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
© 2018 Allan M Zarembski
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.
Integration of Data Acquisition Systems and Data Analytics in Maintenance Actions

Data Management Philosophy

Data acquisition systems
- Satellites
- Drones/UVs
- Commercial trains
- Crowd sensing
- Measurement trains & Mobile inspection systems
- Inspection tablets
- Remote devices
- Wayside sensors

Data management & analytics
- Asset characteristics
- Local conditions
- Usage
- Weather
- Environment
- FMECA
- Maintenance concept
- Tooling
- IRISys
- GIS
- Smart algorithms
- Extracting relevant information (A.I.)
- asset condition
- optimized maintenance plan
- maintenance reports

Control Room
- Trend analysis
- Operation support
- Alerts
- Repair orders

Effective Maintenance

Maintenance products & services

Lean execution methods (automation)

Enhanced tooling for data analytics & decision support

Effective Asset Management & Exciting New Big Data Sources
Predictive Maintenance for Rail Vehicles

New tools, new processes, New ways of working

System Availability

Time or Mileage based Maintenance

RCM (Reliability-Centered Maintenance)

Corrective Maintenance upon failure

CBM (Condition-based Maintenance)

remote monitored

Corrective Maintenance

periodic

Maintenance productivity
(reduction of material & labour cost)

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)
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
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
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
Aligned Data

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
ARIMA Forecast Example

5 Degree Curve

5 deg High Rail - Head Wear

5 deg High Rail - Gage Wear

5 deg High Rail - Head Wear Rates

5 deg High Rail - Gage Wear Rates
ARIMA Forecast Example

7.25 Degree Curve

7.25 deg High Rail - Head Wear

7.25 deg High Rail - Gage Wear

7.25 deg High Rail - Head Wear Rates

7.25 deg High Rail - Gage Wear Rates
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”

<table>
<thead>
<tr>
<th>Wear Rates (in/100MGT)</th>
<th>5 Deg Curve</th>
<th>7.25 Deg Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head</td>
<td>Gage</td>
<td>Head</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0427</td>
<td>0.0003</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.0574</td>
<td>0.0137</td>
</tr>
<tr>
<td>Average</td>
<td>0.0515</td>
<td>0.0081</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.0036</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MGT to Head Wear Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Hit</td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Replacement Date (30 MGT/Yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Hit</td>
</tr>
<tr>
<td>50%</td>
</tr>
<tr>
<td>100%</td>
</tr>
</tbody>
</table>
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
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)

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

\[
\hat{P}_{\text{geometry}} = \frac{e^{-4.98+0.04\cdot BFI_{\text{center}} +0.18\cdot BFI_{\text{right}}-0.92\cdot BLT_{\text{center}}}}{1+e^{-4.98+0.04\cdot BFI_{\text{center}} +0.18\cdot BFI_{\text{right}}-0.92\cdot BLT_{\text{center}}}}
\]
Logit Model Sensitivity Analysis – $P(R_{\text{prof}})$ vs BFI R: 3-D Plot
Statistical Validation

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

<table>
<thead>
<tr>
<th>Predicted condition</th>
<th>Condition positive</th>
<th>Condition negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>213</td>
<td>29</td>
</tr>
<tr>
<td>Predicted condition positive</td>
<td>4</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition positive</td>
<td>213</td>
</tr>
<tr>
<td>Condition negative</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>86.96%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True positive rate (TPR), Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98.16%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>False positive rate (FPR), probability of false alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80.56%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>False negative rate (FNR), Miss rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.84%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>True negative rate (TNR), Specificity (SPC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19.44%</td>
</tr>
</tbody>
</table>
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.