# Big Data in Railroad Engineering

**Dr. Allan M Zarembski**
**Director of Railroad Engineering and Safety Program**
***Department of Civil and Environmental Engineering***
**University of Delaware**
**Newark, 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 20% of worldwide RR industry

# Evolution of Infrastructure Data Collection

- RR inspection and 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.

# Major US Railroads

- Six largest US railroads have between 20,000 and 40,000 miles of track (30,000 and 60,000+ km) each
  - Larger than most national railroads
- Data management and analysis of big data has become of growing importance for these major railroads.

# Infrastructure Inspection

- Most infrastructure inspection is performed from rail inspection vehicles
  - High Speed track geometry inspection vehicles
  - 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
  - Dynamic load measurement systems

# Supplemental Infrastructure Inspection

- Supplemented by track based measurements of vehicle condition such as:
  - Wheel load/impact detectors
  - Lateral force detectors
  - L/V detectors
  - Overheated bearing detectors
  - Dragging equipment detectors
- On a busy mainline a detector would measure over 3 Million wheels a year

# *Track Geometry Data*

- On board measurements every foot. Based on a system average frequency 1 inspection per mile per year, this would represent over 100,000,000 measurements per year with at least 12 channels of data collected at each measurement.
- Recorded exception data, stored in an active data base, represents approximately 70,000 measurements per year with at least 12 channels of data collected at each measurement

# *Rail Defect Data*

- On board measurements on a continuous basis. Based on a system average frequency 1 inspection per km per year, this would represent over 36,000 km of inspection data
- Recorded exception or defect data, stored in an active data base, represents approximately 20,000 data sets per year.

# Rail Inspection Data

- In addition to rail defect data, railroads now collect rail profile and wear data at the same frequency as track geometry data
  - Rail profile measurement systems mounted on track geometry cars
- Within the last 30 years US Railroads have gone from 3GB to almost 3000 GB (3 TB) of rail measurement data per annum
- This will continue to grow to include other elements like special track work where higher inspection density is required.

# *Vertical Track Interaction (VTI) Data*

- Represents vehicle-track dynamic data as recorded by an inspection vehicle
  - Can be unmanned vehicle mounted Vertical Track Interaction measurement systems.
- Data is collected continuously but only values that exceed specific exception values set by the railroad are recorded.
- Based on partial coverage of the network, represents approximately 1,000, 000 stored data records per year.

# Cross-Ties (Sleepers)

- Cross-ties represent another area of Big Data in railroads.
- Typically there are 3250 ties per mile so that a railroad with 22,000 miles would have over 70 million ties.
- These ties are usually inspected on a four to five year cycle,
  - 15 to 17 Million ties per year are inspected as to their condition and whether they need replacement.
- This data collected is uploaded into the railroad system database and then used to determine required annual replacement ties by mile of track.
  - Typically, a railroad of this size would replace 2 Million ties a year.

# Levels of Data Analysis

- At the first level, basic threshold analyses are performed to determine if the measured value exceed a predefined threshold to include both maintenance and safety thresholds
- At the second level, this data is entered into large data bases to allow for historical monitoring, trend analysis and first generation forecasting of rates of degradation or failure
- At the third level, this data is used in state of the art statistical analyses such a multivariate regression analysis or Multivariate Adaptive Regressive Splines (MARS) analysis to develop higher order forecasting and trend analysis
- At the next level, these forecasting models are combined with maintenance planning models for determination of maintenance requirements and scheduling of maintenance activities across railroad
  - Maintenance planning and management models often combine economic analyses with the projected failure analyses to calculate the optimum maintenance and replacement requirements

# Big Data Analyses Analysis

- As the size and extent of the data bases continue to grow, more refined statistical analyses such a multivariate regression analysis or Multivariate Adaptive Regressive Splines (MARS) analysis are used to develop higher order forecasting and trend analysis

# Sample MARS Analysis

- a MARS application to geometry and rail defect data for a big data application, representing over 500,000 data records

$$MGT = 477.37 + 86.6031(BF_2) + 28.6544(BF_4) - 47.8978(BF_5)$$

$BF_2 = max(0, 6 - WARP)$

*WARP = number of WARP Defects*

$BF_4 = max(0, 18 - RAIL\_CANT)$

*RAIL_CANT = number of rail cant defects*

$BF_5 = max(0, ALIGNMENT - 0)$

*ALIGNMENT = number of alignment defects*

*Max MGT ( no defects) = 1513 MGT*

# Summary

- Railroad industry has entered into the era of "Big data" with large data bases and large volumes of data
- Strong need to separate derive information from "mountain of data"
- As inspection technologies improve and become more widespread, this volume of data will continue to increase
- Need for Big Data type analyses tools to address this data challenge
- Mini-conference on Big Data in RR Maintenance Planning scheduled for Decembers 2015 at University of Delaware