFI_{center} + 0.18 \cdot BFI_{right} - 0.92 \cdot BLT_{center}}}$$

Logit Model Sensitivity Analysis – P(Rprof) vs BFI R: 3-D Plot



Statistical Validation

- Confusion Matrix
  - 29 false positives
  - 4 missed positives
- Overall accuracy
  - 87% matched prediction

		True condition	
	Total population	**Condition positive**	**Condition negative**
Predicted condition	**Predicted condition positive**	**213**	**29**
	Predicted condition negative	**4**	**7**

Accuracy
86.96%

True positive rate (TPR), Sensitivity	False Positive Rate (FPR), probability of false alarm
98.16%	80.56%
False negative rate (FNR), Miss rate	True negative rate (TNR), Specificity (SPC)
1.84%	19.44%

Practical Use

- Constant Probability as a function of condition
  - Ballast Fouling Index Right
  - Ballast Layer Thickness Center
- Shows combination of ballast layer thickness and ballast fouling index right at which a defined probability of a right profile – 62 > 0.4 (10 mm) will occur.